21 世纪高等学校电子信息类规划教材

信息论与编码

平西建　童　莉
巩克现　马金全　编著

西安电子科技大学出版社

2009

内容简介

本书系统地讨论了 Shannon 信息理论中的基本概念和相关问题，介绍了信道、信源编码的一般原理和基本方法。全书主要内容分为 8 章，包括绪论、离散无记忆信源与信息熵、离散无记忆信道与互信息、信道与信道容量、有噪信道编码、离散信源及其信息冗余、无失真信源编码和限失真信源编码。

本书可作为信息工程、通信工程、网络工程和计算机应用等相关专业高年级本科生及研究生的教材或教学参考书，也可作为从事信息理论、信息技术和通信系统研究的科研及工程技术人员的参考用书。

★ 本书配有电子教案，有需要者可登录出版社网站，免费提供。

图书在版编目(CIP)数据

信息论与编码/平西建等编著. —西安：西安电子科技大学出版社，2009.2

21 世纪高等学校电子信息类规划教材

ISBN 978-7-5606-2182-1

Ⅰ. 信… Ⅱ. 平… 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(2008)第 208018 号

策　　划　薛　媛

责任编辑　许青青　薛　媛

出版发行　西安电子科技大学出版社(西安市太白南路 2 号)

电　　话　(029)88242885　88201467　　邮　　编　710071

网　　址　www.xduph.com　　电子邮箱　xdupfxb001@163.com

经　　销　新华书店

印刷单位　陕西天意印务有限责任公司

版　　次　2009 年 2 月第 1 版　2009 年 2 月第 1 次印刷

开　　本　787 毫米×1092 毫米　1/16　印　张　15.75

字　　数　370 千字

印　　数　1～4000 册

定　　价　23.00 元

ISBN 978-7-5606-2182-1/TN・0481

XDUP 2474001-1

前　言

随着信息技术的发展，信息的表示、存储、传输和处理成为多媒体技术、通信技术、信息处理、模式识别、网络工程以及计算机应用等领域的关键技术和热点研究课题。信息理论应用概率论、随机过程和数理统计等方法研究信息的描述、传输与处理中的一般规律以及基本关系，还可用于研究信息系统的有效性和可靠性，是信息科学研究和各类信息系统开发中具有重要指导作用的基础理论。

本书主要讨论 Shannon 信息理论中的基本概念和问题，主要内容分为 8 章。第 1 章介绍了信息理论的主要研究内容、发展历史和现状。第 2、3 章讨论 Shannon 信息理论中两个最重要的基本概念，即信息熵和互信息的定义与性质，通过深入分析平均互信息与信源概率分布和信道转移概率分布的凸性关系，引出了信息系统的可靠性和有效性问题。第 4、5 章讨论了信道和信息系统的可靠性问题，首先依据互信息的上凸性关系，引出了给定信道的传输信息能力并给出信道容量的定义，阐述了 Shannon 有噪信道编码定理，然后针对有噪信道中信息可靠传输的要求，比较系统地介绍了纠错编码的基本概念和线性分组码、循环码、BCH 码及卷积码等信道编码的构造原理。第 6～8 章讨论了信源和信息系统的有效性问题。在分析离散无记忆信源及其扩展信源、平稳信源和马尔可夫信源统计关系的基础上，分析并度量了信源的信息冗余，引出了信息系统的有效性问题，针对无失真表示信源信息的应用，讨论了无失真信源编码定理和编码方法。依据互信息的下凸性关系和失真度量方法，讨论了信息率失真函数的定义和性质，阐述了限失真信源编码定理，并介绍了限失真信源编码的基本原理与方法。作为学习上述内容的预备知识，书后附录中介绍了凸函数、Jensen 不等式和线性代数基础(群、环、域)等知识。

本书在内容编排上力图将信息理论与信息系统的最优化问题有机结合，注重信息论的基本概念与信息系统的可靠性、有效性等应用技术原理之间的联系，使相关专业的读者能够运用信息理论的基本概念分析信息系统的最优化问题，系统地了解信道、信源编码的一般原理和基本方法，为读者进一步学习信息与通信工程、网络工程、计算机应用等领域的专业课程，从事图像、语音等多种媒体数据的处理、编码技术研究和信息系统开发工作打下良好的基础。

本书作者长期从事“信息论与信源编码”和“信道编码”课程的教学工作，为了满足信息工程、通信工程、网络工程和计算机应用等相关专业人才培养的教学需要，在多年使用的《信息论基础与信源编码原理》、《信道编码》等教材的基础上编写了本书。本书由平西建、董莉、巩克现、马金全共同完成。刘玉君、肖永隆、邵美珍等老师在前期的教学和教材编写中做了大量的工作，张涛、周利莉、许漫坤对本书的编写提出了有益的建议，汪然、王绪和刘学谦等多名在校研究生参与了书稿的整理工作并绘制了部分插图。

我国信息科学界的不少著名学者在信息理论方面有很好的著作，这些著作对于作者从事信息论课程的教学和在相关技术领域的学术研究以及本书的编写都给予了很大的帮助。西安电子科技大学出版社的薛媛和许青青编辑为本书的出版做了大量的工作。

在此谨向对本书给予帮助和支持的所有人员表示感谢。

由于作者水平和时间所限，书中难免有不妥之处，诚恳期望读者赐教和指正。

作　者

2008 年 10 月

目　　录

第1章 绪 论

信息科学和工程以信息为研究对象，以信息的运动规律和利用信息的原理为主要研究内容，以扩展人类的信息处理能力为主要研究目标，是当代具有强大生命力和推动力的新兴学科。作为半个多世纪科学技术进步的主要标志之一，信息科学和工程广泛地应用于通信理论与技术、计算机科学与技术、电子工程、遥感遥测技术和人工智能等各个领域。信息科学和工程的一个重要的基础理论——信息论，是在长期的信息与通信工程实践中，与概率论、随机过程和数理统计等近代数学学科相结合而建立并发展起来的，它主要研究各类电子信息系统和通信系统中信息的描述、传输和处理的一般规律与基本关系。研究信息系统的有效性和可靠性理论的目的是实现系统的最优化。在今天的“信息化社会”中，信息理论对于各类实用信息系统的研制与应用都具有重要的指导作用。作为信息论的主要分支，信源编码着重研究信源的统计特性和信源的编码理论与数据压缩技术，是提高信息系统有效性的技术手段；信道编码研究如何检测和纠正信息通过有噪信道传输后发生的错误，是提高信息系统可靠性的技术手段。信道编码、信源编码是各类电子信息系统研究和开发中的关键技术课题。

本章我们首先引出信息理论中最基本、最重要的概念——信息，简述信息理论及其主要应用领域——信源编码的主要研究内容、历史发展及现状。

1.1 信息的定义与性质

人类在日常生活和社会活动中，一时一刻也离不开信息的交流。特别是在今天的信息社会中，人们要有效地工作、明智地行动必须拥有充分的信息。但是，究竟什么是“信息”呢？“信息”有什么属性？怎样度量？这些概念通常却是含糊不清的。自从信息论这一学科诞生以来，“信息”这一名词便在工程上有了比较明确的概念，成为一个基本的技术术语。

1.1.1 信息的定义

“信息”是一个常用词，在日常生活中泛指“消息”、“信号”、“情况”、“情报”、“知识”等。例如，人们从报纸或电视新闻中得到某些消息，便说得到了信息；司机根据红绿灯信号得到车辆行驶信息，等等。随着信息的开发和利用，“信息”也逐渐显示出其巨大的威力，为人类和人类社会的进步带来明显的经济效益和社会效益。因此，在日常生活和社会活动中“信息”又被赋予各种动人的比喻，如企业家认为“信息就是效益”，商人认为“信息就是利润”，军人认为“信息就是胜利”等。可见，通常的“信息”概念仅仅是从实用角度出发在某个侧面为“信息”建立的非常粗浅、非常模糊且具有很强主观性的概念，这些概念对于不同

的人在不同的时间、地点和条件下，其意义、程度有很大的差异。因此，关于信息的这些含糊不清的概念必须进行加工、概括、提升和开拓。在我们的专业学习中，信息通常是指代表着某一个抽象的有待传送、交换、存储以及提取和识别的内容，它有严格、确切的含义，有一定的数学模式并能定量地度量。

信息论的主要奠基人香农(C. E. Shannon)以通信系统为物理模型，用概率测度和数理统计的方法研究通信系统中的基本问题，并于 1948 年在《贝尔系统技术》杂志上发表了著名论文《通信的数学理论》，奠定了信息论的理论基础。在这篇论文中他将信息定义为“用来消除不确定性的东西”。

图 1－1 给出了一个一般的通信系统。在这样的通信系统中，发信者(信源)发出的信息在某种物理媒介(信道)中传输并被收信者(信宿)接收。例如在由人与人的信息交流活动所构成的信息系统中，传输的是人体各种感觉器官(如眼、耳、鼻、舌等)所能感知的光、声和字符等信号；在电信系统中，消息(符号)便为不同频率、不同相位、不同强弱的电信号。因此，我们可以将这种一般的通信系统表述为传输消息(符号)的系统。

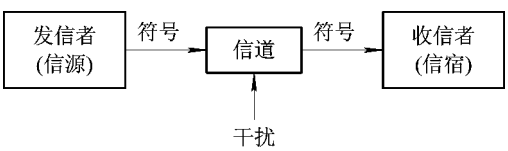

图 1－1　一般的通信系统

在图 1－1 所示的通信系统中，显然存在着一个能够使发信、收信双方都能理解的符号表(如某种文字、语音、图像、图形或电信号)。然而，通信过程是一个随机的、不确定的过程，即谁使用此通信系统、在什么时间发出何种具体符号是不能够事先预料的，而且通信过程中不可避免的干扰也是随机出现的。这些随机性因素使得这种消息或符号的传递过程有一个最基本、最普遍却又十分不引人注意的特点，就是收信者在收到符号之前不知道发信者发出的是哪一个符号，即对于收信者而言具有一定的随机性或不确定性。只有通过通信过程，收信者才能知道是谁向他发出符号，发出的是何种符号，消除关于发信者及其发出何种符号的不确定性。由此可知，信息的传递过程对于接收者而言是一个由不确定到确定的过程，而实现由不确定到确定的转变，所依赖的只能是接收者由通信过程所获得的信息。所以，信息确实是用来消除不确定性的东西，这也就是香农信息定义的含义。

对于一次实际的通信过程，由于信道中存在着随机性的干扰，因此接收者在通信完成之后可能完全消除了关于信源的不确定性，也可能只消除了部分不确定性，甚至仍然具有同样的不确定性。根据香农信息定义，接收者由这一通信过程所获得的信息量是不同的，即可能获得了信源输出的全部信息，也可能只获得部分信息甚至是没有得到信息。所以，香农信息定义一方面指明了信息与消息、信号、情况、情报的差异，即消息、信号、情况、情报只是携带着信息，是信息的载体，并不是信息本身，信息则是这个载体所携带的内容。另一方面，香农信息定义指明了信息是可以度量的，即消息、信号、情况、情报所携带的信息量的大小与接收者通过通信过程消除或减少对于该事物的不确定程度有关。只要原有的不确定性程度有所减少，接收者便由此次通信得到了一定的信息。

香农信息定义揭示了信息的含义，为信息、通信系统的定量分析打下了良好的基础。

但是，该定义只是从信息的功能上对信息加以描述，只是说出了信息能够做什么，有什么用，并没有正面回答信息是什么。

关于信息的定义，学术界有多种描述，但是尚无一种适用范围广、得到普遍公认的定义形式。这说明人们对于信息的研究还没有达到完全揭示本质的程度。但是人们在这一领域所从事的研究、所提出的各种定义仍然在一定程度或从一定的侧面触及到了信息的某些实质问题。例如，控制论的创始人维纳(N. Wiener)1948 年在《控制论——动物和机器中的通信与控制问题》一书中指出："信息就是信息，不是物质，也不是能量"。1950 年，他又在《人有人的用处》中提出："信息就是人和外界互相作用的过程中相互交换的内容和名称"。维纳的定义表明，人与客观世界除了物质、能量的交换之外，还有信息的交换。人类通过语言、文字、图像等各种手段交换信息，学习有关彼此的知识，使人类社会成为一个有机的整体。由此可知，信息及其信息的交换确实也是一种客观存在。同时也应当注意到，信息的交换不仅发生在人与外界相互作用的过程中，在没有人参与的许多相互作用的过程中同样包含信息的交换。

从更一般的意义上可以认为，信息反映了一切事物的存在和运动方式的描述。此时信息的定义表述为：事物运动(包括客观世界和主观世界以及人脑的思维活动)所表现出来的各种状态和方式。

在这种定义下，维纳所说的"人与外界……相互交换的内容"不是别的，正是"事物运动的状态和方式"，即人从外界得到的信息是外界事物的运动状态和方式，而人给外界的信息则是人的思维运动的状态和方式，是人的意志和命令。对于香农定义中的所谓不确定性，则正是对于事物运动的状态和方式所具有的不确定性，即不知道事物处于什么运动状态中，不知道事物正在以什么方式运动。要消除这种不确定性，就需要知道事物运动的状态和方式，需要得到信息。因此，信息——事物运动的状态和方式，就是用来消除不确定性的东西。可见，这一定义不仅统一了多种定义的概念，而且从更一般的意义上较好地揭示了信息的本质。

近十几年来，以计算机、信息处理、网络、通信和微电子技术的发展、应用和普及为标志的信息技术得到了迅猛发展。人类已经进入信息时代，我们对于信息及其在人类的生活和社会活动中的作用有了更加明确、深刻的认识，"信息"已经成为人类在"数字化生存"中的一项基本要素，决定了人在现代社会中的生存方式和生存质量。

由上面的分析我们可以归纳出以下三点：

(1) 信息不是物质或能量，它是自然世界和人类社会中的另一种基本要素；

(2) 信息具有客观存在的属性，是客观事物运动状态和方式的表征，不完全取决于人的主观认识；

(3) 信息又与主观认识有关，是客观事物在人脑中的反映。

在信息处理学科中，信息常表示为对数据含义的一种解释，而数据则是那些可以记录下来、反映事物运动状态和方式的符号，包括数值、曲线、图像、语音等。在信息的处理过程中，输入的是数据，处理后的输出结果仍然是数据，而信息则隐含于这些数据中，针对某种特定的目的，可由数据解释其含义。例如卫星获得的地物遥感影像数据，经数字图像处理、识别与分析后，输出仍是数据。对于同一幅遥感影像数据，通过不同方法处理可以得到不同的地物目标分类，农业、矿业、军事等不同的应用部门由输出数据或影像可解译

出不同的结果，获取不同的信息，如农作物面积、矿藏分布或军事设施。在我们的专业学习中，通常遇到的是这种狭义的信息概念。

1.1.2 信息的性质

信息作为客观世界的第三要素，与物质、能量相比具有一些特殊的性质。

(1) 信息是无形的。

信息不同于物质和能量，它是客观世界中的一种看不见、摸不着、没有重量的基本要素。新闻、报纸、信号、密码等只不过是信息的载体。

(2) 信息是可以共享的。

在客观世界中，物质和能量所遵循的基本规律之一是守恒定律，即物质和能量不能创造也不能消灭，只能转换。然而，信息不受守恒定律的制约，没有重量，易于复制，能以极快的速度传播，是一种可以共享的重要的社会资源。信息的交流不但不会使信息的持有者失去原有信息，而且可以获得新的信息。

在现代社会中，飞速发展的通信手段和信息处理技术为信息资源的充分共享提供了优越的条件，同时国家之间的军事对抗、商业活动中的市场争夺、多媒体数字产品的版权归属等，也对信息的安全和保密提出了更高的要求。

(3) 信息是无限的。

客观世界的运动是永恒的。作为描述客观事物的运动状态和运动方式的信息，与客观事物及它们的运动一样，也是永恒的。

在组成客观世界的三大要素中，人类可以使用的物质和能量资源只能供有限的人使用，总有一天会出现短缺危机，而信息则永远在产生、更新和演变，可以多人共享使用，并且使用的人越多，其价值越高。信息已经成为人类社会中的一个取之不尽、用之不竭的知识源泉。

信息的这一无限性在时空上表现为可扩展性。例如，失去了现实应用时效的历史资料可以反映社会发展变化的统计规律，先进地区的科学技术可以带动落后地区的经济发展。

(4) 信息是可开发的。

信息的开发是指信息的表示、存储、传输、处理和利用。显然，信息的开发和利用历来是客观世界发展、人类文明史进化的一个基本条件。在信息社会的今天，信息技术产业已成为推动社会进步，拉动经济、科技、教育等各个领域发展的重要部分。

(5) 信息是可度量的。

信息反映事物的运动状态和方式。通过研究客观世界所遵循的自然规律，人类将不断地认识和掌握事物变化、发展的客观规律，并且找出适当的方法描述事物的运动状态和方式，于是反映客观事物的运动状态和方式的信息也可以按某种方式加以度量。

例如，在香农的信息定义中，信息量与事件发生的不确定程度及信源发出符号的随机性有关，信息是关于事件随机性的一种描述。因此，我们可以用概率统计的方法来度量信息。

设发信者发出的符号 a_i 出现的概率为 $P(a_i)$。$P(a_i)$愈小，则 a_i 的随机性即不确定性愈大，a_i 含有的信息量也就愈大。由此可知，a_i 具有的信息量应当与 a_i 发生的概率成反比，即

$$I(a_i) = \log \frac{1}{P(a_i)} = -\log P(a_i) \tag{1.1}$$

在通信的过程中，信道中不可避免地存在着随机性的干扰。于是，在接收端收到符号 b_j 时，发信者发出符号为 a_i 的可能性可以用条件概率 $P(a_i|b_j)$ 来表示，那么在接收到符号 b_j 后对发信者发出的符号是否是 a_i 仍然存在的不确定性便成为

$$\log \frac{1}{P(a_i \mid b_j)} = -\log P(a_i \mid b_j)$$

于是，通过信息的传递，接收者对于事件 a_i 的不确定性改变了，表明通过这样的信息传输系统我们由 b_j 得到了关于 a_i 的信息。显然，由这一通信过程所获得的信息量的大小应当与关于事件 a_i 的不确定性的改变量有关。于是，通过信息传输，接收者由 b_j 得到的关于 a_i 的信息量可以表示为

$$I(a_i; b_j) = \log \frac{1}{P(a_i)} - \log \frac{1}{P(a_i \mid b_j)} = \log \frac{P(a_i \mid b_j)}{P(a_i)}$$

可见，香农给出的信息度量确实是一种对通信过程中所消除的不确定性的度量。

1.2 信息论的主要研究内容

由上面的讨论我们知道，信息论的研究对象是传输消息(符号)的系统。由于这些消息载荷着信息，因此这种消息传输系统即为信息传输系统，通常简称为通信系统。信息论的创始人香农以狭义的通信系统为物理模型，研究信息的定量描述方法和编码理论，分析通信系统中的有效性和可靠性问题，奠定了信息论的理论体系。随着科学技术的发展，信息论的研究范围、应用领域和指导意义已远远超出了原来的范围，成为研究广义信息系统中一般规律的工程学科，它的主要目的是提高信息系统的可靠性和有效性，实现系统的最优化。

客观世界中存在着各种各样的信息传输、存储和处理系统，如电报、电话、雷达、遥感、计算机系统等工程意义上的各类狭义的通信系统，以及人类社会的管理系统、生物有机体的神经系统、生物遗传系统等。这些系统的形式、用途各不相同，但从信息传输的角度分析，在本质上它们有许多共同之处，即在这些系统中都有信息的发送者和接收者，都存在着信息的传递、交换和处理。对于这些互不相同的信息系统，我们可抽象为一个广义的通信系统来描述，并概括为图 1-2 所示的模型。

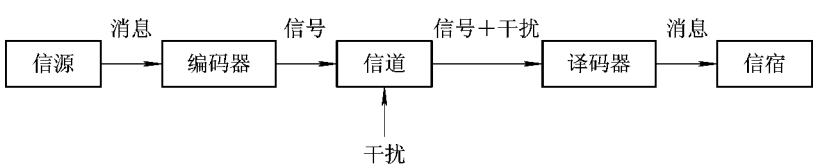

图 1-2 通信系统模型

广义的通信系统模型主要由信源、编码器、信道、译码器和信宿五个部分组成。利用这样的模型我们可以分析、探讨通信系统在传输消息的过程中的一般规律。

(1) 信源：发出信息的客观事物，可以是人，也可以是某种设备或物体。

由于消息是信息的载体，因此信源的输出是消息或消息序列，如电信号、文字、图像、

语音等。消息或消息序列可以是连续的，也可以是离散的。由前面的讨论我们已经知道，信息的基本属性之一是随机性，因此信源输出的消息、消息序列需要用随机过程加以描述。

信源研究的主要问题是其统计模型(统计特性描述)及信源所具有的信息量的定量表示。

(2) 信道：传输信号的物理媒介或通道。

在狭义的通信系统中，实际信道有明线、电缆、波导、光纤、无线电传播空间等。广义的通信系统的信道多种多样。除了传输信号，信道还具有存储信号的作用，如书信在邮递过程中不仅执行信息的传递任务，同时也起到了信号的存储作用。再如，计算机中的硬盘、移动存储设备不仅可以保存数据或文件，存储的数据、文件也可以与其他计算机进行交流，实现信息的传递和共享。

信道研究的主要问题是它能传输、存储多少信息，即它的信道容量有多大。

(3) 编码器：将信源输出的消息或消息序列转换成适合通信系统要求的信号的设备。

不同的通信系统中允许的消息或消息序列的表示形式可能不同。例如，面对面的交谈需将要表达的内容通过发声器官转换成听觉器官能够接收的声波，并在允许声波传播的空间中传输；使用邮政通信需将要传递的消息转换成文字并构成规范化的书信；电报通信则需将文字转换成由点、画构成的莫尔斯码；语音、图像、视频等数字媒体文件需使用二进制符号表示并按照一定的格式进行处理与存储。因此，对于任何通信系统而言，信源输出的信息都必须通过编码实现从消息到信号、数据的转换，才能够通过某种信道进行传输、存储和处理。由消息到信号、数据的转换一般要经过信源编码器和信道编码器。对于狭义通信系统，编码器还包括调制等各种处理。

编码器研究的主要问题是如何进行信源编码，使信源的消息被充分利用并可靠地通信。

(4) 译码器：编码器的逆过程，它将信号转变为消息。

与编码器相对应，译码器分为信道译码器和信源译码器。对于狭义的通信系统，还包括信号的接收与解调。

(5) 信宿：消息的接收者，即接收消息的人或设备。

(6) 噪声源：信号在信道传播的过程中常常会受到噪声的干扰。噪声可分为系统内噪声和系统外噪声，如雷电、宇宙辐射、设备自身的电子噪声和人为电子干扰等。在广义通信系统模型中，这些干扰和噪声都等效地折合为信道干扰，可看做是由一个噪声源产生并作用于在信道中传输的信号。在通信系统中，噪声是影响可靠通信的主要原因之一，可通过各种信道编码技术提高系统的抗干扰能力。

通过对广义通信系统模型的分析可以看出，通信的基本目的是在接收端(信宿)准确地或近似地再现从另一端(信源)选择出来的消息，因而通信系统的基本问题是信源、信道及编码问题，信息论的基本任务则是为设计有效而且可靠的信息系统提供理论依据。

本教材主要涉及狭义的通信系统，即讨论香农的信息理论，在深入分析香农信息论的基本概念和编码定理的基础上，研究提高信息系统有效性的理论与工程手段，即讨论信源编码原理和数据压缩的一般方法。

1.3 信道编码的研究内容与发展简史

信道编码是20世纪40年代末提出、60年代发展起来的一门提高数据传输可靠性的理论与技术，至今已有60余年的历史。随着数字通信的发展，特别是20世纪70年代以来，卫星通信和高速数据网的飞速发展，对数据传输的可靠性提出了越来越高的要求，因此，如何提高数据传输的可靠性一直是通信发展过程中研究的重点之一。

自从1948年Shannon发表了《A Mathematical Theory of Communication》这篇论文后，就开创了两个现代研究领域——信息论和编码理论。他最杰出的贡献就是引入了概率论来分析和研究通信系统。在信道编码领域，他给出了“信息”的一个非常有用的定义，以及几个“信道编码定理”，描述了给定通信系统可靠传输的精确上界(称为信道容量)。

在这篇论文中，最令人感兴趣的结果是“noisy channel coding theorem for continuous channels with average power limitations”，这个定理采用带限加性高斯白噪声信道模型，这个信道模型可以近似描述许多实际的数字通信和存储系统。定理的证明表明，对于任何小于等于信道容量的传输速率，存在一个编码方案，可以达到任意小的错误概率；相反地，如果传输速率大于信道容量，则不存在任何编码方案可以达到可靠的性能。然而，这是一个存在性定理，对于如何寻找合适的编码方案，实现起来多么复杂，该定理并没有给出任何指导。

从1950年Hamming的工作开始，许多通信工程师和编码理论家已经开发了很多种不同的复杂度合理的方法，试图接近Shannon所预期的性能。在这里，我们将沿着过去60余年编码在通信和存储系统中的应用历程来看一下在缩小实际系统和信道容量之间的距离上已经取得的巨大进步。

在Shannon定义了可靠通信的理论极限的同时，Hamming和Golay正在着手开发第一个可实现的差错控制方案，他们的工作开辟了一个欣欣向荣的应用数学的分支——编码理论。通常认为Richard Hamming是发现第一个差错控制码的人。在1946年，Hamming在Bell实验室工作，从事弹性理论的研究，然而，他最终在计算机上花费了大量的精力。那时候，计算机的运行非常不稳定，虽然具备了检错能力，但是当发现错误的时候，程序只能中断执行。那时Hamming的程序很少不中断地从头到尾执行，他尝试了很多方法对输入进行编码，最终使得计算机能够纠正单个错误进而能连续运行。他的编码方法是在每4比特信息位后面插入3比特校验位，校验位是由信息比特线性组合得到的，被后人称做Hamming码(汉明码)。

虽然Hamming码取得了很大的进步，但也存在很多缺点：一是传输效率不高，二是在每个数据块中只能纠正一个错误。Marcel Golay指出了这些问题并将Hamming的工作进行了推广，这就是著名的Golay码。Golay码将数据中每12个比特组成一个数据块，然后插入11个比特的校验位，组成长度为23比特的码字，可以纠正3个比特错误。这类码字的一个共同特点是在每k个信息比特之后插入r个校验比特组成n个比特的码字，这就是众所周知的分组码。在引入信道编码的这60余年当中，已经找到了很多性能优良的分组码，并得到了大量的应用，比如Golay码就曾被应用到木星探测中。当然，现在它早已经被很多性能更好的编码方式代替了。

1957 年 Prange 提出了循环码的概念。这种码字可以用多项式来描述，有很严谨的数学结构。它有一个很大的优点是每个码字循环移位后仍然是一个码字，可以利用这种循环移位特性来降低编码器和译码器的复杂度，尤其是译码，采用一种 Meggitt 译码器的结构可以非常有效地对纠错能力较小的循环码进行译码，但是译码复杂度随着纠错能力的提高指数增加。

循环码的一个很重要的子类是 1959 年 Hocquenghem 以及 1960 年 Bose 和 Ray-Chaudhuri 组成的研究小组发现的，这就是著名的 BCH 码。Reed 和 Solomon 还在多进制上构造出 BCH 码，后人称之为 RS 码。RS 码的出现是编码历史上一个很重要的进步，因为它可以纠正突发错误。但是直到 1967 年 Berlekamp 才提出了一个有效的译码算法，RS 码才逐步得到广泛应用，如 CD、DVD。

虽然分组码得到了成功的应用，但是在使用过程中它也有几个缺点：第一，由于分组码是分帧的，因此在译码之前必须要接收到整个码字才可以进行，这会对系统引入一个不可接收的延迟，尤其是对于码长较长的分组码；第二，分组码需要精确的帧同步；第三，大部分基于代数的译码器输入的是硬判决比特，而不是解调器输出的没有量化的“软”值。如果分组码使用硬判决比特，则信道被假定为二进制信道；如果是软判决比特，则信道的输出是连续取值的。然而，为了达到 Shannon 预测的性能限，需要采用连续取值的信道输出。因此，虽然在相对较好的信道上分组码可以取得卓越的性能，但是当信噪比较低时，它具有相当差的性能。

注意，在低信噪比的情况下，分组码较差的性能并不是码本身的特性，而是由于硬判决译码具有次最优特性，因此分组码也可以用软判决来译码。但在不久以前，人们还认为软判决译码太复杂，目前这种想法正在发生改变，主要是因为最近几年的一些研究工作，如 RS 码的纠错纠删译码算法和 1998 年 S. Lin 等人提出的基于网格软判决译码算法。

1955 年，Elias 提出了卷积码的概念。卷积码可以克服分组码的缺点，卷积编码器采用移位寄存器对连续的比特流增加冗余。第一个有价值的译码算法是 1961 年 Wozencraft 和 Reiffen 提出的序列译码算法。之后，1963 年 Fano 和 1969 年 Jelinek 分别对其做了改进。虽然 Fano 和 Jelinek 的工作对译码算法进行了改进，但是直到 1967 年引入了 Viterbi 算法以后，最佳的译码方案(在最大似然的意义上)才可实际操作。

卷积码最致命的弱点是对突发错误非常敏感。

卷积码抗突发错误能力较差的问题可以通过交织器来减轻。通过打乱发送端的比特的顺序，并在接收端恢复比特的顺序，突发错误可以被展开，对译码器来说，就可以看成随机错误。

这种串行级联的形式是 David Forney 于 1966 年首先提出来的，并用于 NASA 和 ESA 在 1987 年制定的深空网络标准中。在这个标准中($q=8$，$n=255$，$k=223$，$t=16$)RS 码同 Odenwalder 卷积码级联使用。

1993 年 7 月在 Geneva Switzerland 召开的 International Conference on Communications (ICC)会议上，实际编码系统与 Shannon 理论极限之间的距离更加接近了。在这次会议上，有两篇论文讨论了一种新的编码方式和相关的译码技术，作者 Berrou、Glavieux 和 Thitimajshima 造了个新词“Turbo Codes”来描述这类新的码字，他们的这个发现产生了很大的影响，并获得了很高赞誉。

低密度奇偶校验码(Low Density Parity Check Codes，LDPC 码)是一种可以用非常稀疏的校验矩阵来定义的线性分组纠错码。Gallager 于 1960 年在其博士论文中首次提出 LDPC 码，并证明了它的距离特性。LDPC 码具有性能接近香农极限，误码率极小，抗干扰能力强等特点。但是 LDPC 码的运算复杂度较大，在接下来的时间里就逐渐被人们淡忘了。

D. J. C. MacKay、M. Neal 和 N. Wiberg 等人对 Gallager 早在 20 世纪 60 年代初就提出的 LDPC 码重新进行了研究，发现 LDPC 码同样具有逼近香农极限的性能，而且复杂度更低。于是越来越多的学者开始关注并研究 LDPC 码。

LDPC 码的优异性能及其在信息可靠传输中的良好应用前景(如第 4 代移动通信系统、高速与甚高速数字用户线、磁记录系统等)，已引起各国学术界和 IT 界的广泛关注。LDPC 码将取代 Turbo 码的趋势已很明显，研究 LDPC 码将具有很大的学术价值及工程应用与商业价值。

1.4 信源编码的研究内容与发展简史

将信源输出的消息或消息序列转换成电子信息系统和通信系统要求的码字的过程称为信源编码。为了提高系统的有效性，信源编码的主要目的是选出一组适当的码字，使表示信源输出消息和消息序列的码率最低。这种以压缩数码率、提高系统有效性为目的的技术手段亦称做数据压缩。

1.4.1 数据压缩的必要性

随着大规模集成电路(微电子技术)与计算机技术的飞速发展和广泛应用，将各类模拟的连续信号转换为数字信号并实现数字信号的存储、传输和处理日益显示出其巨大的优越性。将模拟信号转换成为数字信号并构成数字信息系统，其突出优点是抗干扰能力强，宜于再生与存储，宜于误码保护与加密，宜于多路复用、分组与组合，并可应用现代计算技术与手段进行信息的提取、滤波、分类、识别等信息处理。在当今的信息时代，各种信源信息的存储与传输业务与日俱增。文件和图形的传真、电视电话、会议电视、数字电视、高清电视、视听图文数据库系统等静止图像、语音数据存储与传输的研究和应用已有突飞猛进的发展。以图像、图形、语音信息的存储、传输为代表的数字信息系统的发展和应用必将给人们的日常生活乃至人类社会的进步带来巨大的变革，在政治、军事、经济、教育等各个领域中发挥重要作用。

然而，数字的信息系统也有其明显的缺点，即模拟信号转换成数字信号后，数据量大，占用频带宽。例如，数字电话系统的通频带宽通常为模拟电话带宽的 8 倍，常规电视图像数据的传输速率达 165 Mb/s。各类数字信息系统的应用(特别是图像信息的通信与存储)将使其数据量大的缺点更为突出。虽然卫星通信、光纤通信等现代通信手段使得通信频带成倍拓宽，激光存储设备为海量数据存储提供了更加充分的硬件环境，但是，人类社会对于各种媒体信息的存储、传输和利用的需求仍然超过、甚至远远超过存储设备的容量和通信系统的传输能力。因此，通过信源编码在无失真或满足某种失真要求的条件下，降低数码率的数据压缩处理已成为现代通信系统和数字信息系统的关键技术课题之一。

1.4.2 数据压缩的可能性和基本方法

作为信息传输系统中一般规律的基础理论学科，香农信息论在提高系统有效性方面进行了深入的理论分析，除了应用数理统计方法对各类信源进行描述、分类及定量地计算输出信息量之外，针对信源输出中含有不包含信息的多余成分的现象，从信源符号的概率分布和信源符号序列中各分量的相关性出发，研究信源输出信息的能力并定义了信源的信息冗余度，深入分析了信息系统中信息传输率和失真的关系并构成了信息率失真理论。在这些理论的基础上，香农第一编码定理和香农第三编码定理分别指出了无失真信源编码和限失真信源编码的最优码的存在性以及最小信息传输率的理论极限。香农信息论中的这些讨论为信源编码原理和应用技术的研究奠定了重要的理论基础。

依据香农信息论中的这些理论，去除信源输出中的冗余数据即可实现无失真的数据压缩。若进一步根据人的生理特性和心理特性，在信息系统中引入一定限度内的失真，即丢弃包含信息量少的次要成分，保留信息量丰富的主要成分，则可实现限失真条件下的高效信源数据压缩，大大提高信息系统的有效性。

由此可知，信源编码原理包括无失真信源编码和限失真信源编码两种类型，相应地数据压缩应用技术也可以分为无失真数据压缩(如传真通信系统)和限失真数据压缩(如图像、语音系统)。对于多数信息系统，由于人的心理特性和生理特性的影响，一定限度内的失真对于接收者而言或者并无觉察，或者认为是可以接受的，而这种少量的失真却可以换取明显的数据压缩效果。因此数据压缩研究的重点是在信息率失真理论指导下的限失真编码方法。

1.4.3 信源编码研究的发展简史与现状

从工程意义上讲，通过信源编码进行数据压缩并不是新的课题。自从电信系统出现后，以提高系统有效性为目的的编码技术就开始了研究与应用。例如，作为早期数据压缩技术应用的成功范例，1832 年电报系统中的莫尔斯(Morse)码提高了电报通信的速度。20 世纪 20 年代，将图像编码用于伦敦和纽约之间的海底电缆传输系统，使横跨大西洋传递一幅新闻图片所需的时间由一个多星期减少到小于三个小时。1939 年达德利(H. Dudley)发明声码器，使数字语音能在频带非常窄的电话信道内传输。这些早期数据压缩技术的研究和应用不仅提高了通信系统的有效性，而且为信息理论的诞生与发展提供了实践依据和理论研究基础。

自 20 世纪 40 年代末开始，以香农信息论为理论指导、以计算机为主要处理工具的信源编码理论与技术得到了系统的研究和广泛的应用。20 世纪 60 年代到 70 年代，预测编码和正交变换编码理论迅速形成并快速发展，在数字语音和数字图像信号的频带压缩传输方面做了大量模拟试验。1969 年，Milliard 等人首先研制带宽约 1 MHz 的电视电话 DPCM 编译码机，并进行了数百公里的传输实验(6.3012 Mb/s)，同年 Mounts 进行了电视电话帧间编码的计算机模拟。20 世纪 70 年代末，利用动态估值进行运动补偿的活动图像编码研究非常活跃，随后方块截尾编码和矢量量化等方法取得了一定发展。在这一阶段的应用技术开发除了语音和静止图像外，还包括电视电话、会议电视和常规电视。近 30 年来，随着大规模集成电路和现代通信技术、互联网技术的进步，信源编码的研究与应用又有了突飞

猛进的发展。集信源编码四五十年的研究精华，国际标准化组织制定了一系列编码标准建议。1983 年提出了 32 kb/s 长途电话质量的语音 DPCM 标准的编码系统；1988 年基本确定了电视电话/会议电视 CCITT H.261 建议和静止图像压缩的 JPEG 建议，稍后又提出了活动图像压缩的 MPEG－2、MPEG－4 标准。在专用芯片的研制方面，1983 年设计出 ADM 语言压缩器和 LPC 语言压缩器，并且运用于轻便电话和移动通信系统。目前已有多种可实现图像、视频数据压缩的专用芯片以及压缩视频和音频信号的多媒体处理器。

在经典编码方法成功应用于各编码系统的国际标准的同时，新的方法、新的理论也不断涌现。子带编码、塔形编码及小波变换编码利用一维或二维数字滤波进行分解与合成，实现了语音和图像的多分辨率编码。作为现代编码技术的模型，分形法和神经网络编码成为图像编码的研究热点，它们的显著特点是突破了经典信源编码的理论框架，在压缩比、图像、语音的主观质量上带来了显著的改进。

信源编码研究的历史和现状表明，高效、高保真数据压缩的信源编码是当代信息科学和工程领域中十分活跃的一个分支，在这一领域中有大量的理论课题和实用装备需要研究和开发。信源编码这一“不十分起眼”的“小学科”对当今“信息化社会”的进步将发挥重要的作用，做出其应有的贡献。

第 2 章　离散无记忆信源与信息熵

信息理论的研究对象是以各类信息的获取、表示、传输和处理为目的的信息系统。图 2－1 给出了一个典型的通信系统物理模型。在这样的通信系统中，一个贯穿始终的、最基本的问题便是信息，即信源输出的是信息，在系统中传输的是信息，接收者获得的也是信息。可见，在信息理论的学习和研究中，首先需要对信息给出一个明确的、可以度量的工程概念。

本章我们从离散无记忆信源的统计特性入手，给出信息理论中的基本概念——信息熵，讨论信源信息熵的基本性质。

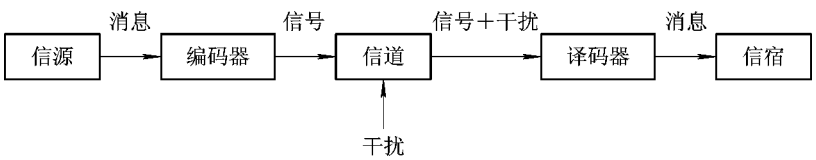

图 2－1　通信系统物理模型

2.1　离散无记忆信源

1．信源的概念

信源是信息的来源，以输出符号(消息)的形式输出信息(见图 2－2)。由于我们关心的是信源输出的信息，因此信源研究的主要对象是载荷信息的消息，而对信源内部的结构和输出消息的机理一般不需要作深入了解。此处我们从最简单的信源入手，讨论信源的描述和主要类型，建立离散无记忆信源的数学模型。

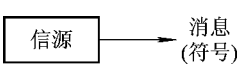

图 2－2　信源模型

由于信源输出的消息载荷着信息，因此这种消息所具有的一个基本的属性便是随机性。正如绪论中所述，通信过程中收信者在收到消息之前，对信源发出的消息是不确定的。例如，掷一个均匀的骰子。虽然我们知道骰子上的数字只能有 1、2、3、4、5、6，即投掷骰子所得到的消息的集合中只有这六种消息，但是骰子稳定之前我们并不知道骰子将呈现出哪一种消息。因此，对于信源输出的这种随机出现的消息，需要用随机变量加以表示。如

果我们已知信源的消息集合(即样本空间或值域)和消息发生的概率分布，那么便可以使用由样本空间和它的概率分布所构成的概率空间来描述信源。

设信源输出的消息 X 为随机变量，其样本空间(值域)为 R。若在此值域上随机变量 X 的概率分布为 P，则此信源可以表示为$[R, P]$或$[X, P]$，并将这种表示称为信源的概率空间。

因此，信源可以用概率空间作为其数学模型。

2. 离散无记忆信源的概念

在本课程的教学中我们主要涉及数字信息系统中的信息传递。在这种数字信息系统中，信源输出消息是一个离散随机变量或由离散随机变量构成的随机变量序列。本教材中我们主要以离散信源作为研究对象，讨论离散信源的信息概念与度量，以及离散信息系统中信息传递的特点与定量关系。这种考虑与目前以计算机和信息处理技术为手段的电子信息系统和通信系统的应用是相一致的。

为了便于学习和理解信息理论中的基本概念和基本关系，我们首先从最简单的离散信源，即离散无记忆信源入手展开讨论。

离散无记忆信源是最简单的离散信源，其主要特点是离散和无记忆。

离散：信源可能输出的消息的种类是有限的或者是可数的。消息的样本空间 R 是一个离散集合。由于信源的每一次输出都是按照消息发生的概率输出 R 中的一种消息，因此信源输出的消息可以用离散随机变量 X 表示。

无记忆：不同的信源输出消息之间相互独立。

对于离散无记忆信源，如果已知信源输出消息的取值集合：

$$R = [a_1, a_2, \cdots, a_n]$$

和每一消息发生的概率：

$$P(a_1) = p_1, P(a_2) = p_2, \cdots, P(a_n) = p_n$$

则该离散无记忆信源的概率空间为

$$[X, P] = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ p_1 & p_2 & \cdots & p_n \end{bmatrix} \tag{2.1}$$

其中，每一消息发生的概率满足：

$$\begin{cases} p_i \geqslant 0 \qquad i = 1, 2, \cdots, n \\ \sum\limits_{i=1}^{n} p_i = 1 \end{cases} \tag{2.2}$$

在上面给出的投掷骰子输出消息的随机试验中，每次试验的结果必然是 1、2、3、4、5、6 六种不同消息中的一个，并且每一次试验中出现哪一种消息是随机的，但是必定出现其中某一种消息，并且两次不同投掷所出现的消息之间没有关联。如果这颗骰子是均匀的，则由大量的试验可以知道，这个信源的六种消息发生的概率均等，即分别为 1/6。因此，这样的信源是一个离散无记忆信源，其概率空间为

$$[X, P] = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ \dfrac{1}{6} & \dfrac{1}{6} & \cdots & \dfrac{1}{6} \end{bmatrix}$$

2.2 自信息和熵

在狭义的通信系统的讨论中，我们把信源看做一个输出消息和消息序列的设备。2.1节我们根据信源输出消息所具有的随机性关系，从信源的外部特性给出了信源的描述方法，即信源的概率空间。然而，信源的基本功能和作用是输出信息，从本质上讲，信源是一个产生并输出信息的设备，消息仅仅是信源所输出的信息的载体。因此，对于信源我们更感兴趣的是它输出的消息中含有多少信息，它输出信息的能力有多大。因此，只有透过信源的外部表现——消息及消息的概率空间，研究信源的信息特性，我们才能掌握信源的本质，并在信息的输出、传输和接收过程中对信息给出定量的理论分析。

下面我们将从信息的基本属性出发，分析和度量信源的信息特性。

首先，我们以离散无记忆信源为对象，引出自信息和熵的定义与度量，并对信息理论中最重要的基本概念——熵的物理含义和数学性质进行深入分析。

2.2.1 自信息

观察一个由 n 种符号构成的离散无记忆信源：

$$[X, P] = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ p_1 & p_2 & \cdots & p_n \end{bmatrix}$$

由于信源按照符号发生的概率随机地输出消息，因此在信源输出消息之前，我们不能够确定信源将要输出的是哪一个消息，即收信者对信源的输出具有不确定性。只有当信源输出某一消息且这个消息在无干扰通信系统中传输并被收信者收到后，收信者才能消除这种不确定性。由香农关于信息的定义可知，这种能够消除不确定性的东西就是信息。由此还可进一步看出，如果某一消息发生的不确定性大，则一旦它发生并被接收到，收信者消除的不确定性就大，从该消息获得的信息量也就多。因此，信源输出消息所载荷的信息量的大小与该消息发生的不确定性的大小相联系。由于信源输出的消息 a_i 依据概率 $P(a_i)$ 发生，因此 a_i 是否发生的不确定性便与其发生的概率 $P(a_i)$ 有关。

可以看出，消息 a_i 发生所含有的信息量 $I(a_i)$ 应当是 a_i 发生的先验概率 $P(a_i)=p_i$ 的函数，即

$$I(a_i) = f[P(a_i)] = f[p_i] \tag{2.3}$$

由于这种函数关系反映的是信源输出某种消息时该消息所载荷的信息量，因此称 $I(a_i)$ 为信源输出消息 a_i 时，消息 a_i 所载荷的自信息。

显然，函数 $f[P(a_i)]$ 应当满足下面四个条件。

(1) $f[P(a_i)]$ 是 $P(a_i)$ 的单调函数。

消息发生的信息量是消息发生不确定性的度量。如果消息发生的概率小，则我们对该消息是否会发生的不确定性就大(难以猜测)，该消息一旦发生，它所含有的信息量也就大；反之，若消息发生的概率大，则猜测它发生的可能性就大而不确定性小。因此，作为度量消息不确定性的函数，$I(a_i)$ 应当是 $P(a_i)$ 的单调减函数。

这一要求与日常生活中的情况也是相符合的。例如新闻发布两种消息，A：某地区下了暴雨；B：某地区下了百年不遇的特大暴雨。显然，人们对 A、B 两种消息的关注或震惊

程度有很大差异。由于消息 B 发生的可能性很小，因此它的发生使人们得到的信息量要比消息 A 的信息量大得多。

(2) 当 $P(a_i)=1$ 时，$f[P(a_i)]=0$。

对于发生概率为 1 的必然事件，在其发生前已可确知，不存在不确定性。因此，发生概率为 1 的必然事件所含有的信息量也就为 0。

(3) 当 $P(a_i)\to 0$ 时，$f[P(a_i)]\to\infty$。

对于几乎不可能发生的事件，一旦它意外地发生了，就会带来极大的信息量。

(4) 信息的度量应当具有可加性。

在收到信源发出的数个相互独立的消息时，收信者所获得的信息量应为各消息所含有的信息量之和。

依据信息度量的函数关系和应当满足的四个条件，我们可以很容易地确定这一函数关系能够取对数形式。

下面我们给出信源消息所含自信息的定义。

设离散信源 X 为

$$[X,\ P]=\begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ p_1 & p_2 & \cdots & p_n \end{bmatrix}$$

如果消息 a_i 已发生，则该消息发生所含有的自信息定义为

$$I(a_i)=\log\frac{1}{P(a_i)}=\log\frac{1}{p_i} \tag{2.4}$$

可以很容易地证明，自信息的定义满足上面提出的四个要求。

说明：

(1) 此自信息的定义是根据消息发生的概率建立的一个工程定义，而不是根据这个消息对人的实际意义而建立的定义。这一纯粹技术性的定义仅仅抓住了“信息”一词在通常概念中所包含的丰富内容的一小部分。

(2) 自信息 $I(a_i)$ 有两种含义：

在消息 a_i 发生之前，自信息 $I(a_i)$ 表示 a_i 发生的不确定性；

在消息 a_i 发生以后，自信息 $I(a_i)$ 表示 a_i 所含有的(或提供的)信息量。

(3) 在式(2.4)中关于对数的底未作明确规定。这是因为对数的底仅仅影响到度量的单位，实际中可根据具体情况和习惯来确定。

如果取对数的底为 2，则所得信息量的单位为比特(bit，binary unit)，此时 $\log x$ 用 $\mathrm{lb}x$ 表示。

如果取对数的底为 e，则所得信息量的单位为奈特(nat，nature unit)，此时 $\log_e x$ 用 $\ln x$ 表示。

如果取对数的底为 10，则所得信息量的单位为哈特(Hart，Hartley)，此时 $\log_{10} x$ 用 $\lg x$ 表示。

本书一般采用底为 2 的对数表示，并简写为 log，此类情况全书不再说明(仅在强调或部分定理证明和公式推理中有特别注明时采用其他对数底)。

根据对数的换底公式：

$$\log_a X=\frac{\log_b X}{\log_b a}$$

可以得到不同底数时信息量的换算关系：

$$\begin{cases} 1 \text{ 奈特} = 1.443 \text{ 比特} \\ 1 \text{ 哈特} = 3.322 \text{ 比特} \end{cases} \tag{2.5}$$

(4) 在上面的讨论中，消息 a_i 是随机变量 X 的一个样值，因而消息 a_i 的自信息 $I(a_i)=I(X=a_i)$是随机变量 X 的函数，因此 $I(a_i)$也是一个随机变量。

$I(a_i)$是一个随机变量并不难理解。因为 a_i 发生可以使收信者获得大小为 $I(a_i)$的自信息，然而在信源未发出消息之前，收信者不仅对 a_i 是否发生具有不确定性，而且对于能够获得多少自信息也是不确定的。因此，伴随着 $X=a_i$ 的随机发生而发生的自信息 $I(a_i)$是一个随机变量，并且与随机变量 X 具有相同的概率分布，即自信息 $I(a_i)$是一个发生概率为 $P(X=a_i)$的随机变量。

【例 2.1】 有一个输出两种消息的离散无记忆信源，其概率空间为

$$[X, P] = \begin{bmatrix} a_1 & a_2 \\ 0.99 & 0.01 \end{bmatrix}$$

如果在信源输出消息之前我们猜测 a_1 或 a_2 发生，显然这两种猜测的困难程度不同。由于 a_1 发生的概率接近于 1，即 a_1 发生的可能性很大，因此我们对 a_1 是否发生的不确定性将比较小。同理，因为 a_2 发生的可能性小，所以对 a_2 是否会发生的不确定性就比较大。

由定义式(2.4)计算消息 a_1、a_2 的自信息，可以得到：

$$I(a_1) = \log \frac{1}{0.99} = -\log 0.99 = 0.014 \text{ 比特}$$

$$I(a_2) = \log \frac{1}{0.01} = -\log 0.01 = 6.644 \text{ 比特}$$

可见，自信息 $I(a_1)$、$I(a_2)$确实是关于 a_1、a_2 发生的不确定性的度量。

如果已知信源输出消息 a_1，则由“a_1 已发生”可使我们消除大小为 $I(a_1)$的不确定性，故 a_1 发生了载荷大小为 0.014 比特的信息量。同理，如果 a_2 发生，则它所包含的信息量为 6.644 比特。

通过这个例子我们可进一步明确自信息的两种含义，即在信源输出消息 a_i 之前，自信息 $I(a_i)$是关于 a_i 发生的不确定性的度量，而在信源输出消息 a_i 之后，自信息 $I(a_i)$表示 a_i 所含有的信息量。

例 2.1 中的信源只有两种不同的消息，我们称这类信源为二进制离散信源。二进制离散信源所具有的两种消息可以分别使用二进制数 0 和 1 表示。

【例 2.2】 在一个抛硬币的随机试验中，正反两面出现的概率相等。如果将正、反两面分别看做 0 和 1 两种不同的消息，则这一随机试验可表示为一个等概率输出 0 或 1 的二进制离散无记忆信源。此二进制离散无记忆信源的概率空间为

$$[X, P] = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

在这一随机实验中，可以求出消息 0、1 发生所包含的自信息为

$$I(X=0) = I(X=1) = \log \frac{1}{1/2} = \log 2 = 1 \text{ 比特}$$

由于消息 0 和 1 发生的概率相同，因此在信源输出消息之前，我们关于信源输出 0 或

1 的不确定性是相同的，而在信源输出 0 或 1 之后，它们所携带的信息量也相同。因此，消息 0 和消息 1 的自信息相同(均为 1 比特)。

应当注意：信息单位的“比特”与计算机术语中的“比特”的意义是不同的。在计算机技术中，比特表示二进制数码中的“位”，为 binary digit 的缩写，而信息理论中的比特是使用以 2 为底的对数关系进行信息度量时的信息单位。

在例 2.2 中，由于信源是等概率的二进制信源，因此在用以 2 为底的对数关系对其输出信息量进行度量时，它所输出的每一位二进制码所含有的自信息量恰为一个比特。

2.2.2　信源的信息熵

式(2.4)定义的自信息给出了信源输出的某一个消息所含有的信息量。自信息从信源输出的每一个消息所含有的信息量描述了信源的信息特性。然而，自信息是一个伴随着信源随机地输出某一消息而产生的随机变量。由于不同消息的自信息不同，因此自信息只是从局部依消息发生的概率对个别信源消息的信息特性给出的一种描述。在实际信息系统分析中，人们往往关心的并不是信源输出的个别消息，而是由整个消息集合所构成的信源的信息特性。因此我们需要给出一个确定的量，能够从信源总体上来度量信源输出信息的大小，描述信源总体输出信息的能力。

为此，我们定义信源 X 中每一消息所包含自信息的数学期望为信源 X 的平均信息量，也称为信源的信息熵(简称熵)，记做 $H(X)$，即

对于概率空间为

$$[X,\ P]=\begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ p_1 & p_2 & \cdots & p_n \end{bmatrix}$$

的离散无记忆信源 X，其平均信息量(即信源的信息熵)为

$$\begin{aligned} H(X) &= E[I(a_i)] = E\left[\log\frac{1}{p_i}\right] \\ &= \sum_{i=1}^{n} p_i \log\frac{1}{p_i} = -\sum_{i=1}^{n} p_i \log p_i \end{aligned} \tag{2.6}$$

在下面的讨论中，有时将熵的定义式写做：

$$H(X) = \sum_{X} P(x)\log\frac{1}{P(x)} \tag{2.7}$$

这时，X 也表示信源 X 的消息所构成的集合；x 为集合中的某一消息，即 X 的某一取值；$\sum\limits_{X}$ 表示对 X 的全空间进行求和运算。

说明：

(1) 在定义式(2.6)中，关于对数的底我们并未作明确规定。

应用中可根据实际情况和习惯确定对数的底，如以 2、e、10 等为底数。由于熵是信源的平均信息量，因此相应于底 2、e 或 10，熵的单位分别为比特/符号、奈特/符号或哈特/符号。

(2) 对于一般的信源 X，其消息集合中可能包含不可能发生的消息(即该消息发生的概率为 0)。

根据

$$\lim_{z\to 0} z\log z = \lim_{z\to 0}\frac{\log z}{\frac{1}{z}} = \lim_{z\to 0}\frac{\frac{1}{z}}{-\frac{1}{z^2}}\log e = 0$$

若消息 a_k 发生的概率为 0，在定义式(2.7)表示的熵的计算中，规定

$$P(a_k)\log\frac{1}{P(a_k)} = 0$$

【例 2.3】 某班下午的工作安排有三种情况：a 自习，b 上课，c 文体活动。假设每天下午独立地、随机地发出一个工作安排的通知，且这三种安排发生的概率分别是 1/2、3/8 和 1/8。若某同学得到了工作安排为 a、b 或 c 的一个通知，则此通知中含有多少信息量？如果每天发出一个通知，则一个通知中含有的平均信息量为多少？

解：此信源的概率空间为

$$[X,\ P] = \begin{bmatrix} a & b & c \\ \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \end{bmatrix}$$

可以求出这三种安排的通知含有的自信息分别为

$$I(a) = \log 2 = 1 \quad \text{比特}$$

$$I(b) = \log\frac{8}{3} = 1.415 \quad \text{比特}$$

$$I(c) = \log 8 = 3 \quad \text{比特}$$

一个通知中含有的平均信息量即为此信源的熵：

$$\begin{aligned} H(X) &= \frac{1}{2}I(a) + \frac{3}{8}I(b) + \frac{1}{8}I(c) \\ &= \frac{1}{2}\log 2 + \frac{3}{8}\log\frac{8}{3} + \frac{1}{8}\log 8 \approx 1.406 \quad \text{比特 / 通知} \end{aligned}$$

根据自信息的定义，自信息量的大小与消息发生的概率成反比，此处便有：

$$I(a) < I(b) < I(c)$$

然而，考察每一通知的自信息对信源信息熵的贡献：

$$\frac{1}{2}I(a) = 0.5,\ \frac{3}{8}I(b) = 0.531,\ \frac{1}{8}I(c) = 0.375$$

可以看出，虽然某消息 a_i 的自信息随着该消息出现概率的减小而增大，但是由于消息 a_i 出现的概率(频次)较小，因此对信源总体输出的平均信息量 $H(X)$ 的贡献并不是很大。由信息熵的定义可以看出，消息 a_i 对信源 X 输出平均信息量的贡献为

$$-P(a_i)\log P(a_i)$$

在这一函数关系中，由于 $P(a_i)$ 减小的速度比 $-\log P(a_i)$ 增加的速度更快，因此平均来讲它对信源总体的贡献还是减小了，其极限情况为

$$\lim_{P(a_i)\to 0}(-P(a_i)\log P(a_i)) = 0$$

由此可以使我们进一步明确：

自信息 $I(a_i)$ 是对消息 a_i 是否发生的不确定性和消息 a_i 所含有信息量的度量，而熵 $H(X)$ 是自信息的统计平均值，是对信源总体信息特性的度量。

作为从统计平均意义上表征信源总体特性的物理量，信源的信息熵具有下面两种物理含义：

(1) 信息熵 $H(X)$表示了信源输出后每个消息所提供的平均信息量。

(2) 信息熵 $H(X)$表示了信源输出前关于信源的平均不确定性。

例如有两个信源：

$$[X,\ P_1]=\begin{bmatrix} a_1 & a_2 \\ 0.99 & 0.01 \end{bmatrix} \qquad [Y,\ P_2]=\begin{bmatrix} b_1 & b_2 \\ 0.5 & 0.5 \end{bmatrix}$$

在信源 X、Y 未输出消息之前，直观分析可看出，信源 X 输出消息 a_1 的可能性是 99%，因此我们对 X 的平均不确定性较小；对信源 Y，它输出 b_1、b_2 的可能性均为 0.5，因此我们对于信源 Y 输出哪一个消息的平均不确定性较大。利用信源信息熵的定义式(2.6)可以对信源 X、Y 输出消息前的这种不确定性加以度量，则有：

$$H(X)=0.08 \quad \text{比特/符号}, \qquad H(Y)=1 \quad \text{比特/符号}$$

可见，信源的信息熵确实反映出了信源输出消息的平均不确定程度的大小。

说明：信源的信息熵 $H(X)$并不表示接收者所获得的平均信息量，接收者所获得的平均信息量一般也不一定等于信源的信息熵 $H(X)$。只有在传输信源输出消息的信道无噪，接收者正确无误地接收到信源发出的消息，消除了大小为 $H(X)$的平均不确定性时，接收者才由通信系统获得大小为 $H(X)$的平均信息量。

2.3　熵函数的性质

由上面的讨论我们知道，式(2.6)定义的熵从总体上描述了信源的平均不确定性和平均信息量。作为统计平均意义上表征信源总体特性的物理量，熵是表征信源输出随机变量 X 的一个数字特征，只与信源消息的概率分布有关。无论信源是否输出了消息，只要这些消息的发生具有一定的概率分布，则该信源的信息熵就存在，根据信源的这组概率分布便可求出信源的信息熵。

【例 2.4】 设有一个离散无记忆的二进制信源，其概率空间为

$$[X,\ P]=\begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix} \qquad 0\leqslant p\leqslant 1$$

可以知道，此信源的信息熵为

$$H(X)=-p\log p-(1-p)\log(1-p) \tag{2.8}$$

由此可以看出，$H(X)$是概率分布 p、$1-p$ 的函数，常记做 $H_2(p)$或 $H_2(p,\ 1-p)$，下标 2 表示信源是二进制的或称二元信源。

由式(2.8)可知，$H_2(p)$是 p 的函数，可以做出 $H_2(p)$的函数曲线，如图 2-3 所示。

例 2.2 为二进制信源在 0、1 等概率分布时的一个特例，即当信源符号 0、1 等概率分布时，消息的自信息为

$$I(X=0)=I(X=1)=1 \quad \text{比特}$$

由熵函数式可知，对于此等概率分布的二进制信源，信息熵：

$$H(X)=-\frac{1}{2}\log\frac{1}{2}-\frac{1}{2}\log\frac{1}{2}=1 \quad \text{比特/符号}$$

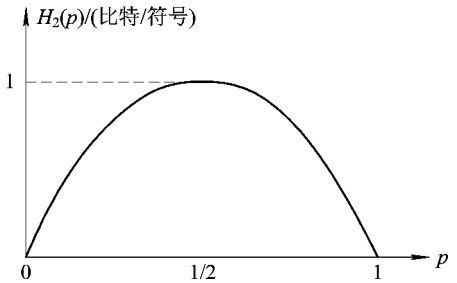

图 2-3 $H_2(p)$的函数曲线

虽然信源消息的自信息与信源的信息熵在数值上相等，但是它们的含义是不同的。自信息 $I(X)=0$、$I(X)=1$ 反映的是信源符号 0、1 含有的信息量为 1 比特，发生概率分别为 $\frac{1}{2}$的随机变量；$H(X)$则表明信源每输出一个符号平均含有 1 比特的信息量，信息熵 $H(X)$是取决于信源概率分布的一个确定量。

作为概率分布 P 的函数，由图 2-3 可以看出熵函数所具有的一些基本性质。如果二元信源的输出是确定的($p=0$ 或 $p=1$)，则该信源不能提供任何信息。当信源符号 0、1 等概率发生时，二元信源的熵达到其最大值，且为 1 比特/符号。

下面我们从信源的信息熵与概率分布的函数关系，讨论熵函数的性质。

若信源 X 的概率空间为

$$[X,\ P]=\begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ p_1 & p_2 & \cdots & p_n \end{bmatrix}$$

它的熵与信源符号的个数 n 及其概率分布 P 有关。信息熵 $H(X)$是 p_1、p_2、…、p_n 的 n 元函数 ($\sum_{i=1}^{n} p_i = 1$，其中有 $n-1$ 个独立变量)，因此也称为熵函数，记做：

$$H(\boldsymbol{P}) = H(p_1,\ p_2,\ \cdots,\ p_n) = -\sum_{i=1}^{n} p_i \log p_i \tag{2.9}$$

其中，$\boldsymbol{P}$ 为概率矢量，它的 n 个分量为 p_1、p_2、…、p_n，并且依概率分布的条件有：

$$\sum_{i=1}^{n} p_i = 1 \qquad (p_i \geqslant 0,\ i = 1,\ 2,\ \cdots,\ n)$$

信源的熵函数 $H(\boldsymbol{P})$具有下面一些基本性质。

1. 对称性

当变量 p_1、p_2、…、p_n 的位置任意互换时，熵函数的值不变，即

$$H(p_1,\ p_2,\ \cdots,\ p_n) = H(p_2,\ p_1,\ \cdots,\ p_n) = H(p_n,\ p_1,\ \cdots,\ p_{n-1}) \tag{2.10}$$

对称性进一步表明，熵是信源总体特性的描述，它只与随机变量的总体统计结构有关。由概率分布律可知，n 个符号的概率矢量由数 1 分割成 n 个正实数取值的组合所构成，而熵函数只与这 n 个实数的组合有关，与这 n 个实数和信源的 n 个符号采取何种对应关系无关。

例如，下面三个信源：

$$[X,\ P]=\begin{bmatrix}a_1 & a_2 & a_3\\ \frac{1}{2} & \frac{3}{8} & \frac{1}{8}\end{bmatrix},\ [Y,\ P]=\begin{bmatrix}a_1 & a_2 & a_3\\ \frac{1}{2} & \frac{1}{8} & \frac{3}{8}\end{bmatrix},\ [Z,\ P]=\begin{bmatrix}b_1 & b_2 & b_3\\ \frac{1}{2} & \frac{3}{8} & \frac{1}{8}\end{bmatrix}$$

我们可以用 a_1、a_2、a_3 分别表示自习、上课和文体活动三个具体消息，而 b_1、b_2、b_3 分别表示晴、雾、雨三个消息。在这三个信源中，X 与 Z 所包含的具体消息(符号)的含义不同，而 X 与 Y 两信源中某一消息的概率不同。但是由于它们的符号个数及其概率组合相同，因此它们的信息熵是相同的，即从信源的总体统计特性看它们是相同的。

由此可以看出，这种熵的定义有其局限性，未能反映出信源输出的消息本身所含有的主观意义。

2. 非负性

$$H(X)\geqslant 0 \tag{2.11}$$

根据概率分布律，随机变量 X 的所有取值的概率 p_i 均在 0 和 1 之间，当取对数的底大于 1 时，有

$$p_i\log\frac{1}{p_i}\geqslant 0 \qquad i=1,\ 2,\ \cdots,\ n$$

故有

$$H(X)=H(p_1,\ p_2,\ \cdots,\ p_n)=\sum_{i=1}^{n}p_i\log\frac{1}{p_i}\geqslant 0$$

这种非负性对于离散信源的熵是合适的。对于连续信源，这种非负性并不存在。

3. 确定性

$$H(1,\ 0)=H(1,\ 0,\ 0)=\cdots=H(1,\ 0,\ 0,\ \cdots,\ 0)=0 \tag{2.12}$$

对于信源的一组概率分布 $\boldsymbol{P}=(p_1,\ p_2,\ \cdots,\ p_n)$，若某信源符号发生的概率 $p_k=1$，则依概率分布满足的条件可知，其他信源符号发生的概率必定全部为 0。因此信源输出的随机变量成为一个确定量，即几乎处处输出发生概率为 1 的那个符号，对于这种信源的不确定性为 0。由于熵是不确定性的度量，因此确知的信源的熵为 0。

4. 扩展性

$$\lim_{\varepsilon\to 0}H_{n+1}(p_1,\ p_2,\ \cdots,\ p_n-\varepsilon,\ \varepsilon)=H_n(p_1,\ p_2,\ \cdots,\ p_n) \tag{2.13}$$

此性质的证明也较简明。根据熵的定义式(2.6)，写出式(2.13)左端的熵函数，应用极限关系

$$\lim_{\varepsilon\to 0}\varepsilon\log\varepsilon=0$$

即可得出此性质成立的结论。

此性质说明，当信源的消息个数增多时，若增加的消息发生的概率极小(接近于 0)，而原有消息的概率在信源总体上几乎没有改变，则信源的熵不变。从某一消息的自信息考虑，虽然这些概率很小的新事件发生后会给予接收者较多的信息，但对于信源总体而言，这种概率很小的消息几乎不会出现，它在熵的计算中占的比重很小。

可见，熵的扩展性也是其总体统计平均性的一种体现。

5. 可加性

统计独立的信源 X 和 Y 的联合信源的熵等于分别熵之和，即：

$$H(XY) = H(X) + H(Y)$$

在对信息概念的讨论中，为了使先后收到的信息可简单地相加，规定信息的度量方式必须满足可加性。下面我们以相互统计独立的情况为例，证明熵函数具有这种性质。

证明：设有两个随机变量 X 和 Y，其中 X 有 n 种取值，概率分布为 q_1，q_2，…，q_n，Y 有 m 种取值，概率分布为 r_1，r_2，…，r_m。

由熵函数的定义式可知 X 和 Y 的熵分别为

$$H(X) = -\sum_{i=1}^{n} q_i \log q_i$$

$$H(Y) = -\sum_{j=1}^{m} r_j \log r_j$$

现在考察由相互独立的 X、Y 构成的联合信源 XY。联合信源 XY 有 nm 种取值可能，它的概率分布为 X、Y 的联合概率分布，即

$$P(X = x_i, Y = y_j) = P(x_i, y_j) = p_{ij} \qquad i = 1, 2, \cdots, n; \ j = 1, 2, \cdots, m$$

联合信源的熵为

$$H(XY) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log p_{ij}$$

由于 X 与 Y 相互独立，有 $p_{ij} = q_i \cdot r_j$，因此：

$$\begin{aligned} H(XY) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} q_i \cdot r_j \log q_i \cdot r_j \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m} q_i \cdot r_j \log q_i - \sum_{i=1}^{n}\sum_{j=1}^{m} q_i \cdot r_j \log r_j \\ &= -\left(\sum_{i=1}^{n} q_i \log q_i\right)\left(\sum_{j=1}^{m} r_j\right) - \left(\sum_{j=1}^{m} r_j \log r_j\right)\left(\sum_{i=1}^{n} q_i\right) \\ &= -\sum_{i=1}^{n} q_i \log q_i - \sum_{j=1}^{m} r_j \log r_j \\ &= H(X) + H(Y) \end{aligned}$$

对于 X 和 Y 非相互独立的一般情况也有类似的结论，只是性质的表达和证明过程要复杂一些。

6. 极值性（最大熵定理）

$$H(p_1, p_2, \cdots, p_n) \leqslant H\left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right) = \log n \tag{2.14}$$

即当信源符号等概率分布时，熵达到其最大值。

为了证明式(2.14)，首先证明下面两个引理。

引理 2.1

$$\ln x \leqslant x - 1 \qquad x > 0 \tag{2.15}$$

证明：构造辅助函数 $f(x) = \ln x - (x-1)$，计算其极大值。

$$f'(x) = \frac{1}{x} - 1$$

令上式为 0，可以求出 $f(x)$ 在 $x=1$ 处取得极值。

由 $f''(x) = -\frac{1}{x^2} < 0$ 可知，当 $x>0$ 时，$f(x)$ 是上凸函数。

因此，在 $x=1$ 处函数 $f(x)$有极大值，$f(x)$的极大值为

$$f(x=1)=0$$

因此

$$f(x)=\ln x-(x-1)\leqslant 0 \qquad x>0$$

即

$$\ln x<x-1$$

证毕。

引理 2.2

$$H(p_1,\ p_2,\ \cdots,\ p_n)\leqslant-\sum_{i=1}^{n}p_i\log r_i \tag{2.16}$$

式中，$\sum_{i=1}^{n}p_i=1$，$\sum_{i=1}^{n}r_i=1$。

证明：考察

$$H(p_1,\ p_2,\ \cdots,\ p_n)+\sum_{i=1}^{n}p_i\log r_i=-\sum_{i=1}^{n}p_i\log p_i+\sum_{i=1}^{n}p_i\log r_i$$

$$=\sum_{i=1}^{n}p_i\log\frac{r_i}{p_i}=-\frac{1}{\ln 2}\sum_{i=1}^{n}p_i\ln\frac{r_i}{p_i}$$

易知 $r_i/p_i>0$，由式(2.15)可知

$$-\frac{1}{\ln 2}\sum_{i=1}^{n}p_i\ln\frac{r_i}{p_i}\leqslant\frac{1}{\ln 2}\sum_{i=1}^{n}p_i\left(\frac{r_i}{p_i}-1\right)=\frac{1}{\ln 2}\left[\sum_{i=1}^{n}r_i-\sum_{i=1}^{n}p_i\right]=0$$

因此

$$H(p_1,\ p_2,\ \cdots,\ p_n)\leqslant-\sum_{i=1}^{n}p_i\log r_i$$

证毕。

在式(2.16)中，若令 $r_i=\frac{1}{n}$，则式(2.16)成为：

$$H(p_1,\ p_2,\ \cdots,\ p_n)\leqslant\log n\sum_{i=1}^{n}p_i=\log n \tag{2.17}$$

此即式(2.14)表示的最大熵定理。

最大熵定理表明，对于一个输出 n 种消息的信源，在其可能的概率分布中，当信源输出的消息为等概率分布时，信源的信息熵最大，且该最大的信息熵为 $\log n$。

对于二进制信源，例 2.4 已经表明，当 $p(0)=p(1)=1/2$ 时信源的信息熵最大，即二进制信源的最大熵为

$$H\left(\frac{1}{2},\ \frac{1}{2}\right)=\log 2=1 \quad \text{比特 / 符号}$$

7. 上凸性

熵函数 $H(p_1,\ p_2,\ \cdots,\ p_n)$是概率矢量 $\boldsymbol{P}=(p_1,\ p_2,\ \cdots,\ p_n)$的上凸函数，即对于概率矢量 $\boldsymbol{P}$、$\boldsymbol{Q}$ 和 $\alpha(0<\alpha<1)$，有：

$$H(\alpha\boldsymbol{P}+(1-\alpha)\boldsymbol{Q})\geqslant\alpha H(\boldsymbol{P})+(1-\alpha)H(\boldsymbol{Q}) \tag{2.18}$$

证明：设 K 为 n 维随机变量的概率分布构成的集合。

在 K 中任意找出两组不同的概率分布：

$$\boldsymbol{P}=(p_1,\ p_2,\ \cdots,\ p_n)$$

$$\boldsymbol{Q}=(q_1,\ q_2,\ \cdots,\ q_n)$$

显然，这两组概率分布满足概率分布律，即

$$\sum_{i=1}^{n}p_i=\sum_{i=1}^{n}q_i=1\qquad 0\leqslant p_i,\ q_i\leqslant 1;\ i=1,\ 2,\ \cdots,\ n$$

在这两组概率分布下，分别有熵函数 $H(\boldsymbol{P})$和 $H(\boldsymbol{Q})$。

取 $0<\alpha<1$，由概率矢量 $\boldsymbol{P}$、$\boldsymbol{Q}$ 的线性组合可构成一组新的概率分布 $\boldsymbol{R}$：

$$\begin{aligned}\boldsymbol{R}&=\alpha\boldsymbol{P}+(1-\alpha)\boldsymbol{Q}\\&=(\alpha p_1+(1-\alpha)q_1,\ \alpha p_2+(1-\alpha)q_2,\ \cdots,\ \alpha p_n+(1-\alpha)q_n)\end{aligned}$$

因为

$$0\leqslant\alpha p_i+(1-\alpha)q_i\leqslant\alpha+(1-\alpha)=1\qquad i=1,\ 2,\ \cdots,\ n$$

且

$$\sum_{i=1}^{n}(\alpha p_i+(1-\alpha)q_i)=\alpha\sum_{i=1}^{n}p_i+(1-\alpha)\sum_{i=1}^{n}q_i=\alpha+(1-\alpha)=1$$

这组新的概率分布 $\boldsymbol{R}$ 满足概率分布律，故有：

$$\boldsymbol{R}=\alpha\boldsymbol{P}+(1-\alpha)\boldsymbol{Q}\in\boldsymbol{K}$$

在由 $\boldsymbol{P}$、$\boldsymbol{Q}$ 的线性组合得到的新的概率分布 $\boldsymbol{R}$ 下，熵函数为

$$\begin{aligned}H(\boldsymbol{R})&=H(\alpha\boldsymbol{P}+(1-\alpha)\boldsymbol{Q})\\&=-\sum_{i=1}^{n}(\alpha p_i+(1-\alpha)q_i)\log(\alpha p_i+(1-\alpha)q_i)\\&=-\alpha\sum_{i=1}^{n}p_i\log[(\alpha p_i+(1-\alpha)q_i)]-(1-\alpha)\sum_{i=1}^{n}q_i\log[(\alpha p_i+(1-\alpha)q_i)]\\&=-\alpha\sum_{i=1}^{n}p_i\log\left[(\alpha p_i+(1-\alpha)q_i)\frac{p_i}{p_i}\right]-(1-\alpha)\sum_{i=1}^{n}q_i\log\left[(\alpha p_i+(1-\alpha)q_i)\frac{q_i}{q_i}\right]\\&=-\alpha\sum_{i=1}^{n}p_i\log p_i-\alpha\sum_{i=1}^{n}p_i\log\left[(\alpha p_i+(1-\alpha)q_i)\frac{1}{p_i}\right]\\&\quad-(1-\alpha)\sum_{i=1}^{n}q_i\log q_i-(1-\alpha)\sum_{i=1}^{n}q_i\log\left[(\alpha p_i+(1-\alpha)q_i)\frac{1}{q_i}\right]\\&=\alpha H(\boldsymbol{P})+(1-\alpha)H(\boldsymbol{Q})-\alpha\sum_{i=1}^{n}p_i\log\left[(\alpha p_i+(1-\alpha)q_i)\frac{1}{p_i}\right]\\&\quad-(1-\alpha)\sum_{i=1}^{n}q_i\log\left[(\alpha p_i+(1-\alpha)q_i)\frac{1}{q_i}\right]\end{aligned}\tag{2.19}$$

由于对数函数是上凸函数，因此式(2.19)的第三项中：

$$\begin{aligned}\sum_{i=1}^{n}p_i\log\left[(\alpha p_i+(1-\alpha)q_i)\frac{1}{p_i}\right]&\leqslant\log\Big(\sum_{i=1}^{n}p_i(\alpha p_i+(1-\alpha)q_i)\frac{1}{p_i}\Big)\\&=\log\Big(\sum_{i=1}^{n}(\alpha p_i+(1-\alpha)q_i)\Big)\\&=\log\Big(\alpha\sum_{i=1}^{n}p_i+(1-\alpha)\sum_{i=1}^{n}q_i\Big)\\&=\log(\alpha+(1-\alpha))=\log 1=0\end{aligned}$$

同理，式(2.19)的第四项中

$$\sum_{i=1}^{n} q_i \log\left[(\alpha p_i + (1-\alpha)q_i)\frac{1}{q_i}\right] \leqslant 0$$

因此式(2.19)中的后两项：

$$-\alpha\sum_{i=1}^{n} p_i \log\left[(\alpha p_i + (1-\alpha)q_i)\frac{1}{p_i}\right] - (1-\alpha)\sum_{i=1}^{n} q_i \log\left[(\alpha p_i + (1-\alpha)q_i)\frac{1}{q_i}\right] \geqslant 0$$

故有：

$$H(\alpha \boldsymbol{P} + (1-\alpha)\boldsymbol{Q}) \geqslant \alpha H(\boldsymbol{P}) + (1-\alpha)H(\boldsymbol{Q})$$

即式(2.18)表示的上凸函数关系成立，熵函数是上凸函数。

证毕。

利用熵函数的上凸性，我们也可以找出信源的最大熵和达到最大熵时信源消息的概率分布。

【例 2.5】 已知信源 X 输出 n 种符号，试用求导法计算此信源的最大熵，以及达到最大熵时 n 个信源符号的概率分布。

解：因为熵函数是上凸函数，所以若它的极值存在，则必为极大值，且该极大值即为给定信源的最大熵。

设 $\boldsymbol{P}=(p_1, p_2, \cdots, p_n)$ 为信源 X 的一组概率矢量，其熵函数：

$$H(\boldsymbol{P}) = -\sum_{i=1}^{n} p_i \log p_i$$

满足条件：

$$\sum_{i=1}^{n} p_i = 1$$

题中所求为熵函数 $H(\boldsymbol{P})$ 在此条件下的条件极值。

首先以 λ 为待定常数作辅助函数：

$$\Phi(\boldsymbol{P}) = -\sum_{j=1}^{n} p_j \log p_j + \lambda\left(\sum_{j=1}^{n} p_j - 1\right)$$

对于 $i=1, 2, \cdots, n$，计算辅助函数 $\Phi(\boldsymbol{P})$ 关于 p_i 的偏导：

$$\frac{\partial \Phi}{\partial p_i} = \frac{\partial\left[-\sum_{j=1}^{n} p_j \frac{\ln p_j}{\ln 2} + \lambda\left(\sum_{j=1}^{n} p_j - 1\right)\right]}{\partial p_i}$$

$$= -p_i \frac{1}{p_i \ln 2} - \log p_i + \lambda = -\frac{1}{\ln 2} - \log p_i + \lambda$$

令

$$\frac{\partial \Phi}{\partial p_i} = 0$$

得到：

$$-\frac{1}{\ln 2} - \log p_i + \lambda = 0 \qquad i = 1, 2, \cdots, n$$

注意，此为一个由 n 个方程构成的方程组。解此方程组得到熵函数 $H(\boldsymbol{P})$ 取得极值的概率分布，即得到：

$$\log p_i = \lambda - \frac{1}{\ln 2}$$

$$p_i = 2^{\lambda-\frac{1}{\ln 2}} = \text{常数} \qquad i = 1, 2, \cdots, n$$

由于 p_i 是与 i 无关的常数，且必须满足概率分布的条件，可知熵函数 $H(\boldsymbol{P})$的极值点概率分布必为

$$p_i = \frac{1}{n} \qquad i = 1, 2, \cdots, n$$

因此熵函数 $H(\boldsymbol{P})$的最大值为

$$H_{\max}(\boldsymbol{P}) = H\left(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right) = n \cdot \frac{1}{n}\log n = \log n$$

因此对于具有 n 种输出符号的信源 X，当信源符号等概率分布时，信源的信息熵 $H(X)$达到其最大值 $\log n$。

此结论与关于熵的极值性的讨论结果是相同的。

2.4 联合事件的熵及其关系

2.3 节我们从信源输出随机变量 X 的概率空间出发，对信源符号的不确定性进行了分析并引出了信源符号的自信息和信源的熵。然而，在对一个通信系统进行分析时，我们遇到的往往并不是单个的随机变量，而是由若干随机变量构成的联合事件。因此需要对由若干随机变量组成的联合事件所具有的不确定性加以讨论。此处，我们以两个随机变量为例，在 2.3 节讨论的熵的概念与定义的基础上，讨论联合事件的熵及其关系。

2.4.1 联合事件的概率空间与概率关系

设有两个离散随机变量 X、Y，其中：

$$X \in \{x_1, x_2, \cdots, x_r\}, \quad Y \in \{y_1, y_2, \cdots, y_s\} \tag{2.20}$$

X 和 Y 构成的联合事件用二维随机变量(X, Y)表示。

二维随机变量(X, Y)的性质不仅与 X 和 Y 有关，而且依赖于两者之间的关系。因此对于由 X 和 Y 组成的联合事件，仅仅单独研究 X 或 Y 是不够的，还需要将(X, Y)作为一个整体来研究。

1. 联合概率分布

已知二维随机变量(X, Y)的取值为(x_i, y_j)，其发生的概率为联合概率 $P(x_i, y_j)$，它们满足：

$$\begin{cases} 0 \leqslant P(x_i, y_j) \leqslant 1 \\ \sum\limits_{i=1}^{r}\sum\limits_{j=1}^{s} P(x_i, y_j) = 1 \end{cases} \quad i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s \tag{2.21}$$

因此，联合事件(X, Y)的概率空间为

$$[XY, P(x_i, y_j)] \tag{2.22}$$

2. 边缘概率分布

对于二维随机变量(X, Y)，X 和 Y 也是随机变量，将随机变量 X 和 Y 各自的概率分布分别称为二维随机变量(X, Y)关于 X 和 Y 的边缘概率分布，并且边缘概率分布由联合

概率分布决定。

依据二维随机变量(X, Y)的联合概率分布$P(x_i, y_j)$，可以求出某一随机变量的边缘概率分布，即

$$P(X=x_i)=P(x_i)=\sum_{j=1}^{s}P(x_i, y_j) \qquad i=1, 2, \cdots, r \tag{2.23}$$

$$P(Y=y_j)=P(y_j)=\sum_{i=1}^{r}P(x_i, y_j) \qquad j=1, 2, \cdots, s \tag{2.24}$$

它们满足：

$$0 \leqslant P(x_i), P(y_j) \leqslant 1 \qquad i=1, 2, \cdots, r; j=1, 2, \cdots, s$$

$$\sum_{i=1}^{r}P(x_i)=\sum_{j=1}^{s}P(y_j)=1$$

3. 条件概率分布

对于二维随机变量(X, Y)，如果已知其联合概率分布$P(x_i, y_j)$和某一事件X或Y的边缘概率分布$P(x_i)$和$P(y_j)$，则可以求出在某事件已发生的条件下另一事件发生的概率，并将此概率称为条件概率。

已知$X=x_i$已发生的条件下$Y=y_j$发生的条件概率分布为

$$P(y_j \mid x_i)=\frac{P(x_i, y_j)}{P(x_i)} \tag{2.25}$$

同理，已知$Y=y_j$已发生的条件下$X=x_i$发生的条件概率分布为

$$P(x_i \mid y_j)=\frac{P(x_i, y_j)}{P(y_j)} \tag{2.26}$$

其中，设

$$P(x_i)>0, P(y_j)>0 \qquad i=1, 2, \cdots, r; j=1, 2, \cdots, s$$

条件概率分布满足：

$$\begin{cases} 0 \leqslant P(y_j \mid x_i), P(x_i \mid y_j) \leqslant 1 \\ \sum_{j=1}^{s}P(y_j \mid x_i)=1, \sum_{i=1}^{r}P(x_i \mid y_j)=1 \end{cases} \qquad i=1, 2, \cdots r; j=1, 2, \cdots, s \tag{2.27}$$

2.4.2　联合熵、无条件熵和条件熵

对于由X、Y组成的联合事件，由于(X, Y)是一个随机矢量，因此我们对联合事件(X, Y)的某一样值是否发生具有不确定性。联合事件(X, Y)的平均不确定性大小可以用联合事件(X, Y)的联合熵进行度量。

若已知联合事件(X, Y)的概率空间为$[XY, P(x_i, y_j)]$，则联合事件(X, Y)的联合熵为

$$H(XY)=\sum_{i=1}^{r}\sum_{j=1}^{s}P(x_i, y_j)\log\frac{1}{P(x_i, y_j)}=-\sum_{i=1}^{r}\sum_{j=1}^{s}P(x_i, y_j)\log P(x_i, y_j) \tag{2.28}$$

若已知某一随机变量的边缘概率分布，则根据熵的定义式(2.6)可以写出联合事件(X, Y)中关于某一随机变量的熵，即

$$H(X) = -\sum_{i=1}^{r} P(x_i)\log P(x_i) = -\sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log P(x_i) \tag{2.29}$$

$$H(Y) = -\sum_{j=1}^{s} P(y_j)\log P(y_j) = -\sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log P(y_j) \tag{2.30}$$

$H[X]$和$H[Y]$分别表示联合事件(X, Y)中关于随机变量X和Y的平均不确定性，它们被称为无条件熵或边缘熵。

对于由多个随机变量组成的联合事件，我们不仅要了解它们的联合不确定性和某一分量的不确定性，还需要了解在某些随机变量已出现的条件下，关于另一随机变量的不确定性。这种平均不确定性即为条件熵。

设已知联合事件(X, Y)中x_i已发生时y_j发生的概率为条件概率$P(y_j|x_i)$，则定义已知X时关于Y的条件熵为

$$H(Y \mid X) = \sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log \frac{1}{P(y_j \mid x_i)} \tag{2.31}$$

它表示在随机变量X已知时，我们对于随机变量Y仍存在的平均不确定性。

同理，已知Y时关于X的条件熵为

$$H(X \mid Y) = \sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log \frac{1}{P(x_i \mid y_j)} \tag{2.32}$$

2.4.3 各种熵之间的关系

1. 条件熵

条件熵满足如下性质：

$$H(Y \mid X) \leqslant H(Y) \tag{2.33}$$

当且仅当X与Y统计独立时等式成立。

证明：

$$\begin{aligned} H(Y \mid X) - H(Y) &= \sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log \frac{1}{P(y_j \mid x_i)} - \sum_{j=1}^{s} P(x_i, y_j)\log \frac{1}{P(y_j)} \\ &= \sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log \frac{P(y_j)}{P(y_j \mid x_i)} \cdot \frac{P(x_i)}{P(x_i)} \\ &= \sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log \frac{P(x_i)\cdot P(y_j)}{P(x_i, y_j)} \end{aligned} \tag{2.34}$$

由于对数函数是上凸函数，因此应用颜森不等式：

$$\begin{aligned} \sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log \frac{P(x_i)\cdot P(y_j)}{P(x_i, y_j)} &\leqslant \log\Big[\sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\,\frac{P(x_i)\cdot P(y_j)}{P(x_i, y_j)}\Big] \\ &= \log\Big[\sum_{i=1}^{r} P(x_i)\sum_{j=1}^{s} P(y_j)\Big] = \log 1 = 0 \end{aligned}$$

当X、Y统计独立时：

$$P(x_i, y_j) = P(x_i)\cdot P(y_j)$$

由式(2.34)第三步可看出等式成立。

同理，存在：

$$H(X \mid Y) \leqslant H(X)$$

此性质表明，在联合事件中某一随机变量的条件熵是有界的，它小于等于该随机变量的无条件熵。

2. 联合熵、无条件熵和条件熵的关系

$$\begin{aligned} H(XY) &= \sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log\frac{1}{P(x_i, y_j)} \\ &= \sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log\frac{1}{P(y_j \mid x_i)\cdot P(x_i)} \\ &= \sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log\frac{1}{P(y_j \mid x_i)} + \sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log\frac{1}{P(x_i)} \end{aligned} \tag{2.35}$$

所以

$$H(XY) = H(Y \mid X) + H(X) \tag{2.36}$$

可见，关于联合事件(X, Y)的不确定性等于关于 X 的不确定性与一旦观测到 X 之后仍然保留的关于 Y 的不确定性之和。

同理，可得到：

$$H(XY) = H(X \mid Y) + H(Y) \tag{2.37}$$

若联合事件(X, Y)中各分量相互独立，又由于 $P(y_j \mid x_i) = P(y_j)$，则

$$\begin{aligned} H(Y \mid X) &= \sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log\frac{1}{P(y_j \mid x_i)} \\ &= \sum_{i=1}^{r}\sum_{j=1}^{s} P(x_i, y_j)\log\frac{1}{P(y_j)} = H(Y) \end{aligned} \tag{2.38}$$

同理，依据 $P(x_i \mid y_j) = P(x_i)$有：

$$H(X \mid Y) = H(X) \tag{2.39}$$

因此，若联合事件(X, Y)中各分量相互独立，则它们的联合熵为

$$H(XY) = H(X) + H(Y) \tag{2.40}$$

【例 2.6】 已知随机变量 X 和 Y 组成联合事件(X, Y)。二维随机变量(X, Y)的联合概率分布如表 2-1 所示。

表 2-1　$P(x_i, y_j)$

X \ Y	y_1	y_2	y_3
x_1	$\frac{1}{8}$	0	$\frac{1}{8}$
x_2	$\frac{1}{16}$	$\frac{1}{16}$	0
x_3	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
x_4	$\frac{3}{16}$	$\frac{1}{16}$	0

(1) 计算联合事件(X, Y)的联合熵 $H(XY)$；

(2) 计算随机变量 X 和 Y 的边缘熵 $H(X)$和 $H(Y)$；

(3) 计算条件熵 $H(Y|X)$和 $H(X|Y)$；

(4) 验证联合熵与边缘熵、条件熵之间的关系。

解：(1) 依据联合事件(X, Y)的联合概率分布计算(X, Y)的联合熵 $H(XY)$为

$$H(XY)=\sum_{i=1}^{4}\sum_{j=1}^{3}P(x_i, y_j)\log\frac{1}{P(x_i, y_j)}$$

$$=\frac{5}{8}\log 8+\frac{3}{16}\log 16+\frac{3}{16}\log\frac{16}{3}=\frac{5}{8}\log 8+\frac{6}{16}\log 16-\frac{3}{16}\log 3$$

$$=\frac{5}{8}\times 3+\frac{6}{16}\times 4-\frac{3}{16}\log 3\approx 3.078 \quad \text{比特 / 符号}$$

(2) 依据边缘概率的计算关系，首先求出随机变量 X 的边缘概率为

$$P(x_1)=\sum_{j=1}^{3}P(x_1, y_j)=\frac{1}{8}+0+\frac{1}{8}=\frac{1}{4}$$

$$P(x_2)=\sum_{j=1}^{3}P(x_2, y_j)=\frac{1}{16}+0+\frac{1}{16}=\frac{1}{8}$$

$$P(x_3)=\sum_{j=1}^{3}P(x_3, y_j)=\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}$$

$$P(x_4)=\sum_{j=1}^{3}P(x_4, y_j)=\frac{3}{16}+\frac{1}{16}=\frac{1}{4}$$

因此随机变量 X 的边缘熵 $H(X)$为

$$H(X)=\sum_{i=1}^{4}P(x_i)\log\frac{1}{P(x_i)}=\frac{1}{4}\log 4+\frac{1}{8}\log 8+\frac{3}{8}\log\frac{8}{3}+\frac{1}{4}\log 4$$

$$=\frac{5}{2}-\frac{3}{8}\log 3\approx 1.906 \quad \text{比特 / 符号}$$

同理，可以求出随机变量 Y 的边缘概率和边缘熵 $H(Y)$：

$$P(y_1)=\sum_{i=1}^{4}P(x_i, y_1)=\frac{1}{8}+\frac{1}{16}+\frac{1}{8}+\frac{3}{16}=\frac{1}{2}$$

$$P(y_2)=\sum_{i=1}^{4}P(x_i, y_2)=0+\frac{1}{16}+\frac{1}{8}+\frac{1}{16}=\frac{1}{4}$$

$$P(y_3)=\sum_{i=1}^{4}P(x_i, y_3)=\frac{1}{8}+0+\frac{1}{8}+0=\frac{1}{4}$$

计算得到随机变量 Y 的边缘熵为

$$H(Y)=\sum_{j=1}^{3}P(y_j)\log\frac{1}{P(y_j)}$$

$$=\frac{1}{2}\log 2+\frac{1}{4}\log 4+\frac{1}{4}\log 4=1.5 \quad \text{比特 / 符号}$$

(3) 已知联合事件(X, Y)的联合概率 $P(x_i, y_j)$和随机变量 X 的边缘概率 $P(x_i)$。依据条件概率的计算关系 $P(y_j|x_i)=\dfrac{P(x_i, y_j)}{P(x_i)}$，可以求出在已知 $X=x_i$ 发生的条件下关于 $Y=y_j$ 发生的条件概率分布 $P(y_j|x_i)$，其中 $i=1, 2, 3, 4$，$j=1, 2, 3$，如表 2-2 所示。

表 2－2　$P(y_j|x_i)$

X \ Y	y_1	y_2	y_3
x_1	$\frac{1}{4}$	0	$\frac{1}{2}$
x_2	$\frac{1}{8}$	$\frac{1}{4}$	0
x_3	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$
x_4	$\frac{3}{8}$	$\frac{1}{4}$	0

可以求出已知 X 时关于 Y 的条件熵为

$$\begin{aligned}H(Y\mid X)&=\sum_{i=1}^{4}\sum_{j=1}^{3}P(x_i,\ y_j)\log\frac{1}{P(y_j\mid x_i)}\\&=\frac{1}{8}\log2+\frac{1}{8}\log2+\frac{1}{16}\log2+\frac{1}{16}\log2+\frac{1}{8}\log3\\&\quad+\frac{1}{8}\log3+\frac{1}{8}\log3+\frac{3}{16}\log\frac{4}{3}+\frac{1}{16}\log4\\&=\frac{7}{8}+\frac{3}{16}\log3\approx1.172\quad\text{比特 / 符号}\end{aligned}$$

同理，依据 $P(x_i|y_j)=\dfrac{P(x_i,\ y_j)}{P(y_j)}$**可以得到已知** $Y=y_j$ **发生的条件下，关于** $X=x_i$ **发生的条件概率分布** $P(x_i|y_j)$**(如表** 2－3 **所示)**。

表 2－3　$P(x_i|y_j)$

X \ Y	y_1	y_2	y_3
x_1	$\frac{1}{2}$	0	$\frac{1}{2}$
x_2	$\frac{1}{2}$	$\frac{1}{2}$	0
x_3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
x_4	$\frac{3}{4}$	$\frac{1}{4}$	0

可以求出已知 X 时关于 Y 的条件熵为

$$H(X \mid Y)=\sum_{i=1}^{4}\sum_{j=1}^{3}P(x_i,\ y_j)\log\frac{1}{P(x_i \mid y_j)}$$

$$=\frac{1}{8}\log4+\frac{1}{8}\log2+\frac{1}{16}\log8+\frac{1}{16}\log4+\frac{1}{8}\log4$$

$$+\frac{1}{8}\log2+\frac{1}{8}\log2+\frac{3}{16}\log\frac{8}{3}+\frac{1}{16}\log4$$

$$=\frac{15}{8}-\frac{3}{16}\log3\approx1.578 \quad \text{比特/符号}$$

(4) 由上述计算结果可以验证联合熵、边缘熵和条件熵之间满足的关系。

下面验证某一随机变量的无条件熵与条件熵的关系。

由计算结果可以看出，在联合事件(X, Y)中，关于随机变量 X 的边缘熵为

$$H(X)=-\sum_{i=1}^{4}\sum_{j=1}^{3}P(x_i,\ y_j)\log P(x_i)=1.906 \quad \text{比特/符号}$$

当已知随机变量 Y 时，关于随机变量 X 的条件熵为

$$H(X \mid Y)=\sum_{i=1}^{4}\sum_{j=1}^{3}P(x_i,\ y_j)\log\frac{1}{P(x_i \mid y_j)}=1.578 \quad \text{比特/符号}$$

显然，随机变量 X 的条件熵与其无条件熵的关系满足：

$$H(X \mid Y)<H(X)$$

同理，随机变量 Y 的条件熵与其无条件熵的关系也满足：

$$H(Y \mid X)<H(Y)$$

由此可以看出，对于联合事件(X, Y)，某一随机变量的条件熵小于等于该随机变量的无条件熵。

下面由计算结果验证边缘熵、条件熵和联合熵之间的关系。

由此例计算得到的联合事件(X, Y)的边缘熵：

$$H(X)=1.906 \quad \text{比特/符号}, \quad H(Y)=1.5 \quad \text{比特/符号}$$

条件熵：

$$H(Y \mid X)=1.172 \quad \text{比特/符号}, \quad H(X \mid Y)=1.578 \quad \text{比特/符号}$$

联合熵：

$$H(XY)=3.078 \quad \text{比特/符号}$$

可以看出，它们之间满足：

$$H(X)+H(Y \mid X)=1.906+1.172=3.078=H(XY)$$

和

$$H(Y)+H(X \mid Y)=1.5+1.578=3.078=H(XY)$$

即联合事件(X, Y)的不确定性(联合事件的联合熵 $H(XY)$)等于关于某一随机变量的不确定性与一旦观测到该随机变量后仍然保留的关于另一随机变量的不确定性之和。

2.5 连续信源的信息测度

至此，我们所研究的都是取值为有限或可数的离散信源，它们输出的消息属于时间离散、取值有限或可数的随机序列，其统计特性可以用联合概率分布来描述，而某些实际信

源的输出是时间和取值都是连续的消息。例如语音信号 $X(t)$、电视信号 $X(x_0, y_0, t)$等都是时间的连续波形。当固定某一时刻 $t=t_0$ 时，它们的可能取值也是连续的。这样的信源称为随机波形信源。随机波形信源输出的消息是随机的，因此，可以用随机过程$\{x(t)\}$来描述。

2.5.1　连续信源的差熵

基本连续信源的输出是取值连续的单个随机变量，可用变量的概率密度、变量间的条件概率密度和联合概率密度来描述。变量的一维概率密度函数为

$$p_X(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}$$

$$p_Y(y) = \frac{\mathrm{d}F(y)}{\mathrm{d}y}$$

一维概率分布函数为

$$F(x_1) = P[X \leqslant x_1] = \int_{-\infty}^{x_1} p_X(x)\mathrm{d}x$$

条件概率密度函数为

$$p_{Y|X}(y \mid x),\ p_{X|Y}(x \mid y)$$

联合概率密度函数为

$$p_{XY}(x_1y_1) = \frac{\partial^2 F(x_1, y_1)}{\partial x_1 \partial y_1}$$

它们之间的关系为

$$p_{XY}(xy) = p_X(x)p_{Y|X}(y \mid x) \tag{2.41}$$

$$p_{XY}(xy) = p_Y(y)p_{X|Y}(x \mid y) \tag{2.42}$$

这些边缘概率密度函数满足：

$$p_X(x) = \int_{\mathbf{R}} p_{XY}(xy)\mathrm{d}y \tag{2.43}$$

$$p_Y(y) = \int_{\mathbf{R}} p_{XY}(xy)\mathrm{d}x \tag{2.44}$$

其中，X 和 Y 的取值域为全实数集 **R**。

若概率密度在有限区域内分布，则可认为在该区间之外所有概率密度函数为零。上述密度函数中的脚标表示所牵涉的变量的总体，而自变量(如 x，y，…)则是具体取值。因为概率密度函数是不同的函数，所以用脚标来加以区分，以免混淆。为了简化书写，往往省去脚标，但在使用时要注意上述问题。

基本连续信源的数学模型为

$$X = \begin{bmatrix} \mathbf{R} \\ p(x) \end{bmatrix}$$

并满足：

$$\int_{\mathbf{R}} p(x)\mathrm{d}x = 1$$

其中，**R** 是全实数集，是连续变量 X 的取值范围。根据前述的离散化原则，连续变量 X 可

量化分层后用离散变量描述。量化单位越小，则所得的离散变量和连续变量越接近。因此，连续变量的信息测度可以用离散变量的信息测度来逼近。

假定连续信源 X 的概率密度函数 $p(x)$ 如图 2-4 所示。

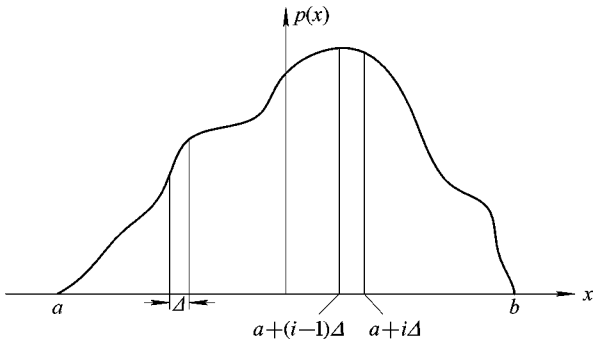

图 2-4 概率密度分布

我们把取值区间$[a, b]$分割成 n 个小区间，各小区间设有等宽 $\Delta=\left(\frac{b-a}{n}\right)$。那么，$X$ 处于第 i 区间的概率 P_i 是

$$\begin{aligned} P_i &= P\{a+(i-1)\Delta \leqslant x \leqslant a+i\Delta\} \\ &= \int_{a+(i-1)\Delta}^{a+i\Delta} p(x)\mathrm{d}x = p(x_i)\Delta \qquad i=1, 2, \cdots, n \end{aligned} \tag{2.45}$$

其中，x_i 是 $a+(i-1)\Delta$ 到 $a+i\Delta$ 之间的某一值。当 $p(x)$ 是 x 的连续函数时，由积分中值定理可知，必存在一个 x_i 值使式(2.45)成立。这样，连续变量 X 就可用取值为 $x_i(i=1, 2, \cdots, n)$的离散变量 X_n 来近似，连续信源 X 就被量化成离散信源：

$$\begin{bmatrix} X_n \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1)\Delta & p(x_2)\Delta & \cdots & p(x_n)\Delta \end{bmatrix}$$

且

$$\sum_{i=1}^{n} p(x_i)\Delta = \sum_{i=1}^{n}\int_{a+(i-1)\Delta}^{a+i\Delta} p(x)\mathrm{d}x = \int_a^b p(x)\mathrm{d}x = 1$$

这时离散信源 X_n 的熵是

$$\begin{aligned} H(X_n) &= -\sum_i P_i \log P_i = -\sum_i p(x_i)\Delta \log[p(x_i)\Delta] \\ &= -\sum_i p(x_i)\Delta \log p(x_i) - \sum_i p(x_i)\Delta \log\Delta \end{aligned}$$

当 $n\to\infty$，$\Delta\to 0$ 时，离散随机变量 X_n 趋于连续随机变量 X，而离散信源 X_n 的熵 $H(X_n)$的极限值就是连续信源的信息熵，即

$$\begin{aligned} H(X) &= \lim_{n\to\infty} H(X_n) \\ &= \lim_{\Delta\to 0} -\sum_i p(x_i)\Delta \log p(x_i) - \lim_{\Delta\to 0}(\log\Delta)\sum_i p(x_i)\Delta \\ &= -\int_a^b p(x)\log p(x)\mathrm{d}x - \lim_{\Delta\to 0}\log\Delta \end{aligned} \tag{2.46}$$

一般情况下，式(2.46)的第一项是定值。当 $\Delta\to 0$ 时，第二项是趋于无限大的常数。所以避开第二项，定义连续信源的熵为

$$h(X) = -\int_{\mathbf{R}} p(x)\log p(x)\mathrm{d}x \tag{2.47}$$

由式(2.47)可知，所定义的连续信源的熵并不是实际信源输出的绝对熵，而连续信源的绝对熵应该还要加上一项无限大的常数项。这一点是可以理解的，因为连续信源的可能取值数是无限多个，若设取值是等概率分布，那么信源的不确定性为无限大。当确知输出为某值后，所获得的信息量也将为无限大。可见，$h(X)$已不能代表信源的平均不确定性大小，也不能代表连续信源输出的信息量。既然如此，为什么要定义连续信源的熵为式(2.47)呢？一方面，因为这样定义可与离散信源的熵在形式上统一起来，另一方面，因为在实际问题中常常讨论的是熵之间的差值问题，如平均互信息。在讨论熵差时，无限大常数项将有两项，一项为正，一项为负，只要两者离散逼近时所取的间隔 Δ 一致，这两个无限大项就将互相抵消掉。因此在任何包含有熵差的问题中，式(2.47)定义的连续信源的熵具有信息的特性。由此可见，连续信源的熵称为差熵，以区别于原来的绝对熵。

同理，我们可以定义两个连续变量 X、Y 的联合熵和条件熵，即

$$h(XY) = -\iint_{\mathbf{R}} p(xy)\log p(xy)\ \mathrm{d}x\ \mathrm{d}y \tag{2.48}$$

$$h(Y \mid X) = -\iint_{\mathbf{R}} p(x)p(y \mid x)\log p(y \mid x)\ \mathrm{d}x\ \mathrm{d}y \tag{2.49}$$

$$h(X \mid Y) = -\iint_{\mathbf{R}} p(x)p(y \mid x)\log p(x \mid y)\ \mathrm{d}x\ \mathrm{d}y \tag{2.50}$$

这样定义的熵虽然在形式上和离散信源的熵相似，但在概念上不能把它作为信息熵来理解。连续信源的差熵只具有熵的部分含义和性质，而丧失了某些重要的特性。

2.5.2　差熵的基本性质

差熵仍然具有可加性、凸状性和极值性，但却不存在非负性和变换不变性等。因此，连续信源的差熵具有如下性质。

(1) **可加性**。

$$h(XY) = h(X) + h(Y \mid X) \tag{2.51}$$

$$h(XY) = h(Y) + h(X \mid Y) \tag{2.52}$$

并且类似于离散情况，可以证得：

$$h(X \mid Y) \leqslant h(X) \quad \text{或} \quad h(Y \mid X) \leqslant h(Y) \tag{2.53}$$

当且仅当 X 与 Y 统计独立时，式(2.52)和式(2.53)成立。

进而可得：

$$h(XY) \leqslant h(X) + h(Y) \tag{2.54}$$

当且仅当 X 与 Y 统计独立时，等式成立。

(2) **凸状性和极值性**。

差熵 $h(X)$是输入概率密度函数 $p(x)$的上凸函数。因此，对于某一概率密度函数，可以得到差熵的最大值。

(3) **差熵可为负值**。

在某些情况下，差熵可以为负值。

例如，若概率密度函数为

$$p(x)=\begin{cases}\dfrac{1}{b-a} & a\leqslant x\leqslant b\\ 0 & x>b,\ x<a\end{cases}$$

则

$$h(X)=-\int_a^b\frac{1}{b-a}\log\frac{1}{b-a}\mathrm{d}x=\log(b-a)\tag{2.55}$$

若 $b-a<1$，则熵 $h(X)<0$。

下面介绍高斯信源的熵值。

基本高斯信源是指信源输出的一维随机变量 X 的概率密度分布是正态分布，即

$$p(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)\tag{2.56}$$

其中，m 是 X 的均值，σ^2 是 X 的方差。这个连续信源的熵为

$$\begin{aligned}h(X)&=-\int_{-\infty}^{+\infty}p(x)\log p(x)\mathrm{d}x\\&=-\int_{-\infty}^{+\infty}p(x)\log\left[\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)\right]\mathrm{d}x\\&=-\int_{-\infty}^{+\infty}p(x)(-\log\sqrt{2\pi\sigma^2})\mathrm{d}x+\int_{-\infty}^{+\infty}p(x)\left[\frac{(x-m)^2}{2\sigma^2}\right]\mathrm{d}x\cdot\log\mathrm{e}\\&=\log\sqrt{2\pi\sigma^2}+\frac{1}{2}\log\mathrm{e}=\frac{1}{2}\log2\pi\mathrm{e}\sigma^2\end{aligned}\tag{2.57}$$

其中，因为

$$\int_{-\infty}^{+\infty}p(x)\mathrm{d}x=1$$

和

$$\int_{-\infty}^{+\infty}(x-m)^2p(x)\mathrm{d}x=\sigma^2$$

可见，正态分布的连续信源的熵与数学期望$\sqrt{\hat{P}}$无关，只与其方差 σ^2 有关。当均值 $m=0$ 时，X 的方差 σ^2 就等于信源输出的平均功率 P。由式(2.57)得

$$h(X)=\frac{1}{2}\log2\pi\mathrm{e}P\tag{2.58}$$

习 题 2

2.1 若以 5 作为信息量对数的底，试求该信息量单位与比特(bit)和奈特(nat)的换算关系。

2.2 一信源发出 m_1、m_2、m_3、m_4 四个消息，其概率分别为 1/2、1/4、1/8、1/8，计算每一消息的自信息量和信源的平均信息量。

2.3 国际间通用的莫尔斯(Morse)电码用点和画表示英文字母，设画发生的概率是点的 1/3，计算一个点、画各自的信息量，计算点画码的平均信息量。

2.4 英文字母表中各字母出现的概率如表 2-4 所示。试问：

(1) 哪个字母携带的信息量最大？

(2) 哪个字母携带的信息量最小？

(3) 如果用单词告诉你一件事情，单词第一个字母是 T 或 X，那么哪个字母对你更有帮助？

表 2-4　英文字母表中各字母出现的概率

A	0.081	H	0.051	O	0.079	V	0.009
B	0.016	I	0.072	P	0.023	W	0.020
C	0.032	J	0.001	Q	0.002	X	0.002
D	0.037	K	0.005	R	0.060	Y	0.019
E	0.124	L	0.040	S	0.066	Z	0.001
F	0.023	M	0.022	T	0.096		
G	0.016	N	0.072	U	0.031		

2.5　某信源输出符号的概率空间为$[X, P]=\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{6} & \frac{1}{9} & \frac{1}{12} & \frac{1}{18} \end{bmatrix}$，求信源的熵。

2.6　同时掷一对均匀的骰子，各面出现的概率为 1/6。

(1) 当得知两骰子“面朝上点数之和为 2”、“面朝上点数之和为 8”或“面朝上点数分别是 3 和 4”时，这三种情况分别获得多少信息量？

(2) 求两个点数之和的熵。

2.7　试证明，若 $\sum_{i=1}^{L} p_i = 1$，$\sum_{i=1}^{m} q_i = p_L$，则

$$H(p_1, p_2, \cdots, p_{L-1}, q_1, q_2, \cdots, q_m) = H(p_1, p_2, \cdots, p_L) + p_L H\left(\frac{q_1}{p_L}, \frac{q_2}{p_L}, \cdots, \frac{q_m}{p_L}\right)$$

并说明等式的物理意义。

2.8　一信源有 4 种输出数符：$x_i=i$，$i=0, 1, 2, 3$，且 $P_i=1/4$。设信源向信宿发出 x_3，但由于传输中的干扰，接收者收到 x_3 后，认为其可信度为 0.9。然后信源再次向信宿发送该字符，信宿无误收到。问信源在两次发送中发出的信息量各是多少？信宿在两次接收中得到的信息量又各是多少？(提示：先计算第二次传输中收、发的信息量。)

2.9　两个信源 S_1 和 S_2 均有两种输出：$X=0, 1$ 和 $Y=0, 1$，概率分别为 $P_{X_0}=P_{X_1}=1/2$，$P_{Y_0}=1/4$，$P_{Y_1}=3/4$。试计算 $H(X)$ 和 $H(Y)$。设 S_1 发出序列 0101，S_2 发出序列 0111，其传输过程无误，则第一个字符传送结束后，相应的两个信宿分别收到多少信息量？当整个序列传送结束后，收到的总信息量及平均每次发送的信息量又各是多少？(设信源先后发出的数字相互独立。)

2.10　设离散无记忆信源 S 的符号集 $A=\{a_1, a_2, \cdots, a_q\}$，其相应的概率分布为 $P=\{p_1, p_2, \cdots, p_q\}$，另一离散无记忆信源 S' 的符号集元素为 S 信源符号集的两倍，$A'=\{a_i' \mid i=1, 2, \cdots, 2q\}$，并且各符号的概率分布满足：

$$p_i' = (1-\varepsilon)p_i \qquad i = 1, 2, \cdots, q$$
$$p_i' = \varepsilon p_{i-q} \qquad i = 1+q, 2+q, \cdots, 2q$$

试写出信源 S' 信息熵与信源 S 信息熵的关系。

2.11 设有一概率空间 X，其概率分布为 $P=\{p_1, p_2, \cdots, p_q\}$，并有 $p_1>p_2$。若取 $p_1'=p_1-\varepsilon$，$p_2'=p_2+\varepsilon(0<2\varepsilon<p_1-p_2)$，其他概率不变。试证明由此所得新概率空间 X' 的熵是增加的，并用熵的物理意义加以解释。

2.12 X 是一个概率密度函数为 $p_X(x)$ 的随机变量，$Y=aX+b$ 是 X 的一个线性变换，此处 a 和 b 是两个常数。试根据 $h(X)$ 计算差熵 $h(Y)$。

2.13 计算下列情况下连续随机变量 X 的差熵。

(1) X 是指数分布的随机变量，参数 $\lambda>0$，即

$$f_X(x) = \begin{cases} \lambda^{-1} e^{-\frac{x}{\lambda}} & x > 0 \\ 0 & \text{其他} \end{cases}$$

(2) X 是拉普拉斯分布的随机变量，参数 $\lambda>0$，即

$$f_X(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$$

(3) X 是三角分布的随机变量，参数 $\lambda>0$，即

$$f_X(x) = \begin{cases} \dfrac{x+\lambda}{\lambda^2} & -\lambda \leqslant x \leqslant 0 \\ \dfrac{-x+\lambda}{\lambda^2} & 0 < x \leqslant \lambda \\ 0 & \text{其他} \end{cases}$$

第 3 章　离散无记忆信道与互信息

3.1　单符号离散无记忆信道及其转移概率

前面我们已经从信源输出随机变量 X 的不确定性出发，讨论了信源的平均信息量，即信源的信息熵。在实际的信息传递过程中，信源输出的信息总是需要通过信道的传输来完成，即人们获取信源输出的信息是通过对信道的输出进行观测而实现的。因此在前面的讨论中，事实上我们做了一个假定，即信道中没有随机性干扰，通过信道的传输，接收端收到的符号与信源的输出完全一致。然而，在一般的通信系统(见图 3－1)中，信道中是存在随机性干扰的。由于信道干扰的影响，其输出随机变量 Y 与输入随机变量 X(信源的输出)并非总是一样。因此对于信息传输过程的考察，不仅需要了解关于信源 X 的不确定性，而且需要分析经过系统的信息传递，在接收端观测到 Y 之后对 X 仍然存在的不确定性。

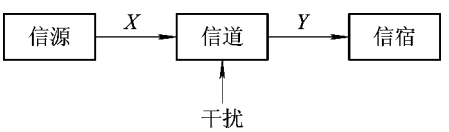

图 3－1　一般的通信系统

为了便于讨论和建立相应的概念，此处首先建立一个最简单的信道模型——单符号离散无记忆信道。

1. 单符号离散无记忆信道

图 3－2 给出了一个单符号离散无记忆信道，其输入和输出分别为离散随机变量 X 和 Y。

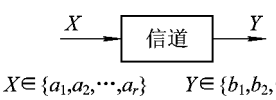

图 3－2　单符号离散无记忆信道

单符号离散无记忆信道有如下两个基本特性。

(1) 输入和输出随机变量的取值都是离散的，即

$$X \in \{a_1, a_2, \cdots, a_r\}$$

$$Y \in \{b_1, b_2, \cdots, b_s\}$$

信道的输入、输出构成了一个离散的随机变量对(X, Y)，可以借助二维随机变量加以

描述。

(2) 某一时刻信道的输出 Y 仅取决于即时信道的输入 X，与前面时刻信道的输入和输出无关。

2. 单符号离散无记忆信道的转移概率分布

由于信道中存在着随机性干扰，信道的输出 Y 统计依赖于信道的输入 X。因此，当信道的输入 $X=a_i$ 时，输出 $Y=b_j$ 发生的概率为条件概率，即

$$P(Y=b_j \mid X=a_i)=P(b_j \mid a_i) \quad i=1,2,\cdots,r;\ j=1,2,\cdots,s \tag{3.1}$$

这组条件概率 $P(b_j|a_i)$ 反映了信道输入 X 与输出 Y 之间的统计依赖关系，被称为信道的转移概率。

信道转移概率满足：

$$\begin{cases} 0 \leqslant P(b_j \mid a_i) \leqslant 1 & i=1,2,\cdots,r;\ j=1,2,\cdots,s \\ \sum\limits_{j=1}^{s} P(b_j \mid a_i)=1 & i=1,2,\cdots,r \end{cases} \tag{3.2}$$

对于有 r 种输入和 s 种输出的单符号离散无记忆信道，反映其输入 X 与输出 Y 之间的统计依赖关系的条件概率共有 $r\times s$ 个。这样的 $r\times s$ 个条件概率可以排成一个 r 行 s 列的矩阵

$$[P(b_j \mid a_i)]=\begin{bmatrix} P(b_1 \mid a_1) & P(b_2 \mid a_1) & \cdots & P(b_s \mid a_1) \\ P(b_1 \mid a_2) & P(b_2 \mid a_2) & \cdots & P(b_s \mid a_2) \\ \vdots & \vdots & & \vdots \\ P(b_1 \mid a_r) & P(b_2 \mid a_r) & \cdots & P(b_s \mid a_r) \end{bmatrix}$$

这个矩阵被称为信道的转移概率矩阵或简称为信道矩阵。

3.2 信道疑义度

信道中存在随机性的干扰。信源输出的符号经信道传递，在这种随机性干扰的影响下，输出 Y 成为输入 X 的一个“干扰”变形。因此，在接收端观测信道的输出 Y 之后，收信者仍然不能够确定信源输出的是哪一个符号，对于信源 X，将仍然存在一定程度的不确定性。由第 2 章的讨论可知，在不借助于对信道输出 Y 的观测的条件下，关于信源 X 的先验不确定性由信息熵 $H(X)$ 来度量。因此，信源的信息熵 $H(X)$ 也称做先验熵。那么，在接收端观测到信道的输出 Y 之后，仍然存在的关于输入 X 的不确定性应该怎样度量呢？

首先，假设在信道的输出端观测到 $Y=b_j$(j 固定)的情况。

由信道的转移概率分布可知，当 $Y=b_j$ 时，信源输出 $X=a_i$ 的概率为条件概率：

$$P(b_j \mid a_i) \qquad i=1,2,\cdots,r$$

因此，在接收到 $Y=b_j$ 之后，我们关于 X 的后验不确定性为

$$H(X \mid Y=b_j)=\sum_{i=1}^{r} P(a_i \mid b_j)\log\frac{1}{P(a_i \mid b_j)} \qquad j=1,2,\cdots,s \tag{3.3}$$

$H(X|Y=b_j)$ 表示了接收到 $Y=b_j$ 时关于 X 的不确定性，称为接收到 $Y=b_j$ 时关于信源 X 的后验熵。

由于 $j=1, 2, \cdots, s$，$H(X|Y=b_j)$是一个伴随着随机变量 $Y=b_j$ 的发生而发生，并且与 $Y=b_j$ 有相同概率分布 $P(b_j)$的随机变量，因此我们需要一个确定的量，能够从总体上来度量在接收端观测到消息集合 Y 时关于信源 X 的平均后验不确定性。

为此，定义 $H(X|Y=b_j)$的数学期望为条件熵 $H(X|Y)$：

$$\begin{aligned} H(X \mid Y) &= E[H(X \mid Y=b_j)] = \sum_{j=1}^{s} P(b_j) H(X \mid Y=b_j) \\ &= \sum_{j=1}^{s} \sum_{i=1}^{r} P(b_j) P(a_i \mid b_j) \log \frac{1}{P(a_i \mid b_j)} \\ &= \sum_{j=1}^{s} \sum_{i=1}^{r} P(a_i, b_j) \log \frac{1}{P(a_i \mid b_j)} \end{aligned} \tag{3.4}$$

由于信道中存在随机性干扰，因此在对信道输出 Y 进行观测后，我们对信道的输入 X 仍具有某种程度的不确定性。条件熵 $H(X|Y)$反映了观测到 Y 之后对 X 仍然保留的不确定性。由于经过信息传递仍然存在的这种不确定性是信道中存在的随机性干扰，因此条件熵 $H(X|Y)$被称为信道疑义度。

如果信道中不存在随机性噪声，即 Y 与 X 有一一对应的确定关系，则在信道的输出端接收到符号集 Y 之后，便可以完全消除关于符号集 X 的不确定性，故信道疑义度 $H(X|Y)=0$。

如果信道中存在随机性干扰，则在信道的输出端接收到符号集 Y 之后，不能够完全消除关于信源 X 的不确定性，即仍然存在一定程度的剩余不确定性。但是，由第 2 章给出的条件熵的基本性质：

$$H(X \mid Y) \leqslant H(X)$$

可以知道，观测信道的输出 Y 之后，信道疑义度必小于等于信源的信息熵。因此，在统计平均意义下，对信道输出 Y 的观测对于减小关于信源 X 的不确定性总会有所帮助。

由下面例子，我们可以更加明确地了解信道疑义度 $H(X|Y)$满足的关系。

【例 3.1】　图 3－3 为一个二进制可抹信道。

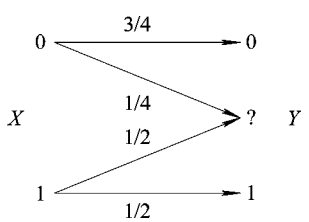

图 3－3　二进制可抹信道

已知输入随机变量 X 的概率空间为$[X, P]=\begin{bmatrix} 0 & 1 \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}$，输出随机变量$Y \in \{0, ?, 1\}$。

计算：$H(X|Y=0)$、$H(X|Y=?)$、$H(X|Y=1)$和 $H(X|Y)$。

解：给定二进制可抹信道的转移概率 $P(y|x)$可以列表表示(见表 3－1)。

表 3-1 $P(y|x)$

	0	?	1
0	$\frac{3}{4}$	$\frac{1}{4}$	0
1	0	$\frac{1}{2}$	$\frac{1}{2}$

信源的先验熵为

$$H(X)=H\left(\frac{2}{3},\frac{1}{3}\right)=0.9183 \quad \text{比特/符号}$$

依概率关系：

$$P(x,y)=P(y\mid x)P(x)$$

$$P(y)=\sum_{X}P(x,y)$$

$$P(x\mid y)=\frac{P(x,y)}{P(y)}$$

可以求出随机变量对(X,Y)的联合概率分布 $P(x,y)$和反向信道参数 $P(x|y)$(见表 3-2 和表 3-3)。

表 3-2 $P(x,y)$

	0	?	1
0	$\frac{1}{2}$	$\frac{1}{6}$	0
1	0	$\frac{1}{6}$	$\frac{1}{6}$
$P(y)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

表 3-3 $P(x|y)$

	0	?	1
0	1	$\frac{1}{2}$	0
1	0	$\frac{1}{2}$	1

于是有：

$$H(X\mid Y=0)=\sum_{X}P(x\mid Y=0)\log\frac{1}{P(x\mid Y=0)}=H(1,0)=0 \quad \text{比特/符号}$$

$$H(X\mid Y=1)=\sum_{X}P(x\mid Y=1)\log\frac{1}{P(x\mid Y=1)}=H(0,1)=0 \quad \text{比特/符号}$$

$$H(X\mid Y=?)=\sum_{X}P(x\mid Y=?)\log\frac{1}{P(x\mid Y=?)}=H\left(\frac{1}{2},\frac{1}{2}\right)=1 \quad \text{比特/符号}$$

计算结果表明，通过观测信道的输出了解信源的输出时，如果得到 $Y=0$ 或 $Y=1$，则对于信源 X 将不再具有不确定性；当观测到 $Y=?$ 时，对于 X 的后验不确定性 $H(X|Y=?)$ 比其先验不确定性 $H(X)$ 更大，即由先验不确定性 0.9183 比特/符号增大为 1 比特/符号。

由此可知，在接收到某一具体符号 $Y=b_j$ 时，可能对于消除关于信源 X 的不确定性有帮助，也可能不仅无帮助，反而会使这种不确定性进一步增大。但是应当明确，条件熵

$H(X|Y=b_j)$伴随着 $Y=b_j$ 的发生而发生，是一个与 $Y=b_j$ 同分布的随机变量。

对于一个通信系统，其传输信息的能力需要从总体上进行考察和度量。条件熵小于等于其无条件熵，即 $H(X|Y)\leqslant H(X)$的基本关系表明，在统计平均意义下，对信道输出 Y 进行观测对于减小或消除关于信源 X 的平均不确定性总会有所帮助。

在此例中，信道疑义度(后验熵)为

$$H(X\mid Y)=\sum_{j=1}^{3}P(y_j)H(X\mid Y=y_j)=\frac{1}{3}\approx 0.333 \quad \text{比特/符号}$$

而关于信源 X 的先验平均不确定(先验熵)为

$$H(X)=0.9183 \quad \text{比特/符号}$$

可见，观测信道的输出 Y 之后，仍然存在的关于信源 X 的平均不确定性减小了。

3.3　范诺不等式

信道疑义度 $H(X|Y)$度量了观测信道输出 Y 之后对信源 X 仍然保留的平均不确定性。范诺(Fano)不等式描述了信道疑义度与信息系统的错误概率之间的关系，是反映通信过程中信道疑义度的产生原因和取值大小的一个重要性质。

为了推出范诺不等式，先给出下面的引理。

引理 3.1　设 X、Y、Z 为随机变量，对 Z 的每一取值 z，定义

$$A(z)=\sum_{XY}P(y)P(z\mid x,\ y) \tag{3.5}$$

则有：

$$H(X\mid Y)\leqslant H(Z)+E[\log A(z)] \tag{3.6}$$

证明：由 $A(z)$的定义可知，$A(z)$、$\log A(z)$都只与随机变量 Z 有关，故 $A(z)$、$\log A(z)$与Z 有相同的概率分布，即概率分布均为 $P(z)$。

考察观测信道的输出 Y 之后，关于信源 X 的平均不确定性即条件熵为

$$H(X\mid Y)=\sum_{XY}P(x,\ y)\log\frac{1}{P(x\mid y)}$$

对于三维随机变量$(X,\ Y,\ Z)$，$(X,\ Y)$发生的边缘概率分布可以由三维随机变量$(X,\ Y,\ Z)$的联合概率分布 $P(x,\ y,\ z)$得到，即

$$P(x,\ y)=\sum_{Z}P(x,\ y,\ z) \tag{3.7}$$

因此

$$\begin{aligned}H(X\mid Y)&=\sum_{XY}P(x,\ y)\log\frac{1}{P(x\mid y)}\\&=\sum_{XY}\sum_{Z}P(x,\ y,\ z)\log\frac{1}{P(x\mid y)}\\&=\sum_{Z}P(z)\sum_{XY}\frac{P(x,\ y,\ z)}{P(z)}\log\frac{1}{P(x\mid y)}\end{aligned} \tag{3.8}$$

已知对数函数是上凸函数，并且$\dfrac{P(x,\ y,\ z)}{P(z)}=P(x,\ y|z)$是一组概率分布，对内和式应用 Jensen 不等式，便有：

$$\sum_{XY}\frac{P(x,y,z)}{P(z)}\log\frac{1}{P(x\mid y)}\leqslant\log\Big[\sum_{XY}\frac{1}{P(z)}\cdot\frac{P(x,y,z)}{P(x\mid y)}\Big]$$
$$=\log\Big[\frac{1}{P(z)}\sum_{XY}\cdot\frac{P(x,y,z)}{P(x\mid y)}\Big]$$
$$=\log\Big[\frac{1}{P(z)}\sum_{XY}\cdot\frac{P(x,y,z)}{P(x,y)}P(y)\Big] \tag{3.9}$$

因此

$$H(X\mid Y)\leqslant\sum_{Z}P(z)\log\Big[\frac{1}{P(z)}\sum_{XY}\frac{P(x,y,z)}{P(x,y)}\cdot P(y)\Big]$$
$$=\sum_{Z}P(z)\log\frac{1}{P(z)}+\sum_{Z}P(z)\log\sum_{XY}\frac{P(x,y,z)}{P(x,y)}\cdot P(y)$$
$$=H(Z)+\sum_{Z}P(z)\log A(z)$$
$$=H(Z)+E[\log A(z)] \tag{3.10}$$

定理 3.1(范诺不等式) 设 X 和 Y 是随机变量，X 和 Y 均取值于离散集合$\{a_1,a_2,\cdots,a_r\}$。令 $p_e=p(X\neq Y)$为错误概率，那么

$$H(X\mid Y)\leqslant H_2(p_e)+p_e\log(r-1) \tag{3.11}$$

证明：在 X 和 Y 有同样符号取值集合的通信系统中，我们感兴趣的是在接收端接收到的符号是否与信源输出的符号一致。为此，在接收到 Y 之后，我们定义随机变量 Z：

$$Z=\begin{cases}0 & X=Y\\ 1 & X\neq Y\end{cases} \tag{3.12}$$

由定理中给出的条件得到随机变量 Z 的概率分布为

$$\begin{cases}P(z=0)=1-p_e\\ P(z=1)=p_e\end{cases} \tag{3.13}$$

根据前面的引理，对于随机变量 Z 的每一取值 z，定义随机变量 Z 的函数：

$$A(z)=\sum_{XY}P(y)P(z\mid x,y)$$

并依据定理中的条件计算随机变量的函数 $A(z)$和 $\log A(z)$的取值。

当 $Z=0$ 时：

$$A(0)=\sum_{XY}P(y)P(z=0\mid x,y)=\sum_{Y}P(y)\sum_{X}P(z=0\mid x,y) \tag{3.14}$$

以 y 作为参变量，遍取 x 时：

$$\begin{cases}P(z=0\mid x=y)=1\\ P(z=0\mid x\neq y)=0\end{cases} \tag{3.15}$$

式(3.14)中的内和式：

$$\sum_{X}P(z=0\mid x,y)=P(z=0\mid x=y)=1 \tag{3.16}$$

故有：

$$A(0)=\sum_{Y}P(y)\sum_{X}P(z=0\mid x,y)=\sum_{Y}P(y)=1 \tag{3.17}$$

和

$$\log A(0)=\log 1=0$$

当 $Z=1$ 时：

$$A(1)=\sum_{Y}P(y)\sum_{X}P(z=1\mid x,\ y) \tag{3.18}$$

以 y 作为参变量，遍取 x 时：

$$\begin{cases}P(z=0\mid x=y)=0\\P(z=0\mid x\neq y)=1\end{cases} \tag{3.19}$$

在固定 y，改变 x 时，满足 $x\neq y$ 的 X 的取值有 $r-1$ 个，有：

$$\sum_{X}P(z=1\mid x,\ y)=(r-1)P(z=0\mid x\neq y)=r-1 \tag{3.20}$$

故有：

$$A(1)=\sum_{Y}P(y)\cdot(r-1)=r-1 \tag{3.21}$$

和

$$\log A(1)=\log(r-1)$$

由于 $A(z)$、$\log A(z)$ 是与 Z 同分布的随机变量，因此可以写出它们的概率空间为

$$[Z,\ P]=\begin{bmatrix}0 & 1\\1-p_e & p_e\end{bmatrix}$$

$$[A(z),\ P]=\begin{bmatrix}A(0) & A(1)\\1-p_e & p_e\end{bmatrix}=\begin{bmatrix}1 & r-1\\1-p_e & p_e\end{bmatrix}$$

$$[\log A(z),\ P]=\begin{bmatrix}\log 1 & \log(r-1)\\1-p_e & p_e\end{bmatrix}=\begin{bmatrix}0 & \log(r-1)\\1-p_e & p_e\end{bmatrix}$$

因而可以求出：

$$H(Z)=(1-p_e)\log\frac{1}{1-p_e}+p_e\log\frac{1}{p_e}=H_2(p_e) \tag{3.22}$$

$$\begin{aligned}E[\log A(z)]&=(1-p_e)\log A(0)+p_e\log A(1)\\&=(1-p_e)\log 1+p_e\log(r-1)\\&=p_e\log(r-1)\end{aligned} \tag{3.23}$$

将式(3.22)和式(3.23)的结果代入引理的结论中，便有：

$$H(X\mid Y)\leqslant H(Z)+E[\log A(z)]=H_2(p_e)+p_e\log(r-1)$$

范诺不等式是反映信道疑义度 $H(X|Y)$ 取值大小关系的一个重要定理，在信息系统分析中有明确意义和作用。

此时信道的输入和输出有相同的符号集合，即 X，$Y\in(a_1,\ a_2,\ \cdots,\ a_r)$，而系统中的错误概率为：$p_e=P(X\neq Y)$。范诺不等式所指出的关系表明，在观测信道的输出 Y 之后，关于随机变量 X 仍然存在的不确定性由两部分组成。

(1) 首先，需要判断 X 是否与 Y 相同。

在范诺不等式的证明中，定义了随机变量：

$$Z=\begin{cases}0 & X=Y\\1 & X\neq Y\end{cases}$$

因此，判断 X 是否与 Y 相同等效于确定随机变量 Z 为 0 还是为 1，即信息传输过程中是否发生了错误。

由于已知信道中发生传输错误的概率为 p_e，因此由输出“对”或“错”所构成的信源的熵为 $H(Z)=H_2(p_e)$，即关于判断“X 是否与 Y 相同”的平均不确定性为 $H_2(p_e)$。

若可判断得出 $X=Y$，则由 Y 可知 X，消除了关于 X 的不确定性。

(2) 若已知 X 与 Y 不相同，则需要确定 X 究竟取何种符号。

如果由第一步判断已得知 $X\neq Y$，则由于 X 仍有 $r-1$ 种可能的取值，因此，接收者对于信源 X 究竟输出了哪一种符号仍然具有不确定性。但是，由熵的极值性可以知道，在已知 $X\neq Y$ 的条件下，关于 X 究竟取哪一种符号的不确定性不会超过 $\log(r-1)$。同时，由于 $X\neq Y$ 发生的概率为 p_e，因此“在 $X\neq Y$ 时，X 究竟取何种符号”的平均不确定性不会超过 $p_e\log(r-1)$。

可见，范诺不等式的确指出了信息系统中信道疑义度 $H(X|Y)$ 取值的上限，即满足：

$$H(X\mid Y)\leqslant H_2(p_e)+p_e\log(r-1)$$

3.4 互信息的定义

在一般的信息传输过程中，收信者获取信源 X 输出的信息，是通过信源输出的符号在信道中传递，观测信道的输出来实现的(见图 3-4)。

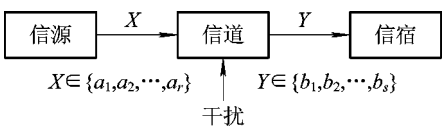

图 3-4 一般的通信系统

由前面的讨论我们已经知道，在收到信源输出消息前对信源 X 的输出具有不确定性。$H(X)$ 表示关于信源 X 的先验不确定性，而 $H(X|Y)$ 则表示对信道输出 Y 观测后关于 X 的后验不确定性。因为 $H(X|Y)\leqslant H(X)$，所以通过信息的传递和对 Y 的观测，接收者对 X 的不确定性由 $H(X)$ 减小为 $H(X|Y)$，表明通过信息系统的信息传递，接收者由 Y 获得了一些关于信源 X 的信息。

在对各类信息系统进行分析和设计中，需要对信源 X 的表示、传输信息的能力等方面进行定量的分析和计算。本节我们将从信息的传递使得关于 X 的不确定性减小入手，建立对一般通信系统中信息传递现象定量描述的基础，引出互信息的定义并深入讨论互信息所具有的性质。

3.4.1 符号间的互信息

设有离散无记忆信源：

$$[X,\ P]=\begin{bmatrix}a_1 & a_2 & \cdots & a_n\\ p_1 & p_2 & \cdots & p_n\end{bmatrix}$$

经离散无记忆信道传输，信道输出符号集为

$$Y\in\{b_1,\ b_2,\ \cdots,\ b_s\}$$

首先考察接收到单个符号 $Y=b_j$ 时对了解信源符号 $X=a_i$ 的帮助，即分析信息系统中单个符号之间的信息传递关系。

如果信道无噪无损，则 X 与 Y 一一对应。由于 Y 与 X 有确定的对应关系，因此如果接收者收到 $Y=b_j$，则可以确定信源的输出符号为 $X=a_i$，即由 $Y=b_j$ 获得了信源输出 $X=a_i$ 的信息。

对于某一信源符号 $X=a_i$，我们具有先验不确定性，即单信源符号的自信息。

已知 $X=a_i$ 发生的概率 $P(a_i)$ 对符号 a_i 的先验不确定性为符号 a_i 的自信息，即

$$I(a_i)=\log\frac{1}{P(a_i)}\qquad i=1,2,\cdots,r \tag{3.24}$$

然而，一般的信道中存在着随机性干扰。在随机干扰的作用下，信道的输出符号 $Y=b_j$ 成为信源输出符号 $X=a_i$ 的一个干扰变型。因此，接收到 $Y=b_j$ 之后，对 $X=a_i$ 仍然存在不确定性。

已知 $Y=b_j$ 时，$X=a_i$ 发生的概率为后验概率 $P(a_i|b_j)$。接收到 $Y=b_j$ 后，接收者关于 $X=a_i$ 的不确定性为 $P(a_i|b_j)$ 的函数，即

$$\log\frac{1}{P\left(\frac{a_i}{b_j}\right)}\qquad i=1,2,\cdots,r;\ j=1,2,\cdots,s$$

可见，接收到 $Y=b_j$ 后，虽然对 $X=a_i$ 仍然具有不确定性，但是，通过对 Y 的观测，关于 $X=a_i$ 的不确定性发生了变化。这种变化表明接收者从接收到 $Y=b_j$ 这一事件中得到了关于 $X=a_i$ 的某些信息。

于是，我们给出符号间互信息的定义：

$X=a_i$ 的先验不确定性与其后验不确定性之差为符号 a_i、b_j 之间的互信息，记做 $I(a_i;b_j)$。

$$\begin{aligned}I(x_i;y_j)&=\log\frac{1}{P(x_i)}-\log\frac{1}{P(x_i\mid y_j)}\\&=\log\frac{P(x_i\mid y_j)}{P(x_i)}\quad i=1,2,\cdots,r;\ j=1,2,\cdots,s\end{aligned} \tag{3.25}$$

互信息 $I(a_i;b_j)$ 代表了由接收到的符号 $Y=b_j$ 获得的关于 $X=a_i$ 的信息量。

符号间的互信息 $I(a_i;b_j)$ 有下面的基本关系。

1. 对称性

$$I(a_i;b_j)=I(b_j;a_i) \tag{3.26}$$

证明：

$$\begin{aligned}I(a_i;b_j)&=\log\frac{P(a_i\mid b_j)}{P(a_i)}=\log\frac{P(a_i\mid b_j)P(b_j)}{P(a_i)P(b_j)}\\&=\log\frac{\frac{P(a_i,b_j)}{P(a_i)}}{P(b_j)}=\log\frac{P(b_j\mid a_i)}{P(b_j)}\\&=I(b_j;a_i)\end{aligned}$$

对称性表明，由 $Y=b_j$ 提供的关于 $X=a_i$ 的信息量等于由 $X=a_i$ 提供的关于 $Y=b_j$ 的信息量。

$I(a_i;b_j)$ 与 $I(b_j;a_i)$ 的这种对称性也称为交互性，因此 $I(a_i;b_j)$ 被称为互信息量。

2. 符号间的互信息量的取值关系

对于信源符号 $X=a_i$，其先验不确定性取决于先验(已知)概率分布 $P(a_i)$。观测到 $Y=b_j$

之后，关于 $X=a_i$ 的后验不确定性则取决于信道的统计特性，即后验概率 $P(a_i|b_j)$。由于 $X=a_i$、$Y=b_j$ 之间的互信息是关于 $X=a_i$ 的先验不确定性和后验不确定性之差，因此实际上此互信息量的取值也取决于信道的特性，即 $X=a_i$ 的后验概率。

在不同的信道中，因其统计特性即后验概率 $P(a_i|b_j)$不同，故 $X=a_i$ 与 $Y=b_j$ 之间的互信息的取值将有不同的特点。

(1) 对于无噪无损信道，$X=a_i$ 与 $Y=b_j$ 之间为一一对应关系，即 $P(a_i|b_j)=1$。此时可由 $Y=b_j$ 唯一地确定信源输出 $X=a_i$，获得 a_i 的自信息量 $I(a_i)$。

(2) 对于 $X=a_i$ 与 $Y=b_j$ 统计独立的信道，因为 $P(a_i, b_j)=P(a_i)P(b_j)$，所以有

$$I(a_i; b_j) = \log \frac{P(a_i \mid b_j)}{P(a_i)} = \log \frac{P(a_i, b_j)}{P(a_i)P(b_j)} = \log 1 = 0 \tag{3.27}$$

可见，由信道的输出 $Y=b_j$ 不能够获得关于 $X=a_i$ 的任何信息，$I(a_i; b_j)=0$。

(3) 对于一般的信道，后验概率 $P(a_i|b_j)$满足：

$$0 \leqslant P(a_i \mid b_j) \leqslant 1 \qquad i = 1, 2, \cdots, r; \ j = 1, 2, \cdots, s$$

此时互信息的取值大小显然取决于 $P(a_i|b_j)$与 $P(a_i)$的关系。

当 $P(a_i) \leqslant P(a_i|b_j) < 1$ 时，通过对信道输出 b_j 的观测，接收者对 $X=a_i$ 的不确定性减小了，即由 b_j 中获得了关于 a_i 的一些信息，此时 $I(a_i; b_j) \geqslant 0$。

如果 $0 < P(a_i|b_j) < P(a_i)$，则表明观测到 $Y=b_j$ 后，对信源是否发出 $X=a_i$ 的不确定性不仅没有减小，反而增大了，即观测系统的输出后，由 $Y=b_j$ 消除的关于 $X=a_i$ 的不确定性为一负值，有 $I(a_i; b_j) < 0$。

因此，在已知某信源符号 $X=a_i$ 的概率关系(先验概率)时，通过观测 $Y=b_j$ 所获得的互信息 $I(a_i; b_j)$的取值可正可负。

3. 任何两个事件之间的互信息量不大于其中任一事件的自信息量

$$\begin{cases} I(a_i; b_j) \leqslant I(a_i) \\ I(a_i; b_j) \leqslant I(b_j) \end{cases} \qquad i = 1, 2, \cdots, r; \ j = 1, 2, \cdots, s \tag{3.28}$$

证明：

由于

$$P(a_i \mid b_j) \leqslant 1 \qquad i = 1, 2, \cdots, r; \ j = 1, 2, \cdots, s$$

因此

$$I(a_i; b_j) = \log \frac{P(a_i \mid b_j)}{P(a_i)} \leqslant \log \frac{1}{P(a_i)} = I(a_i)$$

同理由

$$P(b_j \mid a_i) \leqslant 1 \quad i = 1, 2, \cdots, r; \ j = 1, 2, \cdots, s$$

得出

$$I(a_i; b_j) = \log \frac{P(b_j \mid a_i)}{P(b_j)} \leqslant \log \frac{1}{P(b_j)} = I(b_j)$$

证毕。

【例 3.2】 图 3-5 所示为二进制可抹信道。计算符号之间的互信息 $I(x=0; y=0)$ 和 $I(x=0; y=?)$。

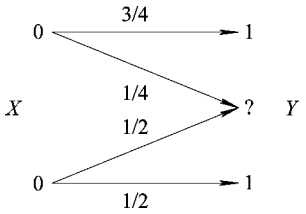

图 3-5　二进制可抹信道

解：由例 3.1 已知 $P(x=0)=\frac{2}{3}$，$P(x=1)=\frac{1}{3}$，**已求出给定二进制可抹信道的信道疑义度。**

根据符号间的互信息的定义，可以求得：

$$I(x=0;\ y=0)=\log\frac{P(x=0\mid y=0)}{P(x=0)}$$

$$=\log\left(\frac{1}{\frac{2}{3}}\right)=\log\left(\frac{3}{2}\right)\approx 0.585 \quad \text{比特}$$

因为 $y=0$ **与** $x=0$ **具有确定的概率转移关系，所以接收到** $y=0$ **后便可以消除信源发出** $x=0$ **的不确定性。因此，由** $y=0$ **所获得** $x=0$ **的信息量为** $x=0$ **的自信息** $I(x=0)$，**而**

$$I(x=0;\ y=?)=\log\frac{P(x=0\mid y=?)}{P(x=0)}$$

$$=\log\left(\frac{\frac{1}{2}}{\frac{2}{3}}\right)=\log\frac{3}{4}\approx -0.415 \quad \text{比特}$$

可见，由于信道中存在干扰，因此在接收到 $y=?$ **时关于** $x=0$ **的不确定性更大，即由** $y=?$ **所获得的关于** $x=0$ **的信息量为负值。**

由二进制可抹信道的后验概率可知，$P(x=0|y=1)=0$，**如果观测到** $y=1$，**则信源不可能发出** 0。**因此无需考虑** $y=1$ **与** $x=0$ **之间的互信息。**

3.4.2　平均互信息

符号间的互信息 $I(a_i;\ b_j)$**描述了在对通信系统的输出进行观测时，由** $Y=b_j$ **得到的关于** $X=a_i$ **的信息量。这一符号间的互信息以单个符号的不确定性的变化(改变量)给出了一种信息的度量方法。但是由于** $X=a_i$ **和** $Y=b_j$ **均为随机变量，因此由** $Y=b_j$ **所获得的关于** $X=a_i$ **的互信息**

$$I(a_i;\ b_j)\qquad i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots,\ s$$

是一个伴随着 $X=a_i$ **与** $Y=b_j$ **同时发生而发生的随机量，且发生的概率为** $X=a_i$ **和** $Y=b_j$ **构成的联合事件发生的联合概率：**

$$P(a_i,\ b_j)\qquad i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots,\ s$$

显然，使用这样的按照一定概率而发生的随机变量来度量一个信息系统传输信息的能

力具有明显的局限性。

对于通信系统总体而言，其传输信息的特性和能力应当从系统的总体上加以度量，即关于信息量的描述和度量不应当是一个随机变量，而应当是能够从信息系统总体进行描述的一个确定的量。因此，关于信息的描述与度量只能在统计平均的意义下进行测度。

定义 3.1 在联合集 X，Y 上，$X=a_i$ 与 $Y=b_j$ 之间的互信息 $I(a_i;\ b_j)$ 在联合概率空间：

$$P(a_i,\ b_j) \qquad i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots,\ s$$

中的统计平均值(期望值)称为平均互信息，记做 $I(X;\ Y)$。

$$\begin{aligned} I(X;\ Y) &= E[I(a_i;\ b_j)] = \sum_{i=1}^{r}\sum_{j=1}^{s}P(a_i,\ b_j)I(a_i;\ b_j) \\ &= \sum_{i=1}^{r}\sum_{j=1}^{s}P(a_i,\ b_j)\log\frac{P(a_i\mid b_j)}{P(a_i)} \\ &= \sum_{i=1}^{r}\sum_{j=1}^{s}P(a_i,\ b_j)\log\frac{1}{P(a_i)} - \sum_{i=1}^{r}\sum_{j=1}^{s}P(a_i,\ b_j)\log\frac{1}{P(a_i\mid b_j)} \end{aligned}$$

即

$$I(X;\ Y) = H(X) - H(X\mid Y)$$

此式表明，“Y 已知”这一事件使我们关于 X 的不确定性减少了 $I(X;\ Y)$，即通过通信系统的传输，接收到符号集 Y 后平均每个符号获得关于 X 的信息量为 $I(X;\ Y)$。

因此，$I(X;\ Y)$ 称为 X 与 Y 之间的平均互信息，简称为互信息。

在下面的讨论中，有时将平均互信息表示为

$$I(X;\ Y) = \sum_{XY}P(x,\ y)\log\frac{P(x\mid y)}{P(x)}$$

此时，x 和 y 分别表示随机变量 X 和 Y 的一个取值符号，而 $\sum_{XY}$ 表示对随机变量 X 和 Y 的全空间求和。

【例 3.3】 接续例 3.1 计算所给信源和二进制可抹信道构成的系统中，由 Y 所得到的关于 X 的平均互信息。

解：前面已求出各种概率的关系，如表 3-4 和表 3-5 所示。

表 3-4 $P(x,\ y)$，$P(x)$

X \ Y	0	?	1	$P(x)$
0	$\frac{1}{2}$	$\frac{1}{6}$	0	$\frac{2}{3}$
1	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
$P(y)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$	

表 3-5 $P(x|y)$

X \ Y	0	?	1
0	1	$\frac{1}{2}$	0
1	0	$\frac{1}{2}$	1

$$H(X)=0.9138 \quad \text{比特/符号}$$

$$H(X|Y)=\frac{1}{3}=0.3333 \quad \text{比特/符号}$$

$$I(X;Y)=\sum_{i=1}^{2}\sum_{j=1}^{3}P(a_i,b_j)\log\frac{P(a_i\mid b_j)}{P(a_i)}$$

$$=\frac{1}{2}\log\left(\frac{1}{\frac{2}{3}}\right)+\frac{1}{6}\log\left(\frac{\frac{1}{2}}{\frac{2}{3}}\right)+\frac{1}{6}\log\left(\frac{\frac{1}{2}}{\frac{1}{3}}\right)+\frac{1}{6}\log\left(\frac{1}{\frac{1}{3}}\right)$$

$$=\frac{1}{2}\log\left(\frac{3}{2}\right)+\frac{1}{6}\log\left(\frac{3}{4}\right)+\frac{1}{6}\log\left(\frac{3}{2}\right)+\frac{1}{6}\log 3$$

$$=\left(\frac{1}{2}+\frac{3}{6}\right)\log 3-\frac{1}{2}-\frac{2}{6}-\frac{1}{6}$$

$$=\log 3-1\approx 0.585\quad \text{比特/符号}$$

亦可：

$$I(X;Y)=H(X)-H(X\mid Y)$$

$$=0.9183-0.3333\approx 0.585\quad \text{比特/符号}$$

由前面的计算结果知道，如果在接收端观测到 $y=?$，则由 $y=?$ 获得的关于 $x=1$ 的互信息为负值。此处的计算表明，对于整个系统而言，所得的平均互信息是大于 0 的，即对符号集 Y 的观测总会使我们对 X 的了解有所帮助。

3.5　平均互信息的基本性质

平均互信息具有下面一些基本性质。

1. 对称性

$$I(X;Y)=I(Y;X)\tag{3.29}$$

证明：

$$I(X;Y)=\sum_{XY}P(x,y)\log\frac{P(x\mid y)}{P(x)}\frac{P(y)}{P(y)}$$

$$=\sum_{XY}P(x,y)\log\frac{P(x,y)}{P(x)P(y)}$$

$$=\sum_{XY}P(x,y)\log\frac{P(y\mid x)}{P(y)}=I(Y;X)$$

因为具有对称性，所以在下面的讨论中我们主要考虑 $I(X;Y)$。

2. 非负性

$$I(X;Y)\geqslant 0\tag{3.30}$$

当且仅当 X、Y 统计独立时等式成立。

证明：由定义式：

$$I(X;Y)=\sum_{XY}P(x,y)\log\frac{P(x\mid y)}{P(x)}\frac{P(y)}{P(y)}=\sum_{XY}P(x,y)\log\frac{P(x,y)}{P(x)P(y)}$$

当 X、Y 统计独立时：

$$P(x,y)=P(x)P(y)$$

故有：

$$I(X;Y)=\sum_{XY}P(x,y)\log 1=0\ (\text{等式成立条件})$$

当 X、Y 非统计独立时，有

$$-I(X;Y)=\sum_{XY}P(x,y)\log\frac{P(x)P(y)}{P(x,y)}$$

因为对数函数是上凸函数，所以应用颜森不等式，有

$$\begin{aligned}-I(X;Y)&\leqslant\log\Big[\sum_{XY}P(x,y)\frac{P(x)P(y)}{P(x,y)}\Big]\\&=\log\Big[\sum_{X}P(x)\sum_{Y}P(y)\Big]=\log 1=0\end{aligned}$$

所以，$I(X;Y)\geqslant 0$ 当且仅当 X、Y 统计独立时等式成立。

证毕。

由关于符号间的互信息 $I(a_i;b_j)$ 的讨论我们知道，$I(a_i;b_j)$ 的取值可正可负，即由于信道中干扰的影响，由某一符号 $Y=b_j$ 获得的关于 $X=a_i$ 的信息有可能是负数。然而，在对全空间 (X,Y) 作统计平均后，由一个通信系统的输出 Y 获得的关于 X 的平均互信息不会为负值。这就是说，从统计平均的意义上讲，观测信道的输出 Y 总会对消除关于信源 X 的不确定性有帮助，即由 Y 总会获得一些关于 X 的信息，而不会损失信息。只有当信道的输入 X 和输出 Y 是统计独立时，由信道的输出 Y 才不能得到关于信源 X 的任何信息。

3. 极值性

$$I(X;Y)\leqslant H(X)\tag{3.31}$$

因为 $H(X|Y)\geqslant 0$，所以有 $I(X;Y)=H(X)-H(X|Y)\leqslant H(X)$。

只有在无损信道中，即 $H(X|Y)=0$ 时，$I(X;Y)$ 才能达到其最大值 $H(X)$（等式成立条件）。

4. 平均互信息和各类熵的关系

平均互信息 $I(X;Y)$ 和条件熵 $H(X|Y)$、$H(Y|X)$、联合熵 $H(X,Y)$ 有如下关系：

$$\begin{aligned}I(X;Y)&=H(X)-H(X\mid Y)\\&=H(Y)-H(Y\mid X)\\&=H(X)+H(Y)-H(X,Y)\end{aligned}$$

这些关系的证明可直接由 $I(X;Y)$ 的定义式得到。例如：

$$\begin{aligned}I(X;Y)&=\sum_{XY}P(x,y)\log\frac{P(x\mid y)}{P(x)}\frac{P(y)}{P(y)}=\sum_{XY}P(x,y)\log\frac{P(x,y)}{P(x)P(y)}\\&=-\sum_{XY}P(x,y)\log P(x)-\sum_{XY}P(x,y)\log P(y)+\sum_{XY}P(x,y)\log P(x,y)\\&=-\sum_{Y}P(x,y)\log P(x)-\sum_{X}P(x,y)\log P(y)+\sum_{XY}P(x,y)\log P(x,y)\\&=H(X)+H(Y)-H(X,Y)\end{aligned}$$

平均互信息和各类熵之间的关系可以由图 3-6 所示的文氏图表示出来。

在图 3-6 中，以 X 为中心的圆的面积表示关于 X 的不确定性即熵 $H(X)$；以 Y 为中心的圆的面积表示关于 Y 的不确定性即熵 $H(Y)$；两圆所包围的关于联合事件 X、Y 的不确定性即 X、Y 的联合熵 $H(X,Y)$；两圆未重叠的部分分别表示在已知某一随机变量时的不确定性，即条件熵 $H(X|Y)$ 和 $H(Y|X)$；这两个圆相互重叠的阴影区则表示出了由一个

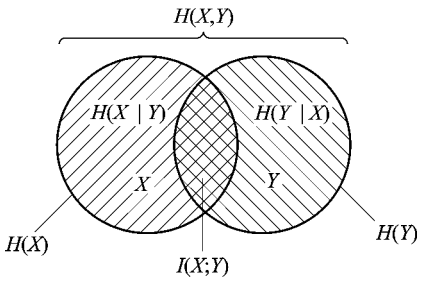

图 3－6　文氏图

随机变量所获得的关于另一个随机变量的信息，即互信息。

文氏图不仅从几何关系上形象地表示出了熵、条件熵、联合熵和互信息之间的关系，也有助于我们对这些概念及它们之间的关系的物理意义的理解。

3.6　平均互信息的凸性

由前面的讨论可知，通过信息传递，我们由信道的输出 Y 获得的关于信源输出 X 的信息为平均互信息 $I(X;Y)$。在图 3－7 所示的通信系统中，如果改变信源(即改变 $P(x)$)，则由 Y 所得到的关于 X 的互信息 $I(X;Y)$也将随之改变。同理，如果改变信道(改变 $P(y|x)$)，那么由 Y 所得到的关于 X 的平均互信息 $I(X;Y)$也将随之改变。

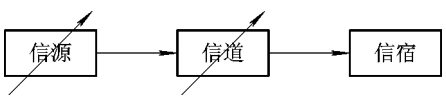

图 3－7　简单的通信系统

由平均互信息 $I(X;Y)$的定义式可以明显地看到这一结论。由于

$$I(X;Y)=\sum_X\sum_Y P(x)P(y\mid x)\log\frac{P(y\mid x)}{P(y)}$$

其中，$P(y)=\sum_X P(x)P(y\mid x)$，因此当信源的概率分布 $P(x)$改变时，或者是信道的转移概率改变时，平均互信息 $I(X;Y)$将随之改变，即平均互信息 $I(X;Y)$是信源概率分布 $P(x)$和信道转移概率分布 $P(y|x)$的函数。这种函数关系可以记做：

$$I(X;Y)=I[P(x);P(y\mid x)] \tag{3.32}$$

由于信源和信道是构成通信系统的主要部分，因此在研究通信系统中的信息传递现象时，深入分析平均互信息 $I(X;Y)$与信源概率分布 $P(x)$和信道转移概率分布$P(y|x)$的依赖关系是十分有意义的。

为此，我们给出平均互信息 $I(X;Y)$与信源概率分布 $P(x)$和信道转移概率分布 $P(y|x)$之间的函数关系所具有的两个重要性质。

定理 3.2　平均互信息 $I(X;Y)$是信源概率分布 $P(x)$的上凸函数。

证明：此处我们考察的是互信息 $I(X;Y)$随信源概率分布 $P(x)$改变时的函数特性。

为了使分析较简明，我们假设给定一个信道(即信道的转移概率 $P(y|x)$固定不变)，而信源的概率分布 $P(x)$是可变的。此时相当于将一个可变信源发出的符号通过一个固定

信道进行传输(如图 3-8 所示的物理模型)。

图 3-8 信源可变、信道固定的通信系统

此时,平均互信息仅与信源概率分布 $P(x)$ 呈函数关系:

$$I(X;Y)=I[P(x)] \tag{3.33}$$

为了证明此时平均互信息函数与信源概率分布 $P(x)$ 的上凸性关系,我们首先假设已给定两种不同的信源概率分布 $P_1(x)$ 和 $P_2(x)$,并使用给定信道分别传输两种不同信源的输出符号。

在这两种不同的信源概率分布下,可以分别得到联合事件 X、Y 的联合概率分布和信道输出 Y 的边缘概率分布:

$$P_1(x,y)=P_1(x)P(y\mid x) \qquad P_1(y)=\sum_X P_1(x)P(y\mid x)$$

$$P_2(x,y)=P_2(x)P(y\mid x) \qquad P_2(y)=\sum_X P_2(x)P(y\mid x)$$

它们均满足概率分布率的条件,即

$$0\leqslant P_1(x,y),P_2(x,y),P_1(y),P_2(y)\leqslant 1$$

$$\sum_{X,Y}P_1(x,y)=\sum_{X,Y}P_2(x,y)=\sum_Y P_1(y)=\sum_Y P_2(y)=1$$

对应于两种信源概率分布 $P_1(x)$ 和 $P_2(x)$,可以计算得到相应的平均互信息 $I(X;Y)$,即

$$I[P_1(x)]=\sum_X\sum_Y P_1(x)P(y\mid x)\log\frac{P(y\mid x)}{P_1(y)}=\sum_X\sum_Y P_1(x,y)\log\frac{P(y\mid x)}{P_1(y)}$$

$$I[P_2(x)]=\sum_X\sum_Y P_2(x)P(y\mid x)\log\frac{P(y\mid x)}{P_2(y)}=\sum_X\sum_Y P_2(x,y)\log\frac{P(y\mid x)}{P_2(y)}$$

为了分析平均互信息 $I(X;Y)$ 与 $P(x)$ 的凸性关系,由两种已知的信源概率 $P_1(x)$ 和 $P_2(x)$,通过线性组合构造第三种信源概率分布 $P(x)$。

令 $0\leqslant\alpha,\beta\leqslant 1$,$\alpha+\beta=1$,组合得到:

$$P(x)=\alpha P_1(x)+\beta P_2(x)$$

已知 $P_1(x)$、$P_2(x)$ 是概率分布,它们满足概率分布率:

$$\begin{cases}0\leqslant P_1(x),P_2(x)\leqslant 1\\ \sum_X P_1(x)=\sum_X P_2(x)=1\end{cases}$$

则有:

$$0\leqslant\alpha P_1(x)+\beta P_2(x)\leqslant 1$$

和

$$\sum_X(\alpha P_1(x)+\beta P_2(x))=\alpha\sum_X P_1(x)+\beta\sum_X P_2(x)=\alpha+\beta=1$$

可知,对于由 $P_1(x)$ 和 $P_2(x)$ 的线性组合构造的第三种信源 $P(x)$ 确为一组概率分布。

使用给定信道传输第三种信源(组合信源)输出的符号,可以求出此时联合事件 X、Y 的联合概率分布:

$$\begin{aligned}P(x,\ y) &= P(x)P(y \mid x)\\ &= \alpha P_1(x)P(y \mid x) + \beta P_2(x)P(y \mid x)\\ &= \alpha P_1(x,\ y) + \beta P_2(x,\ y)\end{aligned}$$

和随机变量 Y 的边缘概率分布：

$$P(y) = \sum_X P(x)P(y \mid x)$$

显然，$P(x,\ y)$和 $P(y)$满足：

$$0 \leqslant P(x,\ y),\ P(y) \leqslant 1$$

$$\begin{aligned}\sum_{XY} P(x,\ y) &= \sum_{XY} \alpha P_1(x,\ y) + \sum_{XY} \beta P_2(x,\ y)\\ &= \alpha \sum_{XY} P_1(x,\ y) + \beta \sum_{XY} P_2(x,\ y) = \alpha + \beta = 1\end{aligned}$$

$$\sum_Y P(y) = \sum_Y \sum_X P(x)P(y \mid x) = \sum_X P(x) \sum_Y P(y \mid x) = 1$$

因此，使用给定信道（$P(y|x)$固定）传输第三种信源（概率分布 $P(x)$）输出的符号时，系统中的平均互信息 $I(X;\ Y)$为

$$I[P(x)] = \sum_X \sum_Y P(x)P(y \mid x)\log \frac{P(y \mid x)}{P(y)} = \sum_X \sum_Y P(x,\ y)\log \frac{P(y \mid x)}{P(y)}$$

如果可以证明上述平均互信息之间满足：

$$\alpha I[P_1(x)] + \beta I[P_2(x)] \leqslant I[\alpha P_1(x) + \beta P_2(x)] = I[P(x)] \tag{3.34}$$

便可知平均互信息 $I(X;\ Y)$是信源概率分布 $P(x)$的上凸函数。

因为

$$\begin{aligned}&\alpha I[P_1(x)] + \beta I[P_2(x)] - I[P(x)]\\ &= \sum_{XY} \alpha P_1(x,\ y)\log \frac{P(y \mid x)}{P_1(y)} + \sum_{XY} \beta P_2(x,\ y)\log \frac{P(y \mid x)}{P_2(y)}\\ &\quad - \sum_{XY} (\alpha P_1(x,\ y) + \beta P_2(x,\ y))\log \frac{P(y \mid x)}{P(y)}\\ &= \alpha \sum_{XY} P_1(x,\ y)\log \frac{P(y \mid x)}{P_1(y)} \frac{P(y)}{P(y \mid x)} + \beta \sum_{XY} P_2(x,\ y)\log \frac{P(y \mid x)}{P_2(y)} \frac{P(y)}{P(y \mid x)}\\ &= \alpha \sum_{XY} P_1(x,\ y)\log \frac{P(y)}{P_1(y)} + \beta \sum_{XY} P_2(x,\ y)\log \frac{P(y)}{P_2(y)}\end{aligned}$$

已知对数函数是上凸函数，应用颜森不等式，则

$$\begin{aligned}&\alpha \sum_{XY} P_1(x,\ y)\log \frac{P(y)}{P_1(y)} + \beta \sum_{XY} P_2(x,\ y)\log \frac{P(y)}{P_2(y)}\\ &\leqslant \alpha \log\Big(\sum_{XY} P_1(x,\ y) \frac{P(y)}{P_1(y)}\Big) + \beta \log\Big(\sum_{XY} P_2(x,\ y) \frac{P(y)}{P_2(y)}\Big)\\ &= \alpha \log\Big(\sum_Y \frac{P(y)}{P_1(y)} \sum_X P_1(x,\ y)\Big) + \beta \log\Big(\sum_Y \frac{P(y)}{P_2(y)} \sum_X P_2(x,\ y)\Big)\\ &= \alpha \log\Big(\sum_Y \frac{P(y)}{P_1(y)} P_1(y)\Big) + \beta \log\Big(\sum_Y \frac{P(y)}{P_2(y)} P_2(y)\Big)\\ &= \alpha \log 1 + \beta \log 1 = 0\end{aligned}$$

因此式(3.34)所表示的互信息之间的关系成立，即对于

$$0 \leqslant \alpha, \beta \leqslant 1, \alpha + \beta = 1$$

满足

$$\alpha I[P_1(x)] + \beta I[P_2(x)] \leqslant I[\alpha P_1(x) + \beta P_2(x)]$$

所以，平均互信息 $I(X; Y)$是信源概率分布 $P(x)$的上凸函数。

证毕。

由平均互信息 $I(X; Y)$的定义知道，平均互信息 $I(X; Y)$是信源概率分布 $P(x)$的函数。当使用某一固定信道(给定信道转移概率分布 $P(y|x)$)传输不同信源(信源概率分布 $P(x)$不同)输出的符号时，在系统的输出端获得的平均信息量是不同的。定理 3.2 指出平均互信息 $I(X; Y)$与信源概率分布 $P(x)$之间的函数呈上凸性关系。这一反映平均互信息 $I(X; Y)$与信源概率分布 $P(x)$之间关系的重要性质表明，对于每一个给定的信道，一定存在一种信源，使得在信道的输出端获得的关于信源的信息量 $I(X; Y)$达到其最大值。

定理 3.3 平均互信息 $I(X; Y)$是信道转移概率 $P(y|x)$的下凸函数。

证明：此处考察的是平均互信息 $I(X; Y)$随信道转移概率分布 $P(y|x)$改变时的函数特性。为此，假设给定一个信源(即信源的概率分布 $P(x)$固定不变)和一个转移概率分布 $P(y|x)$可变的实验信道。将给定信源($P(x)$固定)的输出符号通过 $P(y|x)$可变的实验信道传输。图 3-9 给出了此时的信息系统模型。

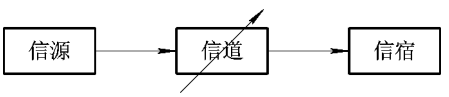

图 3-9 信源固定、信道可变的通信系统

由于 $P(x)$是固定不变的，因此此时系统中的平均互信息 $I(X; Y)$仅与信道转移概率分布 $P(y|x)$呈函数关系，即可简单记做：

$$I(X; Y) = I[P(y \mid x)] \tag{3.35}$$

设给定概率分布为 $P(x)$的信源 X。

首先给定转移概率分布分别为 $P_1(y|x)$和 $P_2(y|x)$的两种已知信道。它们的信道参数分别满足概率分布率的条件，即

$$0 \leqslant P_1(y \mid x), P_2(y \mid x) \leqslant 1$$

$$\sum_Y P_1(y \mid x) = \sum_Y P_2(y \mid x) = 1$$

将给定信源 X 输出的符号经这两种不同的信道构成信息传输系统。对应于两种不同的信息传输系统，输入、输出随机变量 X、Y 所构成的联合事件(X, Y)的联合概率分布分别为

$$P_1(x, y) = P(x)P_1(y \mid x)$$

$$P_2(x, y) = P(x)P_2(y \mid x)$$

相应的边缘概率分布为

$$P_1(y) = \sum_X P(x)P_1(y \mid x)$$

$$P_2(y) = \sum_X P(x)P_2(y \mid x)$$

条件概率分布为

$$P_1(x \mid y) = \frac{P(x)P_1(y \mid x)}{P(x)}$$

$$P_2(x \mid y) = \frac{P(x)P_2(y \mid x)}{P(x)}$$

对应于两种不同的信道 $P_1(y|x)$和 $P_2(y|x)$，可以计算得到相应的平均互信息 $I[P_1(y|x)]$和 $I[P_2(y|x)]$，即

$$I[P_1(y \mid x)] = \sum_X \sum_Y P(x)P_1(y \mid x)\log \frac{P_1(x \mid y)}{P(x)}$$

$$= \sum_X \sum_Y P_1(x, y)\log \frac{P_1(x \mid y)}{P(x)}$$

$$I[P_2(y \mid x)] = \sum_X \sum_Y P(x)P_2(y \mid x)\log \frac{P_2(x \mid y)}{P(x)}$$

$$= \sum_X \sum_Y P_2(x, y)\log \frac{P_2(x \mid y)}{P(x)}$$

为了分析平均互信息 $I(X;Y)$与信道转移概率 $P(y|x)$的凸性关系，由两种已知信道 $P_1(y|x)$和 $P_2(y|x)$通过线性组合构造第三种信道。

设 $0 \leqslant \alpha, \beta \leqslant 1$，$\alpha+\beta=1$，第三种信道的转移概率分布为

$$P(y \mid x) = \alpha P_1(y \mid x) + \beta P_2(y \mid x)$$

因为 $P_1(y|x)$和 $P_2(y|x)$满足概率分布率的条件，所以可推知 $P(y|x)$满足：

$$0 \leqslant \alpha P_1(y \mid x) + \beta P_2(y \mid x) \leqslant 1$$

和

$$\sum_Y (\alpha P_1(y \mid x) + \beta P_2(y \mid x)) = \alpha \sum_Y P_1(y \mid x) + \beta \sum_Y P_2(y \mid x) = \alpha + \beta = 1$$

显然，线性组合得到的第三种信道的转移概率分布率 $P(y|x)$是一组概率分布。

将给定信源 X 输出的符号经组合信道传输构成信息传输系统时，输入、输出随机变量 X、Y 的联合概率分布为

$$P(x, y) = P(x)P(y \mid x) = \alpha P(x)P_1(y \mid x) + \beta P(x)P_2(y \mid x)$$

$$= \alpha P_1(x, y) + \beta P_2(x, y)$$

进一步可以求出此时的边缘概率分布为

$$P(y) = \sum_X P(x)P(y \mid x)$$

条件概率分布为

$$P(x \mid y) = \frac{P(x)P(y \mid x)}{P(x)}$$

进而得到经组合信道传输给定信源的输出符号时的平均互信息 $I(X;Y)$为

$$I[P(y \mid x)] = \sum_X \sum_Y P(x)P(y \mid x)\log \frac{P(x \mid y)}{P(x)}$$

$$= \sum_X \sum_Y P(x, y)\log \frac{P(x \mid y)}{P(x)}$$

$$= \sum_X \sum_Y (\alpha P_1(x, y) + \beta P_2(x, y))\log \frac{P(x \mid y)}{P(x)}$$

考察互信息之间的关系：

$$
\begin{aligned}
&I[P(y\mid x)]-\alpha I[P_1(y\mid x)]-\beta I[P_2(y\mid x)]\\
&=\sum_{XY}(\alpha P_1(x,y)+\beta P_2(x,y))\log\left(\frac{P(x\mid y)}{P(x)}\right)\\
&\quad-\alpha\sum_{XY}P_1(x,y)\log\left(\frac{P_1(x\mid y)}{P(x)}\right)-\beta\sum_{XY}P_2(x,y)\log\left(\frac{P_2(x\mid y)}{P(x)}\right)\\
&=\alpha\sum_{XY}P_1(x,y)\log\left(\frac{P(x\mid y)}{P_1(x\mid y)}\right)+\beta\sum_{XY}P_2(x,y)\log\left(\frac{P(x\mid y)}{P_2(x\mid y)}\right)
\end{aligned}
$$

因为对数函数是上凸函数，所以应用颜森不等式得：

$$
\begin{aligned}
&\alpha\sum_{XY}P_1(x,y)\log\left(\frac{P(x\mid y)}{P_1(x\mid y)}\right)+\beta\sum_{XY}P_2(x,y)\log\left(\frac{P(x\mid y)}{P_2(x\mid y)}\right)\\
&\leqslant\alpha\log\left[\sum_{XY}\frac{P_1(x,y)}{P_1(x\mid y)}P(x\mid y)\right]+\beta\log\left[\sum_{XY}\frac{P_2(x,y)}{P_2(x\mid y)}P(x\mid y)\right]\\
&=\alpha\log\left[\sum_{Y}P_1(y)\sum_{X}P(x\mid y)\right]+\beta\log\left[\sum_{Y}P_2(y)\sum_{X}P(x\mid y)\right]\\
&=\alpha\log\sum_{Y}P_1(y)+\beta\log\sum_{Y}P_2(y)=0
\end{aligned}
$$

因此，对于 $0\leqslant\alpha,\beta\leqslant1$，$\alpha+\beta=1$，满足：

$$
I[\alpha P_1(y\mid x)+\beta P_2(y\mid x)]\leqslant\alpha I[P_1(y\mid x)]+\beta I[P_2(y\mid x)]
$$

所以，平均互信息 $I(X;Y)$是信道转移概率分布 $P(y|x)$的下凸函数。

证毕。

由平均互信息 $I(X;Y)$的定义知道，$I(X;Y)$是信道转移概率分布 $P(y|x)$的函数。如果使用不同的信道(信道转移概率分布 $P(y|x)$可变)传输给定信源(信源概率分布 $P(x)$固定)输出的符号，则在系统的输出端获得的平均信息量是不同的。定理 3.3 指出平均互信息 $I(X;Y)$与信道转移概率分布 $P(y|x)$之间的函数关系呈下凸性关系。这一平均互信息 $I(X;Y)$的重要性质表明，对于每一个给定的信源($P(x)$固定)，一定存在一种信道(转移概率分布为 $P(y|x)$)，使得传输该信源符号时的平均互信息 $I(X;Y)$达到其最小值。

3.7 信息系统的可靠性和有效性问题

对于一个广义的通信系统模型，信息传输的基本目的是在接收端准确地或近似地再现从信源输出的信息，如何有效地表示信源输出的信息，如何充分利用信道的信息传输能力实现信息的可靠传输，是信息系统的设计所追求的基本目标。因此，信息系统的有效性和可靠性是系统最优化技术研究中的基本问题。

在图 3-10 所示的简单通信系统的信息传输过程中，信源 X 发出的符号经信道传递，收信者通过观测信道的输出 Y 获得信源信息。平均而言，由每一个信道输出 Y 所得到的信源 X 的信息量被称为系统的信息传输率 R。由于平均互信息 $I(X;Y)$即为在信道的输出端接收到符号集 Y 后，平均由每一个输出符号中获得的关于信源 X 的信息量，因此，信息传输系统的信息传输率即为平均互信息，即

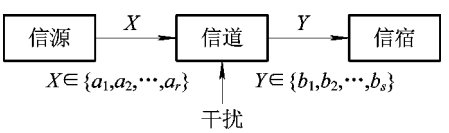

图 3-10 简单通信系统

$$R = I(X;Y) \quad \text{比特 / 符号}$$

由于平均互信息 $I(X;Y)$是信源概率分布 $P(x)$和信道转移概率分布 $P(y|x)$的函数，即

$$I(X;Y) = I[P(x);P(y|x)]$$

因此对于一个信息传输系统，如果信源不同($P(x)$不同)或者信道不同($P(y|x)$不同)，则该系统的信息传输率 R 不同。

由此可知，作为描述信息系统传输信息的特性和能力的一个基本物理量，平均互信息 $I(X;Y)$的凸性对于研究信息传输与处理系统的可靠性和有效性问题，有重要的理论指导意义和实用价值。

下面我们分别讨论平均互信息的凸性所具有的物理意义以及针对信息系统最优化中所反映的基本问题，分析平均互信息的凸性所具有的理论指导意义。

1. 互信息 $I(X;Y)$是信源概率分布的上凸函数的物理意义

由于信道中存在的随机性干扰可能使得信息传输发生错误，因此为了提高信息传输的可靠性，需要通过信道编码实现无差错的信息传输。对于一个给定信道，在充分利用其信息传输能力的条件下的无差错信息传输成为各类信道编码方法研究的目标。因此，给定有噪信道所具有的最大信息传输率 R 是信息系统可靠性研究中的一个基本的理论问题。

由于平均互信息 $I(X;Y)$是 $P(x)$的上凸函数，因此对于每一种给定的信道($P(y|x)$固定)，存在一个平均互信息(信息传输率)的最大值，即

$$\alpha I[P_1(x)] + \beta I[P_2(x)] \leqslant I[\alpha P_1(x) + \beta P_2(x)]$$

平均互信息是信源概率分布 $P(x)$的上凸函数表明，使用给定信道传输不同信源输出的信息时，系统的信息传输率不同。对于每一种给定信道($P(y|x)$固定)，一定存在一种信源概率分布 $P(x)$，能够使得系统的平均互信息达到其最大值。此最大平均互信息本质上依赖于给定信道的转移概率分布 $P(y|x)$，它由给定信道本身的特性所决定，反映的是给定信道传输信息的能力，即给定信道的最大信息传输率。

平均互信息 $I(X;Y)$的上凸性所指出的信息传输率的上界表明，对于一种给定的信道，一定存在一个最大的信息传输率 R。如果系统的信息传输率不超过此上界值，则总可以找到一种编码方法，实现信息的可靠传输。在信息系统工程应用中，平均互信息的上凸性所反映的最大的信息传输率指出了实现信息可靠传输的理论极限，对于信息系统的可靠性分析和信道编码原理研究具有重要的理论意义，在有噪信道的无差错编码技术研究中有明显的指导意义。

2. $I(X;Y)$是信道转移概率分布 $P(y|x)$的下凸函数的物理意义

对于一个信息系统，通过信源编码可以以较少的数据表示信源输出的信息(实际应用中常要求满足一定失真限度)，构成一个具有较高有效性的信息系统。因此，对于一个给定的信源，传输其输出信息时必须有的最小信息率 R 是信息系统有效性研究中的一个基本的理论问题。

平均互信息 $I(X;Y)$是 $P(y|x)$的下凸函数，即

$$I[\alpha P_1(y|x) + \beta P_2(y|x)] \leqslant \alpha I[P_1(y|x)] + \beta I[P_2(y|x)]$$

互信息的下凸性表明：

使用不同信道传输给定信源输出的信息时，系统的信息传输率不同。对于每一种给定信源，一定存在一种信道，使得信宿获得的关于此信源的平均信息率为最小。此最小值取决于给定信源的概率分布 $P(y|x)$，反映的是给定信源输出信息的特性。下凸性所指出的平均互信息 $I(X;Y)$的取值下界给出了传输该信源输出信息时，系统必须有的最小信息传输率 R。同时，平均互信息的下凸性也表明，如果系统的信息传输率 R 小于 $I(X;Y)$下界所指出的最小信息率，则该信源输出的信息将产生失真。

如果将不同信道看做不同的信源编码方法，那么对于一种给定的信源，在满足失真要求的条件下，若某种信源编码方法能够以接近 $I(X;Y)$下界所指出的最小信息率表示该源输出的信息，则该编码方法的有效性较高。

因此，平均互信息 $I(X;Y)$的取值下界指出了在满足一定失真度的要求下，传输该信源输出信息时信息系统必须有的最小信息传输率 R。可见，平均互信息 $I(X;Y)$的下凸性对于信息系统的有效性分析和高效信源编码原理具有重要理论意义，在图像、视频、音频、文本等各类信源编码和数据压缩的应用技术领域有明显的指导意义。

【例 3.4】 设离散无记忆信源：

$$[X,P]=\begin{bmatrix}0 & 1\\ \varepsilon & 1-\varepsilon\end{bmatrix}\qquad 0\leqslant\varepsilon\leqslant 1$$

通过图 3-11 所示的二进制对称信道传输。计算此时的平均互信息 $I(X;Y)$，讨论平均互信息 $I(X;Y)$随信源、信道的变化关系。

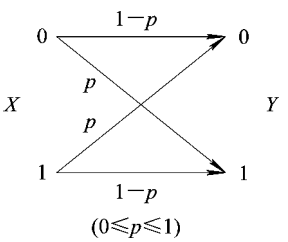

图 3-11 二进制对称信道

解：已知信源概率分布 $P(x)$和信道转移概率分布 $P(y|x)$，可以求出此输入随机变量 X 和输出随机变量 Y 的联合概率分布，以及随机变量 Y 的边缘概率分布 $P(y)$。

此通信系统中的各种概率关系如表 3-6 和表 3-7 所示。

表 3-6 $P(x,y)$，$P(y)$

	0	1	$P(x)$
0	$1-p$	p	ε
1	p	$1-p$	$1-\varepsilon$

表 3-7 $P(x,y)$，$P(y)$

	0	1
0	$\varepsilon(1-p)$	εp
1	$(1-\varepsilon)p$	$(1-p)(1-\varepsilon)$
$P(x)$	$\varepsilon(1-2p)+p$	$1-\varepsilon(1-2p)-p$

可以求出：

$$\begin{aligned}H(Y)&=H(\varepsilon(1-2p)+p,\ 1-\varepsilon(1-2p)-p)\\&=H(\varepsilon+p(1-2\varepsilon),\ 1-\varepsilon-p(1-2\varepsilon))\\&=-[\varepsilon(1-2p)+p]\log[\varepsilon(1-2p)+p]\\&\quad-[1-\varepsilon(1-2p)-p]\log[1-\varepsilon(1-2p)-p]\\&=-[\varepsilon+p(1-2\varepsilon)]\log[\varepsilon+p(1-2\varepsilon)]\\&\quad-[1-\varepsilon-p(1-2\varepsilon)]\log[1-\varepsilon-p(1-2\varepsilon)]\end{aligned}$$

$$\begin{aligned}H(Y\mid X)&=\sum_X P(x,y)\log\frac{1}{P(y\mid x)}\\&=\sum_X P(x)\Big[\sum_Y P(y\mid x)\log\frac{1}{P(y\mid x)}\Big]\\&=\sum_X P(x)H(p(1-p))=H(p,1-p)\\&=-p\log p-(1-p)\log(1-p)\end{aligned}$$

于是得到：

$$\begin{aligned}I(X\mid Y)&=H(Y)-H(Y\mid X)\\&=H(\varepsilon(1-2p)+p,1-\varepsilon(1-2p)-p)-H(p,1-p)\end{aligned}\tag{3.36}$$

当固定信道参数 p 和信源参数 ε 时，分别考察 $I(X;Y)$ 的函数变化关系。

(1) 如果信道给定(信道的参数 p 固定)，而信源可变(信源概率分布可变，即 ε 为变量)，则平均互信息 $I(X;Y)$ 随着 ε 的改变而改变。

当 $\varepsilon=0$ 和 $\varepsilon=1$ 时：

$$I(X;Y)=H(p,1-p)-H(p,1-p)=0\tag{3.37}$$

$I(X;Y)$ 是信源概率分布的上凸函数。随着 ε 在[0，1]范围内改变，$I(X;Y)$ 在其极值点处取得最大值。

于是，对平均互信息 $I(X;Y)$ 求偏导，得到 $I(X;Y)$ 随 ε 改变的极值点概率分布和平均互信息 $I(X;Y)$ 的最大值。

$$\begin{aligned}\frac{\partial I(X;Y)}{\partial\varepsilon}&=\frac{\partial H(Y)}{\partial\varepsilon}-\frac{\partial H(Y\mid X)}{\partial\varepsilon}=\frac{\partial H(Y)}{\partial\varepsilon}\\&=(1-2p)\log\frac{1-\varepsilon(1-2p)-p}{\varepsilon(1-2p)+p}\end{aligned}$$

令上式为 0，有

$$\frac{1-\varepsilon(1-2p)-p}{\varepsilon(1-2p)+p}=1\quad\Rightarrow\quad\varepsilon=\frac{1}{2}$$

由此可知，$\varepsilon=\frac{1}{2}$ 为使得平均互信息 $I(X;Y)$ 取得极大值的信源概率分布。

可以进一步求出此时信道的输出 Y 的边缘概率分布为

$$P(y=0)=P(y=1)=\frac{1}{2}$$

$$H(Y)=H\Big(\frac{1}{2},\frac{1}{2}\Big)=1\quad\text{比特/符号}$$

即当信源概率分布时，平均互信息 $I(X;Y)$ 取得最大值，且此最大值为

$$\begin{aligned}\max I(X;Y)&=\max\{H(Y)-H(Y\mid X)\}\\&=1-H(p,1-p)\quad\text{比特/符号}\end{aligned}$$

由此可以做出互信息 $I(X;Y)$ 随机信源概率分布 $P(x)$ 改变时的函数曲线(如图 3-12 所示)。

函数曲线表明，对于给定信道，在信源概率分布不同时，由信道的每一输出符号平均获得的信源输出的信息量不同。当信源概率为等概分布时，由此信道输出端获得的平均互信息最大。这个最大的平均互信息为

$$\max I(X;Y) = 1 - H(p, 1-p)$$

取决于信道参数 p，反映了给定信道传输信息的能力。

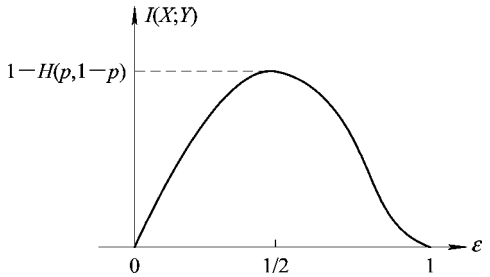

图 3-12 互信息 $I(X;Y)$随机信源概率分布 $p(x)$改变时的函数曲线

(2) 如果使用不同的信道(p 为变量)传输给定信源(ε 固定)输出的符号，则在接收端获得的平均互信息 $I(X;Y)$将随着 p 的改变而改变。

当 $p=0$ 或 $p=1$ 时：

$$\begin{aligned} I(X;Y) &= H(\varepsilon, 1-\varepsilon) - H(0, 1) \\ &= H(\varepsilon, 1-\varepsilon) = H(X) \quad \text{比特/符号} \end{aligned}$$

$I(X;Y)$是信道转移概率分布的下凸函数。随着 p 在[0，1]的范围内变化，平均互信息 $I(X;Y)$的函数关系是下凸的，在其极值点处 $I(X;Y)$取得最小值。

可以求出，当 $p=\frac{1}{2}$时信道中的平均互信息 $I(X;Y)$达到最小值。

由式(3.36)可知，此最小的平均互信息为

$$\begin{aligned} \min I(X;Y) &= \left[H(\varepsilon(1-2p)+p, 1-\varepsilon(1-2p)-p) - H(p, 1-p)\right]_{p=\frac{1}{2}} \\ &= H\left(\frac{1}{2}, \frac{1}{2}\right) - H\left(\frac{1}{2}, \frac{1}{2}\right) \\ &= 0 \quad \text{比特/符号} \end{aligned}$$

这个最小值表明，在图 3-11 所示的二进制对称信道中，当信道参数 $p=\frac{1}{2}$(信源参数 ε 为 0～1 之间的一个定值)时，信道中的干扰非常严重，以至于由信道的输出 Y 无法得到关于信源 X 输出的信息，$I(X;Y)=0$。

由关于平均互信息 $I(X;Y)$凸性的讨论和例题分析可以看出，平均互信息 $I(X;Y)$的凸性是反映信息系统最优化的一个具有对偶性的基本问题。

$I(X;Y)$是信源概率分布 $P(x)$的上凸函数，指出了给定信道的最大信息传输率。由于这一最大的信息传输率度量了给定信道传输信息的能力，因此 $I(X;Y)$的上凸性函数关系反映了信息系统的可靠性问题，指出了信道编码研究中的基本关系和追求目标。

$I(X;Y)$是信源概率分布 $P(y|x)$的下凸函数，指出了传输给定信源必须有的最小信息传输率。由于这一最小的信息传输率度量了给定信源输出信息的特性，因此 $I(X;Y)$的下凸性函数关系反映了信息系统的有效性问题，指出了信源编码方法研究中需要满足的基本关系和追求目标。

作为信息理论中的两个最基本的概念，信源的信息熵 $H(X)$和平均互信息 $I(X;Y)$在信息理论的学习和信息系统设计中有重要的意义和作用。

3.8　连续信道的平均互信息

本节将首先讨论单符号连续信道的平均互信息，然后得到波形信道的信息传输率，而后给出连续信道平均互信息的一些基本性质。

3.8.1　单符号连续信道的平均互信息

单符号连续信道的数学模型为$[X, p(y|x), Y]$，如图 3-13 所示。

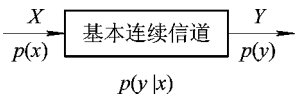

图 3-13　基本连续信道

其输入信源 X 为

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} \mathbf{R} \\ p(x) \end{bmatrix} \qquad \int_{\mathbf{R}} p(x)\mathrm{d}x = 1$$

输入信源 Y 为

$$\begin{bmatrix} X \\ p(y) \end{bmatrix} = \begin{bmatrix} \mathbf{R} \\ p(y) \end{bmatrix} \qquad \int_{\mathbf{R}} p(y)\mathrm{d}y = 1$$

而信道的传递概率密度函数为 $p(y|x)$。

其中，$\int_{\mathbf{R}} p(y \mid x)\mathrm{d}y = 1$。

与 2.5 节讨论的连续信源信息熵的问题一样，我们也可以对连续信道输入和输出随机变量的取值进行量化，将其转换成离散信道，求得此离散信道的平均互信息，然后将量化间隔 Δ 趋于零，就成为连续信道的平均互信息。

因此，单符号连续信道输入 X 和输出 Y 之间的平均互信息为

$$\begin{aligned} I(X;Y) &= H(X_n) - H(X_n \mid Y_n) \\ &= -\int_{\mathbf{R}} p(x)\log p(x)\mathrm{d}x - \lim_{\Delta\to 0}\log\Delta \\ &= -\left[-\iint_{\mathbf{R}} p(x)p(y \mid x)\log p(x \mid y)\ \mathrm{d}x\ \mathrm{d}y - \lim_{\Delta\to 0}\log\Delta\right] \\ &= \iint_{\mathbf{R}} p(xy)\log\frac{p(x \mid y)}{p(x)}\ \mathrm{d}x\ \mathrm{d}y = h(X) - h(X \mid Y) \qquad (3.38) \\ &= \iint_{\mathbf{R}} p(xy)\log\frac{p(y \mid x)}{p(y)}\ \mathrm{d}x\ \mathrm{d}y = h(Y) - h(Y \mid X) \qquad (3.39) \\ &= \iint_{\mathbf{R}} p(xy)\log\frac{p(xy)}{p(x)p(y)}\ \mathrm{d}x\ \mathrm{d}y = h(X) + h(Y) - h(XY) \qquad (3.40) \end{aligned}$$

对于连续信道的平均互信息来说，不但这些关系式和离散信道下平均互信息的关系式完全类似，而且它保留了离散信道的平均互信息的所有含义和性质，只是表达式中用连续信源的差熵替代了离散信源的熵。

由此可看出，将差熵定义为连续信源的熵是有重要的实际意义的。

单符号连续信道的信息传输率为

$$R = I(X;Y) \text{ 比特 / 自由度} \tag{3.41}$$

3.8.2 连续信道平均互信息的性质

两连续型随机变量之间的平均互信息的表达式与两离散随机变量之间的平均互信息的表达式不但类似，而且具有相同的性质。

下面我们给出两连续型随机变量之间的平均互信息的特性。

1. 非负性

$$I(X;Y) \geqslant 0 \tag{3.42}$$

证明：

$$\begin{aligned} h(X) - h(X \mid Y) &= -\int_R p(x)\log p(x)\mathrm{d}x + \iint_R p(xy)\log p(x \mid y)\mathrm{d}x\ \mathrm{d}y \\ &= -\iint_R p(xy)\log p(x)\mathrm{d}x\ \mathrm{d}y + \iint_R p(xy)\log p(x \mid y)\mathrm{d}x\ \mathrm{d}y \\ &= -\iint_R p(xy)\log \frac{p(x)}{p(x \mid y)}\mathrm{d}x\ \mathrm{d}y \end{aligned}$$

其中：

$$\int_R p(xy)\mathrm{d}y = p(x)$$

因为 $-\log X$ 是下凸函数，根据颜森不等式得：

$$\begin{aligned} h(X) - h(X \mid Y) &\geqslant -\log\iint_R p(xy)\,\frac{p(x)}{p(x \mid y)}\mathrm{d}x\ \mathrm{d}y \\ &= -\log\iint_R p(y)p(x)\mathrm{d}x\ \mathrm{d}y = \log 1 = 0 \end{aligned}$$

上式运用 $\int_R p(xy)\mathrm{d}x = 1$ 和 $\int_R p(xy)\mathrm{d}y = 1$ 求得。

从证明过程中可知，等式成立的条件是 $p(x)=p(x|y)$，即连续随机变量 X 和 Y 统计独立。

证毕。

上述证明方法与离散变量的证明方法是一致的，只是在公式中把求和号换成积分号，把概率分布换成概率密度函数。因此，在离散变量中所得的有关结论均可推广到连续变量中。

2. 对称性（交互性）

因为 $p(xy)=p(yx)$，所以

$$\begin{aligned} I(X;Y) &= \iint_R p(xy)\,\frac{p(xy)}{p(x)p(y)}\ \mathrm{d}x\ \mathrm{d}y \\ &= \iint_R p(yx)\,\frac{p(yx)}{p(y)p(x)}\ \mathrm{d}x\ \mathrm{d}y = I(Y;X) \end{aligned}$$

当 X 和 Y 统计独立时：

$$p(x \mid y) = p(x)$$

$$I(X; Y) = I(Y; X) = 0$$

此时不可能从一个随机变量获得关于另一个随机变量的信息。

3. 凸状性

连续变量之间的平均互信息 $I(X; Y)$是输入连续变量 X 的概率密度函数 $p(x)$的上凸函数，$I(X; Y)$又是连续信道传递概率密度函数 $p(y|x)$的下凸函数。

凸状性的证明方法类似于离散情况中的证明，此处从略。

4. 信息不增性

设连续信道输入变量为 X，输出变量为 Y。

若对连续随机变量 Y 再进行处理而成为另一连续随机变量 Z，一般总会丢失信息，最多保持原获得的信息不变，而所获得的信息不会增加。

这也就是数据处理定理，即

$$I(X; Z) \leqslant I(X; Y) \tag{3.43}$$

其中，$z=f(y)$，当且仅当该函数是一一对应时式(3.43)中的等式成立。

证明：设 X、Y、Z 三连续随机变量之间形成两个连续信道的串接，如图 3-14 所示。

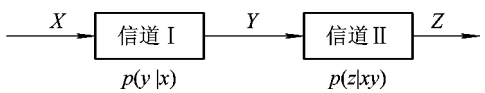

图 3-14　两串接连续信道

因为

$$I(XY; Z) = \iiint_{R} p(xyz) \log \frac{p(z \mid xy)}{p(z)} \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z$$

而

$$I(Y; Z) = \iiint_{R} p(yz) \log \frac{p(z \mid y)}{p(z)} \, \mathrm{d}y \, \mathrm{d}z$$

$$= \iiint_{R} p(xyz) \log \frac{p(z \mid y)}{p(z)} \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z$$

其中，因为

$$\int_{R} p(xyz) = p(yz)$$

所以得到：

$$I(Y; Z) - I(XY; Z) = E\left[\log \frac{p(z \mid y)}{p(z)}\right] - E\left[\log \frac{p(z \mid xy)}{p(z)}\right]$$

$$= E\left[\log \frac{p(z \mid y)}{p(z \mid xy)}\right]$$

其中，$E[\cdot]$为对 X、Y、Z 三个连续概率空间求平均。

运用颜森不等式，可得

$$
\begin{aligned}
I(Y;\ Z)-I(XY;\ Z) &\leqslant \log E\left[\frac{p(z\mid y)}{p(z\mid xy)}\right] \\
&= \log\iiint_{R} p(xyz)\log\frac{p(z\mid y)}{p(z\mid xy)}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \\
&= \log\iiint_{R} p(xy)p(z\mid y)\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z \\
&= \log\iiint_{R} p(xy)\mathrm{d}x\,\mathrm{d}y\int_{R} p(z\mid y)\mathrm{d}z \\
&= \log 1 = 0
\end{aligned}
$$

其中，因为 $\int_{R} p(z\mid xy)\mathrm{d}z=1$，$\int_{R} p(y\mid x)\,\mathrm{d}y=1$，$\iint_{R} p(xy)\,\mathrm{d}x\,\mathrm{d}y=1$，**所以**

$$\int_{R} p(z\mid y)\mathrm{d}z=1$$

由此证得：

$$I(XY;\ Z)\geqslant I(Y;\ Z) \tag{3.44}$$

当且仅当对所有 x、y、z **都有** $p(z|xy)=p(z|y)$，**即** X、Y、Z **是马氏链时，式**(3.44)**中的等式成立。**

同理可证得：

$$I(XY;\ Z)\geqslant I(X;\ Z) \tag{3.45}$$

当且仅当对所有 x、y、z **都有** $p(z|xy)=p(z|x)$**时等式成立。**

一般情况下，两串接连续信道的输入、输出随机变量 X、Y、Z **形成马氏链，即** Z **只通过** Y **与** X **发生联系，与** X **没有直接联系，就有**

$$p(z\mid xy)=p(z\mid y)\quad(\text{对所有 } x、y、z)$$

所以

$$I(XY;\ Z)=I(Y;\ Z)$$

$$I(Y;\ Z)\geqslant I(X;\ Z) \tag{3.46}$$

当且仅当对所有 x、y、z **有** $p(z|xy)=p(z|x)$**时等式成立。**

因为 X、Y、Z **是马氏链，**Z、Y、X **也是马氏链，所以同式**(3.44)**和式**(3.45)**一样可证明得：**

$$I(ZY;\ X)\geqslant I(Y;\ X) \tag{3.47}$$

$$\geqslant I(Z;\ X) \tag{3.48}$$

当且仅当对所有 x、y、z **有** $p(x|yz)=p(x|y)$**时式**(3.47)**中的等式成立，当且仅当对所有** x、y、z **有** $p(x|yz)=p(x|z)$**时式**(3.48)**中的等式成立。**

因为 Z、Y、X **是马氏链，所以有** $p(x|yz)=p(x|y)$，**对所有** x、y、z，**式**(3.47)**中的等式成立，由此可得：**

$$I(Z;\ X)\leqslant I(Y;\ X)$$

即证得：

$$I(X;\ Z)\leqslant I(X;\ Y)$$

当且仅当 $p(x|yz)=p(x|z)$**时对所有** x、y、z，**上式的等号才成立。**

证毕。

此证明过程与离散串接信道的证明类似。

由此性质也可以得出，如果连续随机变量 Z 和 Y 只是坐标发生一一对应的变换，则其平均互信息应该不变。

5. $I(X;Y)$**与** $I(X_i;Y_i)$**的关系**

若平稳连续信源是无记忆的，即 X 中各分量 X_i 相互统计独立，则存在：

$$I(X;Y)\geqslant\sum_{i=1}^{N}I(X_i;Y_i) \tag{3.49}$$

若多维连续信道是无记忆的，则存在：

$$I(X;Y)\leqslant\sum_{i=1}^{N}I(X_i;Y_i) \tag{3.50}$$

若平稳连续信源是无记忆的，多维连续信道也是无记忆的，则

$$I(X;Y)=\sum_{i=1}^{N}I(X_i;Y_i) \tag{3.51}$$

以上结论的证明与离散情况相同，请读者自行证明。

习　题　3

3.1　信道的输入和输出分别为 X 和 Y，已知 X 的符号集为 $\{x_1,x_2,x_3\}$，Y 的符号集为 $\{y_1,y_2\}$，X、Y 的联合概率 $P(x_i,y_j)$ 如表 3-8 所示。求出信道的转移概率 $P(y_j|x_i)$，写出信道矩阵并用图形表示此信道。

表 3-8　题 3.1 表

$P(x_i,y_j)$	y_1	y_2
x_1	0.1	0.2
x_2	0.2	0.1
x_3	0.3	0.1

3.2　设某信道中干扰很强，使得信道的输入、输出随机变量 X 和 Y 相互独立。此时 $H(X|Y)$、$I(X;Y)$ 是多少？能否从 Y 中得到关于 X 的信息？为什么？

3.3　设有一理想无干扰信道，信道的输出 Y 与 X 有一一对应关系。在对 Y 进行观测后，对信源 X 是否存在不确定性？此时获得的信息量是多少？能否获得更多的关于 X 的信息？为什么？

3.4　若发送符号集为 $X=\{x_1,x_2,x_3,x_4\}$，接收符号集为 $Y=\{y_1,y_2\}$，信道如图 3-15 所示，且已知 $P(x_1)=0.3$，$P(x_2)=0.3$，$P(x_3)=0.25$，$P(x_4)=0.15$。

(1) 计算无条件熵 $H(X)$ 和条件熵 $H(X|Y)$、$H(Y|X)$。

(2) 计算 $I(x_1;y_1)$、$I(x_1;y_2)$、$I(x_2;y_1)$、$I(x_2;y_2)$，并求出系统的平均信息量 $I(X;Y)$。

3.5　设离散无记忆信源 $[X,P]=\begin{bmatrix}0 & 1\\ 0.2 & 0.8\end{bmatrix}$ 的输出经过图 3-16 所示的二元信道传

输，试计算 $I(x=0;\ y=1)$、$I(x=1;\ Y)$和 $I(X;\ Y)$。

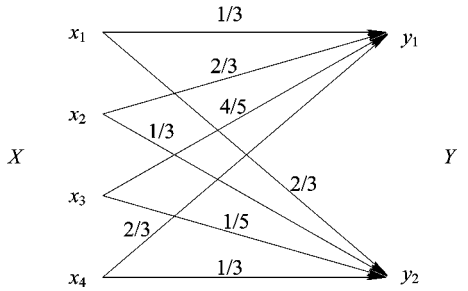

图 3－15　题 3.4 图

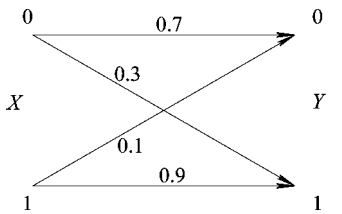

图 3－16　题 3.5 图

3.6　一个电报系统传输两种信号：传号 M 和空号 S。假设在发送端发送 M 和 S 的概率相等，由于信道上有干扰，因此有 1/6 的传号 M 被错传成空号 S，同时有半数的空号 S 被错传成传号 M。典型的收发符号序列如下：

发送：M M M　M M M　S S S　S S S

接收：M M M　M M S　S S S　M M M

试求每个符号传输的平均信息量。

3.7　设信源 $[X,\ P(x)]=\begin{bmatrix} x_1 & x_2 \\ 0.6 & 0.4 \end{bmatrix}$ 通过一个干扰信道，接收符号集为 $Y=\{y_1,\ y_2\}$。信道转移概率如表 3－9。试求：

(1) 信源 X 中事件 x_1、x_2 分别有的自信息；

(2) 收到消息 y_j 后获得的关于 x_i 的信息量$(i,\ j=1,\ 2)$；

(3) X 和 Y 的无条件熵；

(4) 条件熵 $H(X|Y)$和 $H(Y|X)$；

(5) 接收到 Y 后获得的平均信息量 $I(X;\ Y)$。

表 3－9　题 3.7 表

	y_1	y_2
x_1	$\frac{5}{6}$	$\frac{1}{6}$
x_2	$\frac{3}{4}$	$\frac{1}{4}$

3.8　设有离散无记忆信道，输入 X：$\begin{Bmatrix} a_1 & a_2 & \cdots & a_K \\ P(a_1) & P(a_2) & \cdots & P(a_K) \end{Bmatrix}$，输出 Y：$\begin{Bmatrix} b_1 & b_2 & \cdots & b_J \\ P(b_1) & P(b_2) & \cdots & P(b_J) \end{Bmatrix}$，当输入、输出 x 和 y 的互信息 $I(x;\ y)$也为一随机变量时，试证：当 $\mathrm{Var}\{I(x;\ y)\}=0$ 时，平均互信息 $I(X;\ Y)$达到信道容量 C。

第 4 章　信道与信道容量

由前面的讨论我们已经知道，在一般的信息传递过程中，信宿对信源的了解(即获取信源发出的信息)是通过对信道的输出 Y 进行观测而实现的。因此，信道是信息传输系统的重要部分。在一般的信息传输、处理系统的研究中，对于信道所具有的功能和所包括范围的理解应当注意下面两点。

(1) 信道承担了信息的传输、存储和处理等任务。

通常我们认为信道是传输信息的物理媒介。但是对于一般信息系统，在发出信息的信源与接收信息的信宿之间所进行的各种变换与处理，都可以看做此信息系统中信道所具有的功能。

例如，地球资源卫星遥感信息获取与处理系统，放置在资源卫星上的多光谱扫描设备采集地球表面地物的反射、辐射电磁波信号，得到遥感影像数据并以无线通信方式传送，地面接收站接收遥感影像数据并记录在存储介质中，根据应用要求进行滤波、增强、分类、识别，提取各类目标信息。如果将反射、辐射电磁波的地球(信息源)与研究人员(信宿)之间看做信道，则此信道中不仅进行着信息的传递，还包含着对信息的存储和处理。

因此，在一般的信息系统中，信道除了具有在某种物理媒介中传递信号的功能之外，还应当包括对信息的采集、编码表示、存储、重建、提取、识别等多种信息处理功能。

(2) 信道所包括的范围可以根据研究对象和应用条件的不同而确定。

在对信息传输系统进行分析和研究时，针对系统中不同的处理要求和处理方法，往往在分析研究的具体对象、信号特征等方面有较大差异。为了使分析更加简明，更有针对性，常将信道在信息系统中所包括的范围加以适当的限定。

例如，对于一个多媒体数据光盘的刻录、播放系统，信道的范围可包括：视频和音频数据的采集、压缩编码，激光盘片刻录，读盘、解压缩恢复视频和音频数据，显示还原图像、声音信息的全过程。如果仅考虑播放过程，则信源指数字媒体作品光盘，信宿为观看者。因此信道的范围为：数据读取、解压缩并按照一定格式重建视频、音频信号等处理过程。

显然，根据不同的应用要求和研究对象明确信道所包括的范围，这是我们明确信息系统设计的任务，从而可有针对性地开展研究与分析工作。

在信息理论基础研究中，信道的主要任务是以符号的形式传输信息和存储信息，对于信道进行研究的主要问题是信道中能够传送或存储的最大信息量，即信道容量问题。

这一章我们将从信道的统计特性分析入手，给出信道容量的定义并对常见信道讨论其信道容量的计算方法。

4.1 信道的描述和分类

信道是传输载荷信息的消息(信号)的物理媒介或通道。信道的输入(即信源输出的消息)是随机变量，加之通信系统中存在着随机性的噪声或干扰，因此信道的输出也是一个随机变量。更一般地讲，信源的输出是广义的时间连续的随机信号或随机变量序列，常用随机过程来描述，因而信道的输出也是连续的随机信号或随机变量序列，也要用随机过程加以描述。

因此，信道的描述和分类都需要从统计的观点出发，用统计的方法加以分析和表述。

4.1.1 信道的描述

在前面的讨论中，我们已经给出了最简单的信道，即单符号的离散信道的数学模型。如果信道的输入是有 r 种取值可能的随机变量 $\boldsymbol{X}$，输出是有 s 种取值可能的随机变量 $\boldsymbol{Y}$，则信道输入、输出之间的统计关系可以由转移概率：

$$P(y_j \mid x_i) \qquad i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s \tag{4.1}$$

表示。

然而，在实际的信息传输系统中，信源的输出为一个随机变量序列 $\boldsymbol{X}=(X_1X_2\cdots X_N)$，在信道的输出端相应地为输出随机变量序列 $\boldsymbol{Y}=(Y_1Y_2\cdots Y_N)$。如果 X_i 有 r 种取值，则 $\boldsymbol{X}$ 的样矢量有 r^N 种；如果 Y_i 有 s 种取值，则 $\boldsymbol{Y}$ 的样矢量有 s^N 种。因此信道的数学模型应当由输入随机矢量 $\boldsymbol{X}$ 和输出随机矢量 $\boldsymbol{Y}$ 的转移概率关系 $P(\boldsymbol{y}|\boldsymbol{x})$ 来描述。这组 $r^N \times s^N$ 个条件概率反映出了输入随机变量与输出随机变量每一对样值之间的统计依赖关系，反映出了一般信道所具有的最基本的统计特性。

4.1.2 信道的分类

根据信道的输入、输出随机变量或随机矢量的取值类型、它们的相互关系以及在某些方面的特点，信道可以划分为不同的类型。

1. 按随机变量的取值类型划分

(1) 离散信道：信道的输入、输出随机变量的取值均为离散事件的集合。离散信道有时也称做数字信道。

(2) 连续信道：信道的输入、输出随机变量 X 和 Y 都是连续事件。这类信道也常称为模拟信道。

(3) 半离散半连续信道：信道的输入和输出随机变量空间一个是离散集，另一个是连续集。

例如，模/数转换系统的输入是连续的随机变量，输出是离散随机变量，称为输入连续、输出离散的信道。反之，数/模转换系统是输入离散、输出连续的信道。

2. 按信道的输入、输出的个数划分

(1) 单用户信道：只有一个输入信号和一个输出信号构成的信道。

(2) 多用户信道：信道的输入或输出中至少有一个或两个以上事件构成的集合。

3. 按输入随机变量和输出随机变量的统计关系的特点划分

条件概率 $P(y|x)$具有不同的特点时，信道的输入、输出具有不同的概率转移关系，信息的传递也具有不同的特性。对于离散信道，根据条件概率的特点，可分为以下三种情况。

1）无干扰信道

无干扰信道中没有随机性的干扰，输出矢量 $\boldsymbol{Y}$ 与输入矢量 $\boldsymbol{X}$ 之间有确定的对应关系。这种确定关系可以是下面三种形式的某一种。

（1）无噪无损信道。此时 $\boldsymbol{Y}$ 是 $\boldsymbol{X}$ 的一一对应的函数，如图 4－1 所示，有 $y=f(x)$，即

$$P(\boldsymbol{y}\mid\boldsymbol{x})=\begin{cases}1 & \boldsymbol{y}=f(\boldsymbol{x})\\ 0 & \boldsymbol{y}\neq f(\boldsymbol{x})\end{cases} \tag{4.2}$$

信道转移概率呈 0、1 分布，$\boldsymbol{X}$、$\boldsymbol{Y}$ 之间的互信息为

$$I(\boldsymbol{X};\boldsymbol{Y})=H(\boldsymbol{X})=H(\boldsymbol{Y}) \tag{4.3}$$

由于这样的信道中无随机性干扰，因此由 $\boldsymbol{Y}$ 所得到的关于 $\boldsymbol{X}$ 的信息与 $\boldsymbol{X}$ 或 $\boldsymbol{Y}$ 的熵相同，即在这样的信道中不会损失信息。

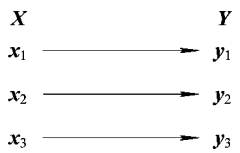

图 4－1　无噪无损信道示例

（2）有噪无损信道。此类信道的 $\boldsymbol{Y}$ 也是 $\boldsymbol{X}$ 的确定函数，但是对于某一个输入随机矢量 $\boldsymbol{X}$ 可能会有多个输出随机矢量，而相应的输出随机矢量只有确定的一个 $\boldsymbol{X}$ 与之对应，如图 4－2 所示（图中 $\alpha+\beta+\gamma=1$）。

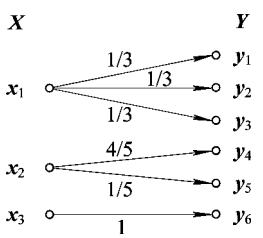

图 4－2　有噪无损信道示例

在有噪无损信道中，由某一个输入随机矢量 x_i 并不能唯一地确定信道的输出，表明信道中存在着干扰。但是由于 y_j 与 x_i 有唯一的确定关系，x_i 载荷的信息可以由 y_j 全部表示，因此信道中仍然没有产生信息的损失，即

$$I(\boldsymbol{X};\boldsymbol{Y})=H(\boldsymbol{X})\leqslant H(\boldsymbol{Y}) \tag{4.4}$$

$$\max I(\boldsymbol{X};\boldsymbol{Y})\rightarrow\max H(\boldsymbol{X}) \tag{4.5}$$

（3）无噪有损信道。此时 $\boldsymbol{Y}$ 是 $\boldsymbol{X}$ 的确定函数，但是不同的输入 $\boldsymbol{X}$ 可能会有同样的输出 $\boldsymbol{Y}$，如图 4－3 所示。

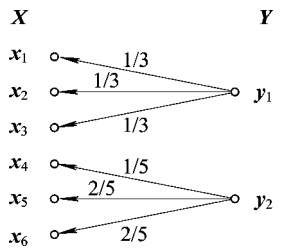

图 4-3 无噪有损信道示例

对于无噪有损信道，由于每一种信道输入 x_i 只有唯一的一种输出，因此信道中无随机性干扰。但是由于多个信道输入 x_i 对应于同一种输出随机矢量 y_j，因此系统传输的信息将会产生损失，即

$$I(\boldsymbol{X};\boldsymbol{Y}) = H(\boldsymbol{Y}) \leqslant H(\boldsymbol{X}) \tag{4.6}$$

2）有干扰无记忆信道

通常，实际信道的输入随机矢量 $\boldsymbol{X}$ 与输出随机矢量 $\boldsymbol{Y}$ 之间并没有确定的对应关系，而是某种概率转移关系。因此，实际信道中常存在着随机性的干扰。

如果信道在某一时刻输出的随机变量仅统计依赖于即时的输入随机变量 X，与前面时刻信道的输入、输出随机变量无关，则称这样的信道为有干扰无记忆信道。满足离散无记忆信道的充要条件为

$$\begin{aligned} P(\boldsymbol{Y} \mid \boldsymbol{X}) &= P(y_1 y_2 \cdots y_N \mid x_1 x_2 \cdots x_N) \\ &= \prod_{i=1}^{N} P(y_i \mid x_i) \end{aligned} \tag{4.7}$$

因此有干扰无记忆信道的特性可以由转移概率矩阵$[P(\boldsymbol{y}|\boldsymbol{x})]$给出。

3）有干扰有记忆信道

有干扰有记忆信道是最一般的信道，信道在某时刻的输出随机变量不仅与信道即时的输入随机变量有关，而且可能与前面时刻信道的输入、输出随机变量有关，即表现为信道的输出是有记忆的。这时信道的条件概率不再满足式(4.7)。

处理这类有记忆信道时，最直观的方法是把记忆较强的 N 个符号当作一个矢量符号来处理，而各矢量符号之间认为是无记忆的，这样就转化成无记忆信道的问题。当然，这样处理一般会引入误差，因为实际上第一个矢量的最后几个符号与第二个矢量的最前面几个符号是有关联的。N 取得越大，误差将越小。

另一种处理方法是把 $P(y|x)$看成 Markov 链的形式，这是有限记忆信道的问题。把信道某时刻的输入和输出序列看成为信道的状态，那么，信道的统计特性可用在已知当前时刻的输入符号和前一时刻信道所处状态的条件下，信道的输出符号和所处状态的联合条件概率来描述，即用 $P(y_n, S_n|x_n, S_{n-1})$来描述。

在本课程的讨论中，我们仅仅考虑简单的单符号离散无记忆信道以及高斯白噪声波形信道。

4.1.3 离散无记忆信道

设离散信道的输入随机变量为 X，取值空间为$\{a_1, a_2, \cdots, a_r\}$，输出随机变量为 Y，

取值空间为$\{b_1, b_2, \cdots, b_s\}$，信道的转移概率为一组条件概率，即

$$P(y \mid x) = P(y = b_j \mid x = a_i) = P(b_j \mid a_i) \quad i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s \tag{4.8}$$

其中，

$$\begin{cases} P(b_j \mid a_i) \geqslant 0 & i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s \\ \sum\limits_{j=1}^{s} P(b_j \mid a_i) = 1 & i = 1, 2, \cdots, r \end{cases} \tag{4.9}$$

由于信道中存在随机性干扰，因此有可能使传输产生错误。这种信道干扰对传输的影响可以用转移概率 $P(b_j|a_i)(i=1, 2, \cdots, r;\ j=1, 2, \cdots, s)$来描述。于是，信道矩阵 $\boldsymbol{P}$ 实际上是一个转移概率矩阵，也称为信道转移概率矩阵。即

$$\boldsymbol{P} = \{P(b_j \mid a_i) = p_{ij},\ i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s\} \tag{4.10}$$

或

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rs} \end{bmatrix} \tag{4.11}$$

并且满足 $p_{ij} \geqslant 0$，$\sum\limits_{j=1}^{s} p_{ij} = 1$，$i = 1, 2, \cdots, r$。

【例 4.1】　二进制对称信道(Binary Symmetrical Channel，BSC)如图 4-4 所示。

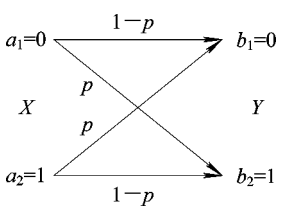

图 4-4　二进制对称信道

在图 4-4 中，输入、输出随机变量 X，$Y=\{0, 1\}$。因为 $r=s=2$，所以我们将此类信道称为二进制对称信道。

由前面的讨论可知，二进制对称信道的转移概率为

$$\begin{cases} P(b_1 \mid a_1) = P(0 \mid 0) = 1 - p = \overline{p} \\ P(b_2 \mid a_2) = P(1 \mid 1) = 1 - p = \overline{p} \\ P(b_1 \mid a_2) = P(0 \mid 1) = p \\ P(b_2 \mid a_1) = P(1 \mid 0) = p \end{cases} \tag{4.12}$$

可以看出，$\overline{p}$ 表示单个符号无错误传输的概率，而 p 表示单个符号在传输中发生错误的概率。

如果用信道矩阵写出各转移概率，则有：

$$\boldsymbol{P} = \begin{bmatrix} \overline{p} & p \\ p & \overline{p} \end{bmatrix} \tag{4.13}$$

此矩阵为一对称矩阵，因此这种信道被称为二进制对称信道。

4.1.4 高斯白噪声加性波形信道

信道中的噪声是高斯白噪声。具有高斯分布的白噪声称为高斯白噪声。高斯噪声和白噪声是从不同的角度来定义的。高斯噪声是指它的 N 维概率密度函数服从高斯分布，并不涉及其功率谱密度的形状；白噪声则是就其功率谱密度是均匀分布而言的，而不论它服从什么样的概率分布。

一般情况下，把既服从高斯分布而功率谱密度又是均匀分布的噪声称为高斯白噪声。电阻内的热噪声就是高斯白噪声的一个典型实例。因此，通信系统中的波形信道常假设为高斯白噪声信道。

低频限带高斯白噪声有一个很重要的性质，即低频限带高斯白噪声$\{n'(t)\}$(均值为零，功率谱密度为 $N_0/2$)是通过一个理想低通滤波器后所得到的。如果理想低通滤波器的带宽为 F Hz，那么它的传递函数的频率响应为

$$K(\omega)=\begin{cases}1 & -2\pi F\leqslant\omega\leqslant 2\pi F\\ 0 & \text{其他}\end{cases}$$

(考虑双边谱密度)因此低频限带高斯白噪声的功率谱密度为

$$P_{n'}(\omega)=P_n(\omega)K(\omega)=\begin{cases}\dfrac{N_0}{2} & -2\pi F\leqslant\varepsilon\leqslant 2\pi F\\ 0 & \text{其他}\end{cases}$$

其自相关函数为

$$R_{n'}(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}P_{n'}(\omega)\mathrm{e}^{\mathrm{j}\omega\tau}\mathrm{d}\omega=N_0F\,\frac{\sin(2\pi F\tau)}{2\pi F\tau}$$

其中，$N_0F=P_{n'}$为噪声平均功率。因为其均值为零，所以$\{n'(t)\}$的平均功率为

$$P_{n'}=E[n'^2]=\sigma_{n'}^2=R(0)=N_0F$$

则

$$R_{n'}(\tau)=\begin{cases}N_0F & \tau=0\\ 0 & \tau=\dfrac{n}{2F}(n=\pm1,\pm2,\cdots)\end{cases}$$

这表示，在时间间隔 $\Delta=\dfrac{1}{2F}$的两个样本点之间的相关函数等于零，即

$$R_{n'}(\Delta)=R_{n'}\left(\frac{1}{2F}\right)=0$$

所以各样本点之间不相关。也就是说，频率受限(频率限制在$-F\sim F$ 之间或限制在 $0\sim F$ 之间)的高斯白噪声$\{n'(t)\}$，其在有限 T 时间内取样后分解成 $N=2FT$ 个连续型随机变量 n_1、n_2、…、n_N 时，这些随机变量相邻的时间间隔为 $\Delta=\dfrac{1}{2F}$，所以它们的相关函数等于零，它们之间是不相关的。又因为各随机变量是高斯概率密度分布，所以随机变量之间统计独立。每个随机变量 $n_i(i=1,2,\cdots,N)$都是均值为零，方差为$\dfrac{N_0FT}{2FT}=\dfrac{N_0}{2}$的高斯变量。

4.2　信道容量的定义

4.2.1　信息传输率

在信息传输系统中，信道每传递一个符号所能够传递(载荷)的平均信息量称为信道的信息传输率，记做 R。

在图 4-5 所示的系统中，平均互信息 $I(X;Y)$即为在信道的输出端接收到符号集 Y 后，由每一个输出符号中获得的关于信源 X 的信息量。因此信道的信息传输率就是这样的信息系统中的平均互信息，即

$$R = I(X;Y) \quad \text{比特 / 符号} \tag{4.14}$$

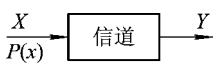

图 4-5　信息传输系统模型

有时我们关心的是信道在单位时间内能够传递多少信息量，或者是在传输一个符号的时间内信道中平均传递的信息量。此时，若平均传输一个符号需要 t 秒，而每一符号传送的信息量为 $I(X;Y)$，则信道每秒传输的信息量便为

$$R_t = \frac{1}{t} I(X;Y) \quad \text{比特 / 秒} \tag{4.15}$$

通常将 R_t 称为信息传输速率。

4.2.2　信道容量

由第 3 章的讨论可知，平均互信息 $I(X;Y)$不但与信道的转移概率 $P(y|x)$有关，而且与信源的概率分布 $P(x)$有关。因此 $R=I(X;Y)$只表示出了传输信源符号时信道所传送的信息率。在研究一个通信系统时我们希望能够找到这样的量，它能够反映出给定信道所具有的传输信息的能力，即该信道传输信息的最大能力。

在对互信息的讨论中，我们已经知道，对于

$$I(X;Y) = I(P(x), P(y \mid x)) \tag{4.16}$$

当给定信道转移概率 $P(y|x)$之后，$I(X;Y)$是信源概率分布 $P(x)$的上凸函数。

因此，对于一个信道转移概率为 $P(y|x)$的给定信道，总存在着一种信源(其概率分布为 $P(x)$)，使传输每一个符号平均获得的信息量最大。这说明对于每一个给定的信道，都具有一个最大的信息传输率，这个最大的信息传输率反映出了该信道传输信息的极限能力。

定义这个最大的信息传输率为该信道的信道容量 C，即

$$C = \max_{P(x)} \{I(X;Y)\} \quad \text{比特 / 符号} \tag{4.17}$$

如果平均传输一个符号需要 t 秒钟，则给定信道在单位时间内平均传输的最大信息量(即信息容量)为

$$C_t = \frac{1}{t} \max_{P(x)} \{I(X;Y)\} \quad \text{比特 / 秒} \tag{4.18}$$

由上述分析和定义可以看出，信道容量 C 与信源的概率分布 $P(x)$ 并无函数关系，它只是信道转移概率的函数，只与信道的统计特性 $P(y|x)$ 有关。所以，C 是描述信道传输信息能力的一个参数。信道容量的计算可以通过找出适当的 $P(x)$，使 $I(X;Y;P(x))$ 为极大值来完成。

【例 4.2】 对于例 4.1 给出的二进制对称信道，设输入信源的概率空间为

$$[X,P]=\begin{bmatrix}0 & 1\\ w & 1-w\end{bmatrix}$$

在这样的通信系统中，可以求出平均互信息为

$$\begin{aligned}I(X;Y)&=H(Y)-H(Y\mid X)\\&=H(Y)-\sum_X P(x)\sum_Y P(y\mid x)\log\frac{1}{P(y\mid x)}\\&=H(Y)-\sum_X P(x)\left[p\log\frac{1}{p}+(1-p)\log\frac{1}{1-p}\right]\\&=H(Y)-\left[p\log\frac{1}{p}+(1-p)\log\frac{1}{1-p}\right]=H(Y)-H_2(p)\end{aligned}$$

根据联合概率分布可求出输出符号 Y 的边缘概率为

$$P(y_j)=\sum_X P(x)P(y_j\mid x)$$

即

$$\begin{aligned}P(y=0)&=w(1-p)+(1-w)p\\P(y=1)&=wp+(1-w)(1-p)\end{aligned}$$

因此

$$\begin{aligned}I(X;Y)&=(w(1-p)+(1-w)p)\log\frac{1}{w(1-p)+(1-w)p}\\&\quad+(wp+(1-w)(1-p))\log\frac{1}{wp+(1-w)(1-p)}-H_2(p)\\&=H(w(1-p)+(1-w)p)-H_2(p)\end{aligned}$$

对于给定的信道，即固定参数 p 时，互信息 $H(X;Y)$ 是 w 的上凸函数，函数曲线如图 4-6 所示。

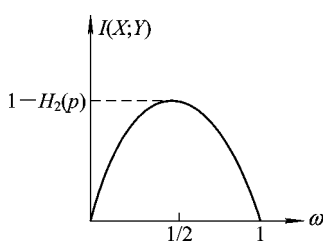

图 4-6 二进制对称信道的互信息

由图 4-6 可以看出，对于给定的二进制对称信道，当信源 X 的概率分布不同时，在接收端平均由每个符号获得的信息量不同。当输入符号集 X 是等概率分布，即

$$P(x=0)=P(x=1)=\frac{1}{2}$$

时，信道输出端平均每个符号才能获得最大信息量，并且这个最大的互信息(即此信道的信道容量)为

$$C = \max_{P(x)}\{I(X;Y)\} = 1 - H_2(p)$$

对于一般的信道，其信道容量的计算就是对互信息 $I(X;Y)$求最大值的问题。由于一般信道的信道容量的计算比较复杂，因此本课程主要针对几类特殊类型的信道，讨论其信道容量的计算方法。

4.3　离散信道的信道容量

4.3.1　对称信道的信道容量

由前面的讨论我们知道，离散信道可以由其转移概率排成的信道矩阵加以描述。因此，若信道矩阵具有不同的特性，那么信道将有不同的特点。

首先我们给出信道的可排列性：

若矩阵的每一行都由同一符号集$\{p_1', p_2', \cdots, p_s'\}$中诸元素的不同排列组成，并且每一列也是由同一符号集$\{q_1', q_2', \cdots, q_r'\}$中诸元素的不同排列组成，则称这样的矩阵具有可排列性。

对于离散无记忆信道，输入随机变量 $X \in \{x_1, x_2, \cdots, x_r\}$，输出随机变量 $Y \in \{y_1, y_2, \cdots, y_s\}$，若其 $r \times s$ 的信道矩阵$[P(y|x)]$具有可排列性，则将这类信道称为对称信道。

【例 4.3】 已知有两个信道，其信道矩阵分别为

$$[P_1(y|x)] = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}, \quad [P_2(y|x)] = \begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{bmatrix}$$

由于这两个矩阵都具有可排列性，因此它们所表示的信道是对称信道。但是矩阵：

$$[P_3(y|x)] = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{3} \end{bmatrix}, \quad [P_4(y|x)] = \begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0.3 & 0.1 & 0.6 \end{bmatrix}$$

不具有可排列性，所以它们所表示的信道不是对称信道。

若输入符号和输出符号个数相同，都等于 r，且信道转移矩阵为

$$\boldsymbol{P} = \begin{bmatrix} \bar{p} & \frac{p}{r-1} & \cdots & \frac{p}{r-1} \\ \frac{p}{r-1} & \bar{p} & \cdots & \frac{p}{r-1} \\ \vdots & \vdots & & \vdots \\ \frac{p}{r-1} & \frac{p}{r-1} & \cdots & \bar{p} \end{bmatrix} \tag{4.19}$$

其中，$p+\bar{p}=1$，则称此信道为强对称信道或均匀信道。这类信道总的错误概率为 p，对称地分配给 $r-1$ 个输出符号，它是离散对称信道的一个特例，其信道矩阵中各列之和也等于 1。二元对称信道就是 $r=2$ 的强对称信道。

下面我们讨论离散对称信道的信道容量。

设信道的转移概率矩阵$[P(y|x)]_{[x]}$具有对称性，在考虑信道的输入 X 与输出 Y 之间的互信息时，设有一个对称信道，其输入 $X\in\{x_1, x_2, \cdots, x_r\}$，输出 $Y\in\{y_1, y_2, \cdots, y_s\}$，$r$ 行、s 列的信道矩阵$[P(y_j|x_i)]$具有对称性。

我们先分析一下此时条件熵的特点。

$$H(Y\mid X)=\sum_{i=1}^{r}\sum_{j=1}^{s}P(x_i)P(y_j\mid x_i)\log\frac{1}{P(y_j\mid x_i)}=\sum_{i=1}^{r}P(x_i)H(Y\mid x_i) \tag{4.20}$$

其中：

$$H(Y\mid x_i)=\sum_{j=1}^{s}P(y_j\mid x_i)\log\frac{1}{P(y_j\mid x_i)}\quad i=1, 2, \cdots, r \tag{4.21}$$

是由$[P(y_j|x_i)]$中第 i 行元素构成的熵函数。

因为$[P(y_j|x_i)]$中的每一行都由 $p_1', p_2', \cdots, p_s'$ 的不同排列所构成，所以若以 i 为参变量，则由对称性可知：

$$\begin{aligned}H(Y\mid x_i)&=\sum_{j=1}^{s}P(y_j\mid x_i)\log\frac{1}{P(y_j\mid x_i)}\\&=H(p_1', p_2', \cdots, p_s')\quad i=1, 2, \cdots, r\end{aligned}$$

是与 i 无关的常量。因此

$$H(Y\mid X)=\sum_{i=1}^{r}P(x_i)H(Y\mid x_i)=H(p_1', p_2', \cdots, p_s') \tag{4.22}$$

条件熵 $H(Y|X)$与信源 X 的分布 $P(x_i)$无关。

此时

$$I(X; Y)=H(Y)-H(Y\mid X)=H(Y)-H(p_1', p_2', \cdots, p_s')$$

根据信道容量的定义，应当有

$$C=\max_{P(x)}[H(Y)-H(p_1', p_2', \cdots, p_s')] \tag{4.23}$$

此时，当信源概率分布 $P(x)$改变时只有随机变量 Y 的熵 $H(Y)$随之改变，互信息 $I(X; Y)$最大值的计算便转化为求 $H(Y)$的最大值问题。

因为 $Y\in\{y_1, y_2, \cdots, y_s\}$，所以由最大熵定理有：

$$H(Y)\leqslant\log s$$

当且仅当 $P(y_j)=\dfrac{1}{s}$时等式成立。

因此，应当找出使 $P(y_j)=\dfrac{1}{s}$时信源的概率分布。

由对称信道的特点可知，$[P(y_j|x_i)]$的每一列都是由 $q_1', q_2', \cdots, q_r'$ 的不同排列构成的，即有：

$$\sum_{i=1}^{r}P(y_j\mid x_i)=\text{常量}\quad j=1, 2, \cdots, s$$

若信源 X 为等概率分布，即

$$P(x_i)=\frac{1}{r}\quad i=1,2,\cdots,r$$

则

$$P(y_j)=\sum_{i=1}^{r}P(x_i)P(y_j\mid x_i)=\sum_{i=1}^{r}\frac{1}{r}P(y_j\mid x_i)$$

$$=\frac{1}{r}\sum_{i=0}^{r}q_i'=\text{常数}\quad j=1,2,\cdots,s$$

由概率分布律可知，此时 $P(y_j)=\frac{1}{s}$，$j=1,2,\cdots,s$，所以当信源 X 为等概率分布时，对称信道具有最大的信息传输率。因此，对称信道的信道容量为

$$C=\log s-H(p_1',p_2',\cdots,p_s')\quad \text{比特/符号} \tag{4.24}$$

【例 4.4】 设某信道为强对称信道，其转移概率矩阵为式(4.19)。根据其信道矩阵的特点，可知其信道容量为

$$\begin{aligned}C&=\log r-H(p_1',p_2',\cdots,p_s')\\&=\log r-H\left(\bar{p},\frac{p}{r-1},\frac{p}{r-1},\cdots,\frac{p}{r-1}\right)\\&=\log r+\bar{p}\log\bar{p}+\frac{p}{r-1}\log\frac{p}{r-1}+\frac{p}{r-1}\log\frac{p}{r-1}+\cdots+\frac{p}{r-1}\log\frac{p}{r-1}\\&=\log r+\bar{p}\log\bar{p}+(r-1)\frac{p}{r-1}\log\frac{p}{r-1}\\&=\log r-p\log(r-1)+\bar{p}\log\bar{p}+p\log p\\&=\log r-p\log(r-1)-H(p)\end{aligned} \tag{4.25}$$

对于强对称信道，p 表示总的错误传输概率，而 $\bar{p}$ 表示正确传输概率。显然，强对称信道在其输入为等概率分布时达到其信道容量 C。由式(4.24)可知，对于 $r=2$ 的二进制对称信道，其信道容量为

$$C=1-H(p)=1+p\log p+(1-p)\log(1-p)\quad \text{比特/符号}$$

4.3.2　准对称信道的信道容量

设信道的输入、输出随机变量取值为：$X\in\{x_1,x_2,\cdots,x_r\}$，$Y\in\{y_1,y_2,\cdots,y_s\}$，已知 $r\times s$ 的信道概率转移矩阵 $\boldsymbol{P}=[P(y_j|x_i)]$不具有可排列性。

但是，如果$[p(y_j|x_i)]$的 s 列可以划分为 m 个不相交的子集($m<s$)，各子集分别有 $s_1,s_2,\cdots,s_m$ 个元(其中 $s_1+s_2+\cdots+s_m=s$)，由 s_k 列组成的子矩阵 $\boldsymbol{P}_k(k=1,2,\cdots,m)$都具有可排列性，则称

$$\boldsymbol{P}=[p(y_j\mid x_i)]\quad i=1,2,\cdots,r;\ j=1,2,\cdots,s$$

所表示的信道是准对称信道。

由准对称信道的定义可以看出，具有可排列性的子矩阵 $\boldsymbol{P}_k$ 是通过对 $\boldsymbol{P}$ 的划分而构成的，因此 $\boldsymbol{P}$ 的每一行是具有可排列性的，即 $\boldsymbol{P}$ 中的每一行元素都是由同一符号集 $\{p_1',p_2',\cdots,p_s'\}$的不同排列而构成的。

【例 4.5】 已知某信道的转移概率矩阵为

$$\boldsymbol{P}=[P(y_j\mid x_i)]=\begin{bmatrix}\frac{3}{8}&\frac{1}{4}&0&\frac{1}{8}&0&\frac{1}{4}\\ \frac{1}{8}&0&\frac{3}{8}&\frac{1}{4}&\frac{1}{4}&0\\ 0&\frac{1}{4}&\frac{1}{4}&\frac{3}{8}&0&\frac{1}{8}\\ \frac{1}{4}&0&\frac{1}{8}&0&\frac{1}{4}&\frac{3}{8}\end{bmatrix}$$

虽然 $\boldsymbol{P}$ 的每一行中的元素具有可排列性，但是列不满足其他列元素重排的要求。故矩阵 $\boldsymbol{P}$ 不具有可排列性，此信道不是对称信道。

但是若将此矩阵的六列划分为两个不相交的子集，即

$$s_1:\{2,\ 5\}$$

$$s_2:\{1,\ 3,\ 4,\ 6\}$$

则由每一子集构成的子矩阵：

$$\boldsymbol{P}_1=\begin{bmatrix}\frac{1}{4}&0\\ 0&\frac{1}{4}\\ \frac{1}{4}&0\\ 0&\frac{1}{4}\end{bmatrix},\ \boldsymbol{P}_2=\begin{bmatrix}\frac{3}{8}&0&\frac{1}{8}&\frac{1}{4}\\ \frac{1}{8}&\frac{3}{8}&\frac{1}{4}&0\\ 0&\frac{1}{4}&\frac{3}{8}&\frac{1}{8}\\ \frac{1}{4}&\frac{1}{8}&0&\frac{3}{8}\end{bmatrix}$$

都具有可排列性。

因此虽然这个信道不是对称信道，但是满足准对称信道的要求，是一个准对称信道。

由于准对称信道的信道矩阵 $\boldsymbol{P}$ 具有行可排列性，因此可以写出：

$$H(Y\mid X)=H(p_1',\ p_2',\ \cdots,\ p_s') \tag{4.26}$$

准对称信道的信道容量应满足：

$$\begin{aligned}C&=\max_{P(x)}[H(Y)-H(Y\mid X)]\\&=\max_{P(x)}H(Y)-H(p_1',\ p_2',\ \cdots,\ p_s')\end{aligned} \tag{4.27}$$

因为 $\boldsymbol{P}$ 中各列不满足可排列性，所以对于任意一种信源概率分布 $P(x)$，都不可能使信道输出 Y 呈现为等概率分布，$\max\limits_{P(x)}H(Y)$ 对应的信源概率分布 $P(x)$ 需要进一步分析才能得到。已知：

$$H(Y)=-\sum_{j=1}^{s}P(y_j)\log P(y_j)$$

根据前面矩阵划分的关系，我们也将求和式划分为 m 项，即

$$H(Y)=-\sum_{k=1}^{m}\sum_{j=1}^{s_k}P_k(y_j)\log P_k(y_j) \tag{4.28}$$

其中，$s_1+s_2+\cdots+s_m=s$。

对于式(4.28)中的第 k 项，集合 $s_k(k=1,\ 2,\ \cdots,\ m)$ 内的符号 $y_j(j=1,\ 2,\ \cdots,\ s_k)$ 发生概率的平均值为

$$P_k(y)=\frac{\sum_{j=1}^{s_k}P_k(y_j)}{s_k}$$

显然有

$$\frac{P_k(y)}{P_k(y_j)}>0$$

根据 $\ln x\leqslant x-1(x>0)$，有：

$$\begin{aligned}\sum_{j=1}^{s_k}P_k(y_j)\ln\frac{P_k(y)}{P_k(y_j)}&\leqslant\sum_{j=1}^{s_k}P_k(y_j)\left[\frac{P_k(y)}{P_k(y_j)}-1\right]\\&=\sum_{j=1}^{s_k}P_k(y)-\sum_{j=1}^{s_k}P_k(y_j)\\&=s_kP_k(y)-s_kP_k(y)=0\end{aligned}$$

即

$$\sum_{j=1}^{s_k}P_k(y_j)\ln P_k(y)-\sum_{j=1}^{s_k}P_k(y_j)\ln P_k(y_j)\leqslant 0$$

$$-\sum_{j=1}^{s_k}P_k(y_j)\ln P_k(y_j)\leqslant-\sum_{j=1}^{s_k}P_k(y_j)\ln P_k(y)$$

$$=-s_kP_k(y)\ln P_k(y)$$

如果每一个子集的信道输出符号发生的概率相同，则 $-\sum_{j=1}^{s_k}P_k(y_j)\log P_k(y_j)$ 可达到其最大值 $-s_kP_k(y)\ln P_k(y)$。

因为子矩阵 $\boldsymbol{P}_k$ 是可排列的，即每一行元素均由 $\{p_1',p_2',\cdots,p_{s_k}'\}$ 的不同排列构成，而且每一列元素也是由同一集合 $\{q_1',q_2',\cdots,q_r'\}$ 的不同排列构成的，所以当信源符号 X 为等概率分布，即

$$P(x_1)=P(x_2)=\cdots=P(x_r)=\frac{1}{r}$$

时，一定能将使子集合 s_k 中的输出符号 y_j 具有相同的概率分布，即式(4.28)中的每一内和式均达到其最大值。

于是我们得到了准对称信道的信道容量：

$$C=[H(Y)-H(Y\mid X)]_{P(x)=\frac{1}{r}}\tag{4.29}$$

即准对称信道的信道容量等于等概率输入的互信息。

【例 4.6】　已知某信道的转移概率如表 4-1 所示，计算此信道的信道容量。

表 4-1　$P(y|x)$

X \ Y	y_1	y_2	y_3	y_4
x_1	$\frac{1}{2}$	0	$\frac{1}{6}$	$\frac{1}{3}$
x_2	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	0

解：此信道的信道矩阵为

$$\boldsymbol{P}=[P(y_j\mid x_i)]=\begin{bmatrix}\frac{1}{2} & 0 & \frac{1}{6} & \frac{1}{3}\\ \frac{1}{6} & \frac{1}{3} & \frac{1}{2} & 0\end{bmatrix}$$

$\boldsymbol{P}$ 不满足可排列性，故此信道不是对称信道。

但是若将矩阵的四个列分为两个不相交的子集：

$$s_1:\{1,3\}$$
$$s_2:\{2,4\}$$

则由每一子集构成的子矩阵：

$$\boldsymbol{P}_1=\begin{bmatrix}\frac{1}{2} & \frac{1}{6}\\ \frac{1}{6} & \frac{1}{2}\end{bmatrix},\quad \boldsymbol{P}_2=\begin{bmatrix}0 & \frac{1}{3}\\ \frac{1}{3} & 0\end{bmatrix}$$

都是可排列的。由此可知，此信道是准对称信道。

根据准对称信道的信号的信道容量计算关系，取输入随机变量 X 为等概率分布，即 $P(x_1)=P(x_2)=\frac{1}{2}$，此时$(x,y)$的联合概率分布及 Y 的边缘概率分布如表 4-2 所示。

表 4-2 $P(x,y)$

X \ Y	y_1	y_2	y_3	y_4
x_1	$\frac{1}{4}$	0	$\frac{1}{12}$	$\frac{1}{6}$
x_2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{4}$	0
$P(y_j)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$

由 Y 的边缘概率分布可看出，在输入 X 等概率分布时，虽然准对称信道的输出 Y 不是等概率分布，但是对于每个子集 s_1 和 s_2，子集中的符号是等概率的。由前面的分析可知，每个子集符号的概率均相等，保证了准对称信道输出随机变量 Y 的熵 $H(Y)$达到其最大量。此时为

$$H(Y)=H\left(\frac{1}{3},\frac{1}{6},\frac{1}{3},\frac{1}{6}\right)=\frac{2}{3}\log 3+\frac{2}{6}\log 6=\frac{1}{3}+\log 3$$

又可求出

$$\begin{aligned}H(Y\mid X)&=\sum_{i=1}^{2}P(x_i)\sum_{j=1}^{4}P(y_j\mid x_i)\log\frac{1}{P(y_j\mid x_i)}\\&=\sum_{j=1}^{4}P(y_j\mid x_i)\log\frac{1}{P(y_j\mid x_i)}\sum_{i=1}^{2}P(x_i)\\&=H\left(\frac{1}{2},\frac{1}{6},\frac{1}{3}\right)\times 1\\&=\frac{2}{3}+\frac{1}{2}\log 3\end{aligned}$$

因此，此准对称信道的信道容量为

$$\begin{aligned} C &= [H(Y) + H(Y \mid X)]_{P(X)=\frac{1}{2}} \\ &= \frac{1}{3} + \log 3 - \frac{2}{3} - \frac{1}{2}\log 3 \\ &= -\frac{1}{3} + \frac{1}{2}\log 3 \approx 0.459 \quad \text{比特 / 符号} \end{aligned}$$

4.3.3　可逆矩阵信道的信道容量

如果信道的输入和输出有相同的符号个数，即

$$X \in \{x_1, x_2, \cdots, x_n\}$$

$$Y \in \{y_1, y_2, \cdots, y_n\}$$

且信道的矩阵$[P(y_j \mid x_i)] = \boldsymbol{P}$的逆矩阵$\boldsymbol{P}^{-1}$存在，则此信道称为可逆矩阵信道。

对于给定的信道，输入概率分布$P(x_i)$改变时平均互信息$I(X; Y)$也将随之改变。由于$I(X; Y)$是输入概率分布的上凸函数，因此一定存在一个极值点，使$I(X; Y)$达到其最大值。根据信道容量的定义，这个极大值就是给定信道的信道容量。

对于可逆矩阵信道，可以在给定的条件下运用拉格朗日乘子法计算互信息$I(X; Y)$的条件极值求出其信道容量。

设给定一个可逆矩阵信道，$n \times n$的信道矩阵为

$$\boldsymbol{P} = [P(y_j \mid x_i)] \qquad i, j = 1, 2, \cdots, n$$

为求出$I(X; Y)$在$\sum_{i=1}^{n} P(x_i) = 1$条件下的极值，构造一个辅助函数$\varphi$：

$$\begin{aligned} \varphi &= I(X; Y, P(x_k)) + \lambda\left(1 - \sum_{k=1}^{n} P(x_k)\right) \\ &= H(Y) - H(Y \mid X) + \lambda\left(1 - \sum_{k=1}^{n} P(x_k)\right) \\ &= -\sum_{j=1}^{n} P(y_j)\log P(y_j) \\ &\quad + \sum_{k=1}^{n}\sum_{j=1}^{n} P(x_k)P(y_j \mid x_k)\log P(y_j \mid x_k) + \lambda\left(1 - \sum_{k=1}^{n} P(x_k)\right) \end{aligned}$$

因为$P(y_j) = \sum_{k=1}^{n} P(x_k)P(y_j \mid x_k)$，$j = 1, 2, \cdots, n$，所以$P(y_j)$也是输入概率分布$P(x_i)$的函数，并且有偏微分：

$$\frac{\partial P(y_j)}{\partial P(x_i)} = P(y_j \mid x_i)$$

$$\begin{aligned} \frac{\partial}{\partial P(x_i)}\log P(y_j) &= \frac{1}{\ln 2 \cdot P(y_j)}\frac{\partial P(y_j)}{\partial P(x_i)} \\ &= \frac{P(y_j \mid x_i)}{\ln 2 \cdot P(y_j)} \quad i = 1, 2, \cdots, n \end{aligned}$$

为了计算互信息的条件极值，将辅助函数对输入概率分布$P(x_i)$求偏导，则有：

$$\frac{\partial \varphi}{\partial P(x_i)}=-\sum_{j=1}^{n}\frac{\partial P(y_j)}{\partial P(x_i)}\log P(y_j)-\sum_{j=1}^{n}P(y_j)\frac{\partial}{\partial P(x_i)}\log P(y_j)$$

$$+\sum_{j=1}^{n}P(y_j\mid x_i)\log P(y_j\mid x_i)-\lambda$$

$$=-\sum_{j=1}^{n}P(y_j\mid x_i)\log P(y_j)-\sum_{j=1}^{n}P(y_j)\frac{P(y_j\mid x_i)}{\ln 2\cdot P(y_j)}+\sum_{j=1}^{n}P(y_j\mid x_i)\log P(y_j\mid x_i)-\lambda$$

$$=\sum_{j=1}^{n}P(y_j\mid x_i)\log\frac{P(y_j\mid x_i)}{P(y_j)}-\frac{1}{\ln 2}\sum_{j=1}^{n}P(y_j\mid x_i)-\lambda$$

$$=\sum_{j=1}^{n}P(y_j\mid x_i)\log\frac{P(y_j\mid x_i)}{P(y_j)}-\frac{1}{\ln 2}-\lambda\quad i=1,2,\cdots,n$$

因为

$$\ln 2=\frac{\log_2 2}{\log_2 \mathrm{e}}=\frac{1}{\log_2 \mathrm{e}}=\frac{1}{\log \mathrm{e}}$$

所以

$$\frac{\partial \varphi}{\partial P(x_i)}=\sum_{j=1}^{n}P(y_j\mid x_i)\log\frac{P(y_j\mid x_i)}{P(y_j)}-\log \mathrm{e}-\lambda$$

为了求出条件极值，令

$$\frac{\partial \varphi}{\partial P(x_i)}=0$$

即

$$\sum_{j=1}^{n}P(y_j\mid x_i)\log\frac{P(y_j\mid x_i)}{P(y_j)}=\log \mathrm{e}+\lambda\quad i=1,2,\cdots,n \tag{4.30}$$

显然，式(4.30)为一个由 n 个方程构成的方程组。求解此方程组可以得到 $I(X;Y)$ 的条件极值点，即使 $I(X;Y)$ 在满足 $\sum_{i=1}^{n}P(x_i)=1$ 的条件下达到其最大值的一组输入极值点概率分布 $P(x_1)$，$P(x_2)$，…，$P(x_n)$。在这组极值点概率分布下，互信息 $I(X;Y)$ 即为给定信道的信道容量。

下面我们计算在这组极值点输入概率分布下的互信息 $I(X;Y)$。

对于式(4.30)表示的由 n 个方程构成的方程组，将第 i 个方程的两端同乘以极值点概率分布 $P(x_i)$，即

$$\sum_{j=1}^{n}P(x_i)P(y_j\mid x_i)\log\frac{P(y_j\mid x_i)}{P(y_j)}=(\log \mathrm{e}+\lambda)\cdot P(x_i)\quad i=1,2,\cdots,n \tag{4.31}$$

将个方程的左、右两端分别相加(相当于将式(4.31)的左、右两端分别对 $i=1,2,\cdots,n$ 求和)，即

$$\sum_{i=1}^{n}\sum_{j=1}^{n}P(x_i)P(y_j\mid x_i)\log\frac{P(y_j\mid x_i)}{P(y_j)}=(\log \mathrm{e}+\lambda)\sum_{i=1}^{n}P(x_i)$$

$$=\log \mathrm{e}+\lambda \tag{4.32}$$

显然，式(4.32)左端即为互信息 $I(X;Y)$ 的表达式。由于此时的互信息 $I(X;Y)$ 是在使其达到最大值的极值点概率分布下得到的，即式(4.32)的左端为

$$\max_{P(x_i)} I(X;Y,p)$$

因此，由信道容量的定义可知，所给信道的信道容量满足下面关系式：

$$C=\log \mathrm{e}+\lambda \tag{4.33}$$

在这个信道容量关系式中，由于 λ 为待定常数，因此我们还不能由此得到信道容量的值，需要利用给定信道的特性作进一步的分布和推导。

对于给定的信道，已知 X、Y 的离散集合有相同数目的元，信道转移概率构成的信道矩阵为一个方阵。在下面的分析中，将信道矩阵简单记做：

$$\boldsymbol{P} = [P(y_j \mid x_i)] = [p_{ij}]$$

其中，$P(y_j \mid x_i) = p_{ij}$ 表示信道矩阵的第 i 行、第 j 列元素。

由于此信道是可逆矩阵信道，因此信道矩阵 $\boldsymbol{P}$ 的逆矩阵 $\boldsymbol{P}^{-1}$ 存在。我们将该信道矩阵的 $n \times n$ 逆矩阵表示为

$$\boldsymbol{P}^{-1} = \boldsymbol{R} = [r_{kj}] \quad k, j = 1, 2, \cdots, n$$

其中，r_{kj} 为矩阵中的第 k 行、第 j 列元素。

已知 $\boldsymbol{R}$ 与 $\boldsymbol{P}$ 是互逆的，它们满足：

$$\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{mn} \end{bmatrix} \cdot \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{mn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

由此互逆关系可以看出，$\boldsymbol{R}$ 的第 k 行与 $\boldsymbol{P}$ 的第 j 列对应元素的乘积并求和为

$$\sum_{i=1}^{n} r_{ki} p_{ij} = \delta_{kj} = \begin{cases} 1 & k = j \\ 0 & k \neq j \end{cases}$$

依据概率分布律应有：

$$\sum_{j=1}^{n} p_{ij} = 1$$

因此

$$\sum_{i=1}^{n} r_{ki} = \left(\sum_{i=1}^{n} r_{ki}\right)\left(\sum_{j=1}^{n} p_{ij}\right) = \sum_{j=1}^{n}\sum_{i=1}^{n} r_{ki} p_{ij} = \sum_{j=1}^{n} \delta_{kj} = 1$$

应用给定可逆矩阵信道的信道矩阵及其逆矩阵具有的这些关系，我们可做进一步分析。将式(4.30)的两端同乘以 r_{ki}，即有：

$$r_{ki} \sum_{j=1}^{n} P(y_j \mid x_i) \log \frac{P(y_j \mid x_i)}{P(y_j)} = (\log \mathrm{e} + \lambda) r_{ki} \quad i = 1, 2, \cdots, n$$

上式左、右两端分别对 $i=1, 2, \cdots, n$ 求和：

$$\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ki} P(y_j \mid x_i) \log \frac{P(y_j \mid x_i)}{P(y_j)} = (\log \mathrm{e} + \lambda) \cdot \sum_{i=1}^{n} r_{ki} = \log \mathrm{e} + \lambda$$

根据式(4.33)所给出的关系，可得到：

$$C = \sum_{i=1}^{n}\sum_{j=1}^{n} r_{ki} P(y_j \mid x_i) \log P(y_j \mid x_i) - \sum_{i=1}^{n}\sum_{j=1}^{n} r_{ki} P(y_j \mid x_i) \log P(y_j) \tag{4.34}$$

在式(4.34)中，只有第二项中的 $P(y_j)$ 是未知量。应用上述可逆矩阵信道的关系，该式第二项为

$$\begin{aligned} \sum_{i=1}^{n}\sum_{j=1}^{n} r_{ki} P(y_j \mid x_i) \log P(y_j) &= \sum_{j=1}^{n}\left(\sum_{i=1}^{n} r_{ki} P_{ij}\right) \log P(y_j) \\ &= \sum_{j=1}^{n} \delta_{kj} \log P(y_j) = \log P(y_k) \end{aligned}$$

则式(4.34)为

$$\log P(y_k) = -C + \sum_{i=1}^{n}\sum_{j=1}^{n} r_{ki} P(y_j \mid x_i)\log P(y_j \mid x_i)$$

此处我们取对数底为 2，并对上式两端分别取以 2 为底的指数运算，便可得到输出 Y 的概率分布，即

$$P(y_k) = 2^{-C+\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ki} P(y_j|x_i)\log P(y_j|x_i)} = 2^{-C} \times 2^{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ki} P(y_j|x_i)\log P(y_j|x_i)} \tag{4.35}$$

其中，$k=1, 2, \cdots, n$。

式(4.35)也是一个由 n 个方程构成的方程组。若将这 n 个方程的左、右两端分别对 $k=1, 2, \cdots, n$ 求和：

$$\sum_{k=1}^{n} P(y_k) = \sum_{k=1}^{n} 2^{-C} \times 2^{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ki} P(y_j|x_i)\log P(y_j|x_i)}$$

则得到：

$$1 = 2^{-C} \times \sum_{k=1}^{n} 2^{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ki} P(y_j|x_i)\log P(y_j|x_i)}$$

$$2^{C} = \sum_{k=1}^{n} 2^{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ki} P(y_j|x_i)\log P(y_j|x_i)}$$

取以 2 为底的对数，得到：

$$C = \log \sum_{k=1}^{n} 2^{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ki} P(y_j|x_i)\log P(y_j|x_i)} \tag{4.36}$$

此即为给定的可逆矩阵信道的信道容量，单位为比特/符号。

【例 4.7】 计算图 4-7 所示信道的信道容量。

解：由图 4-7 中所示信道转移概率关系，可写出信道的信道矩阵为

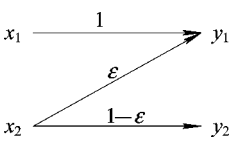

图 4-7

$$\boldsymbol{P} = \begin{bmatrix} 1 & 0 \\ \varepsilon & 1-\varepsilon \end{bmatrix}$$

首先，我们判断此信道是否为可逆矩阵信道，因

$$|\boldsymbol{P}| = \begin{vmatrix} 1 & 0 \\ \varepsilon & 1-\varepsilon \end{vmatrix} = 1-\varepsilon$$

其伴随矩阵为

$$\boldsymbol{P}^{*} = \begin{bmatrix} 1-\varepsilon & 0 \\ -\varepsilon & 1 \end{bmatrix}$$

于是得到信道矩阵 $\boldsymbol{P}$ 的逆阵为

$$\boldsymbol{R} = \boldsymbol{P}^{-1} = \frac{\boldsymbol{P}^{*}}{|\boldsymbol{P}|} = \begin{bmatrix} 1 & 0 \\ -\dfrac{\varepsilon}{1-\varepsilon} & \dfrac{1}{1-\varepsilon} \end{bmatrix}$$

由于此信道矩阵存在，因此此信道是可逆矩阵信道。其信道容量可以由式(4.36)给出的计算关系得到：

$$
\begin{aligned}
A(k=1) &= \sum_{i=1}^{2}\sum_{j=1}^{2} r_{1i}P(y_j \mid x_i)\log P(y_j \mid x_i) \\
&= r_{11}\left[P(y_1 \mid x_1)\log P(y_1 \mid x_1) + P(y_2 \mid x_1)\log P(y_2 \mid x_1)\right] \\
&\quad + r_{12}\left[P(y_1 \mid x_2)\log P(y_1 \mid x_2) + P(y_2 \mid x_2)\log P(y_2 \mid x_2)\right] \\
&= -1 \times H(1,\ 0) - 0 \times H(\varepsilon,\ 1-\varepsilon) = 0 \\
A(k=2) &= \sum_{i=1}^{2}\sum_{j=1}^{2} r_{2i}P(y_j \mid x_i)\log P(y_j \mid x_i) \\
&= -r_{21} \cdot H(1,\ 0) - r_{22}H(\varepsilon,\ 1-\varepsilon) \\
&= \frac{\varepsilon}{1-\varepsilon} \cdot 0 + \frac{1}{1-\varepsilon}\left[\varepsilon\log\varepsilon + (1-\varepsilon)\log(1-\varepsilon)\right] \\
&= \log\left[\varepsilon^{\frac{\varepsilon}{1-\varepsilon}}(1-\varepsilon)\right]
\end{aligned}
$$

因此

$$
\begin{aligned}
C &= \log\Big(\sum_{k=1}^{2} 2^{\sum_{i=1}^{2}\sum_{j=1}^{2} r_{ki}P(y_j \mid x_i)\log P(y_j \mid x_i)}\Big) = \log(2^{A(k=1)} + 2^{A(k=2)}) \\
&= \log\left(2^0 + 2^{\log\left(\varepsilon^{\frac{\varepsilon}{1-\varepsilon}}(1-\varepsilon)\right)}\right) \\
&= \log\left(1 + (1-\varepsilon)\varepsilon^{\frac{\varepsilon}{1-\varepsilon}}\right)
\end{aligned}
$$

此即所给可逆矩阵信道的信道容量。

对于可逆矩阵信道，在使用式(4.36)求出其信道容量 C 之后，由式(4.35)可以求出输出符号 Y 的概率分布，即

$$
P(y_k) = 2^{-C} \times 2^{\sum_{i=1}^{n}\sum_{j=1}^{n} r_{ki}P(y_j \mid x_i)\log P(y_j \mid x_i)} \qquad k = 1,\ 2,\ \cdots,\ n
$$

再利用 $P(y_j) = \sum_{i=1}^{n} P(x_i)P(y_j \mid x_i)$，$j = 1,\ 2,\ \cdots,\ n$ 的关系，可以求出在此时的信道容量 C，即互信息 $I(X;\ Y)$ 为最大值时所对应的输入随机变量 X 的概率分布 $P(x_i)$，$i=1,2,\cdots,n$。

如果求出的输入概率分布 $P(x_i)$ 满足概率分布率要求，即

$$
0 \leqslant P(x_i) \leqslant 1 \quad \left(\sum_{i=1}^{n} P(x_i) = 1 \text{ 的条件在辅助函数的构造中已经考虑}\right)
$$

则表明上述计算是合理的，得出的信道容量 C 有效。

但是，在可逆信道的信道容量计算中，辅助函数 φ 的构造并未对 $0 \leqslant P(x_i) \leqslant 1$，$i=1,\ 2,\ \cdots,\ n$ 加以限制，计算信道容量 C 后得出的输入概率分布中可能会有某些输入符号的概率小于零。这表明在 $\sum_{i=1}^{n} P(x_i) = 1$ 的条件下，$I(X;\ Y)$ 极大值所对应的输入概率分布不满足概率分布律的要求，所得到的 C 并不是给定信道的信道容量，采用这种信道容量的计算方法失效。

上面我们对几种特殊的信道讨论了其信道容量的计算方法。对于更一般的情况，信道矩阵不一定是可逆的，信道的输入与输出符号的数目也并不一定相同。因此这些信道容量的计算方法所适用的范围有明显的限制。

对于一般的信道，信道容量的计算是非常复杂的。这种计算的复杂性由信道容量的

定义：

$$C = \max_{P(x_i)} I(X; Y, P(x_i))$$

即可看出。

这种最大值从互信息 $I(X; Y)$ 与输入概率分布的上凸关系来看一定存在，但对于一个任意的信道，在 $I(X; Y)$ 与 $P(x_1)$，$P(x_2)$，…，$P(x_n)$ 所构成的多元关系中，根据给定信道的特点求出这个最大的互信息显然是困难的，几乎不易得出正确的答案。

4.4 离散无记忆信道容量的迭代算法

对于一般的离散无记忆信道而言，信道容量的计算比较复杂，可以用迭代算法实现。

设离散无记忆信道的输入符号集和相应分布为

$$X = \{x_1, x_2, \cdots, x_r\} \quad \{P(x_i), i = 1, 2, \cdots, r\}$$

输出符号集和相应分布为

$$Y = \{y_1, y_2, \cdots, y_s\} \quad \{P(y_j), j = 1, 2, \cdots, s\}$$

信道转移概率矩阵为

$$\boldsymbol{P} = \{P(y_j \mid x_i), i = 1, 2, \cdots, r; j = 1, 2, \cdots, s\}$$

信道示意图如图 4-8 所示。

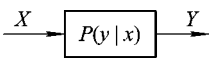

图 4-8 信道模型

信道容量：

$$C = \max_{[X, P]} I(X; Y)$$

其中：

$$I(X; Y) = \sum_j \sum_i P(x_i)P(y_j \mid x_i)\log \frac{P(y_j \mid x_i)}{P(y_j)}$$

$$= \sum_j \sum_i P(x_i)P(y_j \mid x_i)\log \frac{Q(x_i \mid y_j)}{P(x_i)} \tag{4.37}$$

$$P(y_j) = \sum_{i=1}^{r} P(x_i) \cdot P(y_j \mid x_i) \tag{4.38}$$

$$Q(x_i \mid y_j) = \frac{P(x_i)P(y_j \mid x_i)}{P(y_j)} = \frac{P(x_i)P(y_j \mid x_i)}{\sum_{k=1}^{r} P(x_k) \cdot P(y_j \mid x_k)} \tag{4.39}$$

其中，$Q(x_i \mid y_j)(i=1, 2, \cdots, r; j=1, 2, \cdots, s)$ 称为反条件概率，由 $[X, P]$ 和信道转移概率矩阵确定，可以设想为原信道的反向信道，如图 4-9 所示。

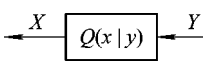

图 4-9 反向信道

该反向信道的转移概率矩阵为

$$\{Q(x_i \mid y_j)\} \quad i = 1, 2, \cdots, r; j = 1, 2, \cdots, s$$

且满足：

$$Q(x_i \mid y_j) \geqslant 0 \quad j = 1, 2, \cdots, s; j = 1, 2, \cdots, r$$

$$\sum_{i=1}^{r} Q(x_i \mid y_j) = 1 \quad j = 1, 2, \cdots, s$$

引入反向信道的概念后，平均互信息量 $I(X; Y)$可以看做是输入概率分布 $P(x)$和反向信道转移概率 $P(x|y)$的函数，可记为：$F\{P(x), P(x|y)\}$。

当 $P(y)$给定时，$F\{P(x), P(x|y)\}$达到最大的分布 $P(x|y)$由下述定理确定。

定理 4.1　对任意 $P(x|y)$，有

$$I(X; Y) \geqslant F\{P(x), P(x \mid y)\} \tag{4.40}$$

当且仅当

$$P(x_i \mid y_j) = \frac{P(x_i)P(y_j \mid x_i)}{P(y_j)} \quad i = 1, 2, \cdots, r; \ j = 1, 2, \cdots, s$$

时等式成立。

证明：令

$$P^*(x_i \mid y_j) = \frac{P(x_i)P(y_j \mid x_i)}{P(y_j)} Q(x_i \mid y_j) \quad i = 1, 2, \cdots, r; \ j = 1, 2, \cdots, s$$

则

$$\sum_i P(x_i \mid y_j) \log P(x_i \mid y_j) \leqslant \sum_i P(x_i \mid y_j) \log Q(x_i \mid y_j) \quad j = 1, 2, \cdots, s$$

当且仅当

$$P(x_i \mid y_j) = Q(x_i \mid y_j) \quad i = 1, 2, \cdots, r; \ j = 1, 2, \cdots, s$$

时等式成立，而

$$I(X; Y) = \sum_j \sum_i P(y_j) Q(x_i \mid y_j) \log \frac{Q(x_i \mid y_j)}{P(x_i)}$$

因此有

$$\begin{aligned} & F\{P(x), P(x \mid y)\} - I(X; Y) \\ &= \sum_j P(y_j) \left\{ \sum_i P(y_j \mid x_i) \log \frac{P(x_i \mid y_j)}{P(x_i)} - \sum_i Q(y_j \mid x_i) \log \frac{Q(x_i \mid y_j)}{P(x_i)} \right\} \leqslant 0 \end{aligned}$$

证毕。

上述定理说明：在给定输入分布 $P(x)$的条件下，可求得使 $F\{P(y), P(x|y)\}$达到最大值的反向转移概率分布 $P^*(x_i|y_j)(i=1, 2, \cdots, r; \ j=1, 2, \cdots, s)$，即

$$I(X; Y) = \max_{P(x|y)} F\{P(x), P(x \mid y)\} = F\{P(x), P^*(x \mid y)\}$$

最佳分布为

$$P^*(x_i \mid y_j) = Q(x_i \mid y_j) \quad i = 1, 2, \cdots, r; \ j = 1, 2, \cdots, s$$

因此

$$C = \max_{P(x)} \{ \max_{P(x|y)} F[P(x), P(x \mid y)] \} \tag{4.41}$$

固定转移概率 $P(x|y)$时，$F\{P(x), P(x|y)\}$可以对输入分布 $P(x)$求最大值。一般而言，信道容量 C 是在

$$P(x_i) \geqslant 0 \quad i = 1, 2, \cdots, r$$

$$\sum_{i=1}^{r} P(x_i) = 1$$

条件下求 $I(X; Y)$的最大值，故可用 Lagrange 乘子法。

定义函数：

$$f(P(x)) = I(X;Y) - \lambda \sum_{i=1}^{r} P(x_i)$$

显然，$f(P(x))$**和** $I(X;Y)$**具有相同的极大值**。

$$\begin{aligned}\frac{\partial}{\partial P(x_i)} f(P(x)) &= \frac{\partial}{\partial P(x_i)} I(X;Y) - \lambda \\ &= \sum_{j=1}^{s} P(y_j \mid x_i) \log P(y_j \mid x_i) \\ &\quad - \sum_{j=1}^{s} \sum_{k=1}^{r} \frac{\partial}{\partial P(x_i)} [P(x_k) P(y_j \mid x_k) \log P(y_j)] - \lambda\end{aligned}$$

又因为

$$\begin{aligned}\frac{\partial}{\partial P(x_i)} \log P(y_j) &= \left\{ \frac{\partial}{\partial P(x_i)} \ln P(y_j) \right\} \log \mathrm{e} \\ &= \left\{ \frac{\partial}{\partial P(x_i)} \ln \sum_{k=1}^{r} P(x_k) P(y_j \mid x_k) \right\} \log \mathrm{e} \\ &= \frac{P(y_j \mid x_i)}{P(y_j)} \log \mathrm{e}\end{aligned}$$

所以

$$\begin{aligned}&\sum_{j=1}^{s} \sum_{k=1}^{r} \frac{\partial}{\partial P(x_i)} [P(x_k) P(y_j \mid x_k) \log P(y_j)] \\ &= \sum_{j=1}^{s} \frac{\partial}{\partial P(x_i)} [P(x_i) P(y_j \mid x_i) \log P(y_j)] + \sum_{j=1}^{s} \sum_{k=1,\, k \neq i}^{r} \frac{\partial}{\partial P(x_i)} [P(x_k) P(y_j \mid x_k) \log P(y_j)] \\ &= \sum_{j=1}^{s} [P(y_j \mid x_i) \log P(y_j)] + \sum_{j=1}^{s} \sum_{k=1,\, k \neq i}^{r} \frac{\partial}{\partial P(x_i)} [P(x_k) P(y_j \mid x_k) \log P(y_j)] \\ &= \sum_{j=1}^{s} [P(y_j \mid x_i) \log P(y_j)] + \sum_{j=1}^{s} \left[P(x_i) P(y_j \mid x_i) \frac{\partial}{\partial P(x_i)} \log P(y_j) \right] \\ &\quad + \sum_{j=1}^{s} \sum_{k=1,\, k \neq i}^{r} \left[P(x_k) P(y_j \mid x_k) \frac{\partial}{\partial P(x_i)} \log P(y_j) \right]\end{aligned}$$

原式 $= \sum_{j=1}^{s} P(y_j \mid x_i) \log \dfrac{P(y_j \mid x_i)}{P(y_j)} - \sum_{j=1}^{s} \sum_{k=1}^{r} P(x_k) P(y_j \mid x_k) \dfrac{P(y_j \mid x_i)}{P(y_j)} \log \mathrm{e} - \lambda$

其中，$i=1, 2, \cdots, r$。**考虑** $\sum_{k=1}^{r} P(x_k) P(y_j \mid x_k) = P(y_j)$，**令偏导数等于** 0 **得**：

$$\sum_{j=1}^{s} P(y_j \mid x_i) \log \frac{P(y_j \mid x_i)}{P(y_j)} = \log \mathrm{e} + \lambda \quad i = 1, 2, \cdots, r \tag{4.42}$$

设该方程组有一组解：$P^*(x_1)$，$P^*(x_2)$，$\cdots$，$P^*(x_r)$，**满足** $\sum_{i=1}^{r} P^*(x_i) = 1$，**求得极值为**

$$C = \lambda + \log \mathrm{e} \tag{4.43}$$

定理 4.2　当且仅当 $P(X)$**满足**

$$\frac{\partial}{\partial P(x_i)} g(P(x)) = \lambda \qquad i = 1, 2, \cdots, r \tag{4.44}$$

时，$g(P(x)) = I(X;Y)$**达到最大值**（**信道容量**）。

证明：

$$\frac{\partial}{\partial P(x_i)}g(P(x)) = \sum_{j=1}^{s} P(y_j \mid x_i)\log\frac{P(y_j \mid x_i)}{P(y_j)} - \log e \quad i = 1, 2, \cdots, r \tag{4.45}$$

先证充分性。

由于

$$\frac{\partial}{\partial P(x_i)}g(P(x)) = \lambda \qquad i = 1, 2, \cdots, r$$

即

$$\sum_{j=1}^{s} P(y_j \mid x_i)\log\frac{P(y_j \mid x_i)}{P(y_j)} = \log e + \lambda \quad i = 1, 2, \cdots, r$$

满足式(4.42)，**故** $P(x)$**为极值点** $P^*(x)$，**得证**。

再证必要性。

由于 $P(x)$**为极值点，因此它为** Lagrange **乘子法求偏导数的方程组中的一个解，式**(4.42)**成立，从而有**

$$\sum_{j=1}^{s} P(y_j \mid x_i)\log\frac{P(y_j \mid x_i)}{P(y_j)} = \log e + \lambda \qquad i = 1, 2, \cdots, r$$

从而式(4.44)**成立**。

同理可证，据式(4.37)，**形式上固定** $P(x|y)$，**分布** $P(x)$**使** $F(P(x), P(x|y))$**达到最大值的充要条件为**

$$\frac{\partial}{\partial P(x_i)}F(P(x), P(x \mid y)) = \lambda \quad i = 1, 2, \cdots, r$$

而

$$\begin{aligned}
\lambda &= \frac{\partial}{\partial P(x_i)}F(P(x), P(x \mid y)) \\
&= \frac{\partial}{\partial P(x_i)}\Big\{\sum_j \sum_k P(x_k)P(y_j \mid x_k)[\log P(x_k \mid y_j) - \log P(x_k)]\Big\} \\
&= \sum_j P(y_j \mid x_i)\log P(x_i \mid y_j) - \sum_j P(y_j \mid x_i)[\log P(x_i) + \log e] \\
&= \sum_j P(y_j \mid x_i)\log P(x_i \mid y_j) - \log P(x_i) - \log e
\end{aligned}$$

不妨取 log **的底为** e，**则**

$$\begin{aligned}
\ln P(x_i) &= \sum_j P(y_j \mid x_i)\ln P(x_i \mid y_j) - \lambda - 1 \\
P(x_i) &= e^{-(\lambda+1)} \times e^{\sum_j P(y_j|x_i)\ln P(x_i|y_j)}
\end{aligned} \tag{4.46}$$

由于约束条件：

$$\sum_i P(x_i) = 1$$

因此

$$P(x_i) = \frac{\exp\Big\{\sum_j P(y_j \mid x_i)\ln P(x_i \mid y_j)\Big\}}{\sum_i \exp\Big\{\sum_j P(y_j \mid x_i)\ln P(x_i \mid y_j)\Big\}} \tag{4.47}$$

这样，通过交替地固定 $P(x)$**和** $P(x|y)$，**利用式**(4.39)**和式**(4.47)**就可以求得使** $F(P(x), P(x|y))$**为最大的分布，从而有以下的迭代算法**。

设初始分布为 $P^{(0)}(x)$，一般可取等概率分布，即

$$P^{(0)}(x_i)=\frac{1}{r}\quad i=1,2,\cdots,r \tag{4.48}$$

令 $k=0,1,2,\cdots$ 为迭代序号，则迭代公式为

$$P^{(k)}(x_i\mid y_j)=\frac{P(y_j\mid x_i)P^{(k)}(x_i)}{\sum_l P(y_j\mid x_l)P^{(k)}(x_l)}\quad j=1,2,\cdots,s;\ i=1,2,\cdots,r \tag{4.49}$$

$$P^{(k+1)}(x_i)=\frac{\exp\left\{\sum_j P(y_j\mid x_i)\ln P^{(k)}(x_i\mid y_j)\right\}}{\sum_i \exp\left\{\sum_j P(y_j\mid x_i)\ln P^{(k)}(x_i\mid y_j)\right\}}\quad i=1,2,\cdots,r \tag{4.50}$$

$$C^{(k+1)}=F(P^{(k+1)}(x),P^{(k)}(x\mid y)) \tag{4.51}$$

迭代步骤如下：

(1) 取初始分布 $P^{(0)}(x)$。

(2) 根据式(4.49)计算 $P^{(k)}(x_i|y_j)(i=1,2,\cdots,r;\ j=1,2,\cdots,s)$。

(3) 根据式(4.50)计算 $P^{(k+1)}(x_i)(i=1,2,\cdots,r)$。

(4) 根据式(4.51)计算 $C^{(k+1)}$。

(5) 若 $|C^{(k+1)}-C^{(k)}|<\delta$，则转向步骤(7)。

(6) 令 $k=k+1$，转向步骤(2)。

(7) 输出 $P^{(k+1)}(x_i)$ 和 $C^{(k+1)}$。

该迭代算法的收敛性由下述定理给出。

定理 4.3 对上述迭代算法有

$$\lim_{k\to\infty}|C-C^{(k)}|=0 \tag{4.52}$$

等价表示：

$$\lim_{k\to\infty}C^{(k)}=C \tag{4.53}$$

证明：不妨令

$$s_i^{(k+1)}=\sum_i \exp\left\{\sum_j P(y_j\mid x_i)\ln P^{(k)}(x_i\mid y_j)\right\}\quad k=0,1,\cdots \tag{4.54}$$

则

$$P^{(k+1)}(x_i)=\frac{s_i^{(k+1)}}{\sum_l s_l^{(k+1)}}$$

而

$$\begin{aligned}C^{(k+1)}&=F(P^{(k+1)}(x),P^{(k)}(x\mid y))\\&=\sum_i\sum_j P(y_j\mid x_i)P^{(k+1)}(x_i)\ln\frac{P^{(k)}(x_i\mid y_j)}{P^{(k+1)}(x_i)}\\&=\sum_i\sum_j P(y_j\mid x_i)P^{(k+1)}(x_i)\ln\frac{P^{(k)}(x_i\mid y_j)}{\frac{s_i^{(k+1)}}{\sum_l s_l^{(k+1)}}}\\&=\sum_i P^{(k+1)}(x_i)\left\{\sum_j P(y_j\mid x_i)\left[\ln P^{(k)}(x_i\mid y_j)-\ln s_i^{(k+1)}+\ln\left(\sum_l s_l^{(k+1)}\right)\right]\right\}\end{aligned}$$

据式(4.54)有

$$C^{(k+1)} = \ln\left(\sum_l s_l^{(k+1)}\right)$$

设 $P^*(x)$是达到信道容量的最佳分布，即

$$C = \sum_j \sum_i P^*(x_i)P(y_j \mid x_i)\ln\frac{P^*(x_i \mid y_j)}{P(x_i)}$$

其中：

$$P^*(x_i \mid y_j) = \frac{P(y_j \mid x_i)P^*(x_i)}{\sum_l P(y_j \mid x_l)P^*(x_l)} \quad i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s$$

$$\begin{aligned}
&\sum_i P^*(x_i)\ln\frac{P^{(k+1)}(x_i)}{P^{(k)}(x_i)} \\
&= \sum_i P^*(x_i)\ln\left[\frac{s_i^{(k+1)}}{\sum_l s_l^{(k+1)}} \times \frac{1}{P^{(k)}(x_i)}\right] \\
&= -C^{(k+1)} + \sum_i P^*(x_i)\ln\frac{1}{P^{(k)}(x_i)} + \sum_i P^*(x_i)\ln s_i^{(k+1)} \\
&= -C^{(k+1)} + \sum_i \left[\sum_j P(y_j \mid x_i)\right]P^*(x_i)\ln\frac{1}{P^{(k)}(x_i)} \\
&\quad + \sum_i P^*(x_i)\left[\sum_j P(y_j \mid x_i)\ln P^{(k)}(x_i \mid y_j)\right] \\
&= -C^{(k+1)} + \sum_i \sum_j P(y_j \mid x_i)P^*(x_i)\ln\frac{P^{(k)}(x_i \mid y_j)}{P^{(k)}(x_i)} \\
&= -C^{(k+1)} + \sum_i \sum_j P(y_j \mid x_i)P^*(x_i)\ln\left[\frac{1}{P^{(k)}(x_i)} \times \frac{P(y_j \mid x_i)P^{(k)}(x_i)}{\sum_l P(y_j \mid x_l)P^{(k)}(x_l)}\right] \\
&= -C^{(k+1)} + \sum_j \left\{\sum_i P(y_j \mid x_i)P^*(x_i)\ln\left[\frac{1}{P^{(k)}(x_i)} \times \frac{P(y_j \mid x_i)P^{(k)}(x_i)}{\sum_l P(y_j \mid x_l)P^{(k)}(x_l)}\right]\right\} \\
&= -C^{(k+1)} + C + \sum_j \left\{\sum_i P(y_j \mid x_i)P^*(x_i)\right. \\
&\quad \left.\cdot \ln\left[\frac{P^*(x_i)}{P^{(k)}(x_i)} \times \frac{1}{P^*(x_i \mid y_j)} \times \frac{P(y_j \mid x_i)P^{(k)}(x_i)}{\sum_l P(y_j \mid x_l)P^{(k)}(x_l)}\right]\right\} \\
&= -C^{(k+1)} + C + \sum_j \left\{\sum_i P(y_j \mid x_i)P^*(x_i)\ln\left[\frac{\sum_l P(y_j \mid x_l)P^*(x_l)}{\sum_l P(y_j \mid x_l)P^{(k)}(x_l)}\right]\right\}
\end{aligned} \tag{4.55}$$

令

$$h_j^{(k)} = \sum_l P(y_j \mid x_l)P^{(k)}(x_l) \quad j = 1, 2, \cdots, s \tag{4.56}$$

$$h_j^* = \sum_l P(y_j \mid x_l)P^*(x_l) \quad j = 1, 2, \cdots, s \tag{4.57}$$

则式(4.55)可写成

$$\sum_i P^*(x_i)\ln\frac{P^{(k+1)}(x_i)}{P^{(k)}(x_i)} = -C^{(k+1)} + C + \sum_j h_j^* \ln\frac{h_j^*}{h_j^{(k)}} \tag{4.58}$$

由式(4.56)和式(4.57)可知：h^* 和 $h^{(k)}$ 分别是对应最佳输入分布和 k 步迭代时输入分布值的输出符号分布函数，故有

$$\sum_j h_j^* \ln\frac{h_j^*}{h_j^{(k)}} \geqslant 0$$

从而

$$C - C^{(k+1)} \leqslant \sum_i P^*(x_i)\ln\frac{P^{(k+1)}(x_i)}{P^{(k)}(x_i)} \quad k = 0, 1, 2, \cdots$$

又因为

$$C \geqslant C^{(k+1)} \quad k = 0, 1, 2, \cdots$$

所以

$$\begin{aligned}\sum_{k=0}^{N-1} |C - C^{(k+1)}| &\leqslant \sum_{k=0}^{N-1}\sum_i P^*(x_i)\ln\frac{P^{(k+1)}(x_i)}{P^{(k)}(x_i)} \\ &= \sum_i P^*(x_i)\ln\frac{P^{(N)}(x_i)}{P^{(0)}(x_i)} \\ &\leqslant \sum_i P^*(x_i)\ln\frac{P^*(x_i)}{P^{(0)}(x_i)} < \infty\end{aligned}$$

不等式右侧与 N 无关，所以级数 $|C-C^{(k)}|_{k=1}^{\infty}$ 是收敛级数，即

$$\lim_{k\to\infty} |C - C^{(k)}| = 0$$

证毕。

该定理表明，迭代算法一定能得到任意接近信道容量的解，接近的程度取决于第 5 步中的值 δ；算法的收敛速度与初始分布的选择有关，初始分布越接近最佳分布，收敛速度越快。当初始分布选为等概率分布时，有

$$\begin{aligned}\sum_{k=0}^{N-1} |C - C^{(k+1)}| &\leqslant \sum_i P^*(x_i)\ln\frac{P^*(x_i)}{P^{(0)}(x_i)} \\ &= \ln r - H_e(P^*(x)) \\ \min_k |C - C^{(k)}| &\leqslant \frac{\ln r - H_e(P^*(x))}{N}\end{aligned}$$

上式表明，在 k 足够大后，$C^{(k)}$ 以 $1/N$ 的速度逼近信道容量 C。

4.5 连续信道的信道容量

和离散信道一样，对固定的连续信道有一个最大的信息传输速率，称之为信道容量，它也是信道可靠传输的最大信息传输速率。一般连续信道的信道容量为

$$C = \max_{p(x)} I(\boldsymbol{X};\boldsymbol{Y}) = \max_{p(x)}[h(\boldsymbol{Y}) - h(\boldsymbol{Y}\mid\boldsymbol{X})] \quad \text{比特/}N\text{个自由度} \tag{4.59}$$

其中，$p(x)$ 为输入矢量 $\boldsymbol{X}$ 的概率密度函数。

一般波形信道的信道容量为

$$C_t = \max_{p(x)}\left[\lim_{T\to\infty}\frac{1}{T}I(\boldsymbol{X};\boldsymbol{Y})\right]$$

$$= \max_{p(x)}\left\{\lim_{T\to\infty}\left[\frac{1}{T}h(\boldsymbol{Y})-h(\boldsymbol{Y}\mid\boldsymbol{X})\right]\right\}\quad \text{比特/秒} \tag{4.60}$$

其中，$p(x)$为输入矢量 $\boldsymbol{X}$ 的概率密度函数。

若研究的是加性信道，则已知在加性信道中的信道传输概率密度函数就是噪声的概率密度函数。条件熵 $h(\boldsymbol{Y}|\boldsymbol{X})$(即噪声熵)就是噪声源的熵 $h(\boldsymbol{n})$。因此，一般多维加性连续信道的信道容量为

$$C = \max_{p(x)}[H(\boldsymbol{Y})-h(\boldsymbol{n})]\quad \text{比特}/N\text{ 个自由度} \tag{4.61}$$

又由式(4.60)得一般加性波形信道的信道容量为

$$C = \max_{p(x)}\left\{\lim_{T\to\infty}\frac{1}{T}[h(\boldsymbol{Y})-h(\boldsymbol{n})]\right\}$$

$$= \lim_{T\to\infty}\frac{1}{T}\{\max_{p(x)}[h(\boldsymbol{Y})-h(\boldsymbol{n})]\}\quad \text{比特/秒} \tag{4.62}$$

式(4.61)和式(4.62)中 $h(\boldsymbol{n})$与输入矢量 $\boldsymbol{X}$ 的概率密度函数 $p(x)$无关(因输入矢量 $\boldsymbol{X}$ 与噪声矢量 $\boldsymbol{n}$ 统计独立)。所以，求加性信道的信道容量就是求某种发送信号的概率密度函数使接收信号的熵 $h(\boldsymbol{Y})$最大。

由于在不同限制条件下，连续随机变量有不同的最大连续差熵值，因此，由式(4.61)和式(4.62)可知，加性信道的信道容量 C 取决于噪声的统计特性和输入随机矢量 $\boldsymbol{X}$ 所受的限制条件。一般实际信道中，无论输入信号还是噪声，它们的平均功率或能量总是有限的。所以本节只讨论在平均功率受限的条件下，各种连续信道和波形信道的信道容量。

4.5.1　单符号高斯加性信道

单符号高斯加性信道是指信道的输入和输出都是取值连续的一维随机变量，而加入信道的噪声是加性高斯噪声。

设信道叠加的噪声 n 是均值为 0，方差为 σ^2 的一维高斯噪声，求得噪声信源的熵为

$$h(n) = \log\sqrt{2\pi e\sigma^2} \tag{4.63}$$

高斯加性信道的信道容量为

$$C = \max_{p(x)}[h(Y)-\log\sqrt{2\pi e\sigma^2}] \tag{4.64}$$

式(4.64)中，只有 $h(Y)$与输入信号的概率密度函数 $p(x)$有关。当信道输出信号 Y 的平均功率限制在 P_0 以下时，由前已知，Y 是均值为零的高斯变量，其熵 $h(Y)$为最大，而输出信号 Y 是输入信号 X 和噪声的线性叠加。又已知噪声是均值为 0、方差为 σ^2 的高斯变量，并与输入信号 X 彼此统计独立。那么，要使 Y 是均值为 0、方差为 P_0 的高斯变量必须要求输入信号也是均值为 0、方差为 $P_s=P_0-\sigma^2$ 的高斯变量(因为由概率理论知，统计独立的正态分布的随机变量之和仍是正态分布的变量，并且和变量的方差等于各变量的方差之和)。因此可得，平均功率受限于高斯信道的信道容量(每个自由度)：

$$C = \log\sqrt{2\pi eP_0}-\log\sqrt{2\pi e\sigma^2} = \frac{1}{2}\log\frac{P_0}{\sigma^2}$$

$$= \frac{1}{2}\log\left(1+\frac{P_s}{\sigma^2}\right) = \frac{1}{2}\log\left(1+\frac{P_s}{P_n}\right) \tag{4.65}$$

其中，P_s 是输入信号 X 的平均功率；$P_n=\sigma^2$ 是高斯噪声的平均功率。只有当信道的输入信号是均值为 0、平均功率为 P_s 高斯分布的随机变量时，信息传输率才能达到这个最大值。

式(4.65)中，P_s/P_n 为信道的信噪功率比。可见，单符号高斯加性信道的信道容量 C 只取决于信道的信噪功率比。

4.5.2 高斯白噪声加性信道的信道容量

高斯白噪声加性波形信道是经常假设的一种波形信道。此信道的输入和输出信号是随机过程$\{x(t)\}$和$\{y(t)\}$，而加入信道的噪声是加性高斯白噪声$\{n(t)\}$(其均值为 0，功率谱密度为 $N_0/2$)，所以输出信号满足：

$$\{y(t)\}=\{x(t)\}+\{n(t)\}$$

此信道又称为可加波形信道。

一段信道的频带宽度总是有限的，设其频带为 W(即$|f|\leqslant W$)，这样信道的输入、输出信号和噪声都是频带受限的随机过程。根据取样定理，可把一个时间连续的信道变换成时间离散的随机序列信道来处理。由于是加性信道，因此随机序列信道也满足：

$$\boldsymbol{Y}=\boldsymbol{X}+\boldsymbol{n}$$

因为信道的频带是受限的，所以加入信道的噪声成为限带的高斯白噪声。而由前已知，低频限带高斯白噪声的各样本值彼此统计独立，所以限频的高斯白噪声过程可分解成 N 维统计独立的随机序列(其中，每个分量 n_i 的均值为 0，方差 $\sigma_n^2=P_n=\dfrac{N_0WT}{2WT}=\dfrac{N_0}{2}$)，则 N 维的联合概率密度为

$$p(\boldsymbol{n})=p(n_1n_2\cdots n_N)=\prod_{i=1}^{N}n_i=\prod_{i=1}^{N}\frac{1}{2\pi\sigma^2}\mathrm{e}^{-n_i^2/2\sigma^2}$$

对于加性信道来说，若上式成立，则有

$$p(\boldsymbol{y}\mid\boldsymbol{x})=p(\boldsymbol{n})=\prod_{i=1}^{N}p(n_i)=\prod_{i=1}^{N}p(y_i\mid x_i)$$

所以信道是无记忆的。那么，随机序列信道就可等效成 N 个独立的并联信道。其中每个独立信道的信道容量由式(4.65)给出，结合无记忆多维连续信道互信息关系，有

$$\begin{aligned}I(\boldsymbol{X};\boldsymbol{Y})&\leqslant\sum_{i=1}^{N}I(X_i;Y_i)\\&\leqslant\frac{1}{2}\sum_{i=1}^{N}\log\left(1+\frac{P_{s_i}}{P_{n_i}}\right)\end{aligned}\tag{4.66}$$

而信道容量：

$$\begin{aligned}C&=\max_{p(x)}I(\boldsymbol{X};\boldsymbol{Y})\\&=\frac{1}{2}\sum_{i=1}^{N}\log\left(1+\frac{P_{s_i}}{P_{n_i}}\right)\quad\text{比特}/N\text{个自由度}\end{aligned}\tag{4.67}$$

现在，高斯白噪声的每个样本值的方差 $\sigma_n^2=P_n=N_0/2$。信号的平均功率受限为 P_s，T 时间内总平均功率为 P_sT，每个信号样本值的平均功率为 $\dfrac{P_sT}{2WT}=\dfrac{P_s}{2W}$。所以，在[0，$T$]时

刻内，信道的信道容量为

$$C=\frac{N}{2}\log\left(1+\frac{\frac{P_s}{2W}}{\frac{N_0}{2}}\right)=\frac{N}{2}\log\left(1+\frac{P_s}{N_0W}\right)=WT\ \log\left(1+\frac{P_s}{N_0W}\right) \tag{4.68}$$

高斯白噪声加性信道单位时间内的信道容量为

$$C_t=\lim_{T\to\infty}\frac{C}{T}=W\ \log\left(1+\frac{P_s}{N_0W}\right)\quad \text{比特/秒} \tag{4.69}$$

其中，P_s 是信号的平均功率；N_0W 为高斯白噪声在带宽 W 内的平均功率(功率谱密度为 N_0)。可见，信道容量与信噪功率比和带宽有关。这就是著名的 Shannon 公式。

习　题　4

4.1　设二进制对称信道的信道矩阵为$[P(y|x)]=\begin{bmatrix}\frac{2}{3} & \frac{1}{3}\\ \frac{1}{3} & \frac{2}{3}\end{bmatrix}$，求出其信道容量和达到信道容量时的输入概率分布。

4.2　某信道矩阵为$[P(y|x)]=\begin{bmatrix}\frac{2}{3} & 0 & \frac{1}{6} & \frac{1}{6}\\ \frac{1}{6} & \frac{1}{6} & \frac{2}{3} & 0\end{bmatrix}$，此信道有何特点？试计算此信道的信道容量。

4.3　求出下列两信道的信道容量，已知 $p+\bar{p}=1$。

$$\begin{bmatrix}\bar{p}-\varepsilon & p-\varepsilon & 2\varepsilon\\ p-\varepsilon & \bar{p}-\varepsilon & 2\varepsilon\end{bmatrix}\qquad\begin{bmatrix}\bar{p}-\varepsilon & p-\varepsilon & 2\varepsilon & 0\\ p-\varepsilon & \bar{p}-\varepsilon & 0 & 2\varepsilon\end{bmatrix}$$

4.4　图 4-10 为一个二进制对称信道，设该信道以 1500 个二进制符号/秒的速率传输输入信号。现有一相互独立的消息序列，共有 1400 个二进制符号，设在该消息中 $P(0)=P(1)=1/2$。问从信息传输的角度考虑，10 秒钟内能否将该消息序列无失真地传完。

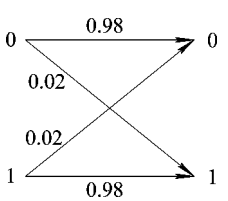

图 4-10　题 4.4 图

4.5　设有输入为 X，输出为 $\boldsymbol{Y}=[Y_1Y_2]$的高斯信道，其中 $Y_1=X+Z_1$，$Y_2=X+Z_2$，X 的最大功率为 P，$(Z_1Z_2)\sim N_2(0,K)$，其中 $K=\begin{pmatrix}\sigma^2 & \rho\sigma^2\\ \rho\sigma^2 & \sigma^2\end{pmatrix}$，试求：

(1) $I(X;Y_1Y_2)=I(X;Y_1)+I(X;Y_2)-I(Y_1;Y_2)+I(Y_1;Y_2|X)$；

(2) $P=1$ 时的信道容量。

第 5 章　有噪信道编码

在无噪无损信道上，只要对信源的输出进行恰当的编码，总能以最大的信息传输率 C（信道容量）无差错地传输信息。但一般信道中总存在噪声或干扰，信息通过传输后会造成损失，那么在有噪信道中怎样才能使信息通过传输后发生的错误最少？在有噪信道中可达的最高信息传输速率是多少？这就是本章研究的内容，即研究通信的可靠性问题，其核心就是 Shannon 第二编码定理，也称为有噪信道编码定理。

5.1　最简单的编码方法

本节从最简单的编码方法入手，了解噪声通信中传输的基本原理和方法。由于噪声的存在，当发送方发送信息比特 0 时，接收方收到的不一定是 0。一个最简单的想法是，当发送比特 m_i 时，重复发送 n 遍，接收方根据接收序列中 0 和 1 的个数多少来判定发送方发送的比特，这种编码方式称做重复码。

【例 5.1】　重复码是信息长度 $k=1$ 的$(n, 1)$码，其中，n 是重复发送的长度，称为码长。经过编码后，码字集合中只有两个长度为 n 的码字$(000\cdots0)$和$(111\cdots1)$。设它们通过二进制对称信道传输，信道的转移概率为 $P=10^{-1}$，码长 n 可能取不同的值，进一步简化情况，只考虑 $n=3$、5、7、…，下面分别讨论。

(1) $n=3$，(3，1)重复码，两个码字是(000)和(111)，可能接收到的向量为{000，001，010，011，100，101，110，111}。当发送 000 时：

没有出现错误，接收到 000，判定为 0；

出现 1 比特错误，即可能接收到 001、010、100 时，仍可以判定为 0，可以纠正 1 比特错误；

出现 2 比特错误，即接收到 110、101、011 时，判定为 1，此时就判错了；

出现 3 比特错误，即接收到 111 时，只能判定为 1，此时也判错了；

也就是说，它只能纠正 1 比特的错误，如果出现 2 比特的错误，如 000 变成了 110，那么就译错了。此时译码错误概率为

$$P_e = 1 - P(0) - P(1) = 1 - (1-p)^3 - 3p(1-p)^2 = 2.8 \times 10^{-2}$$

(2) $n=5$，(5，1)重复码，两个码字是(00000)和(11111)，可以纠正 2 个错，译码错误概率为

$$P = C_5^3 p^3 (1-p)^2 + C_5^4 p^4 (1-p) + p^5 = 8.1 \times 10^{-3}$$

(3) $n=7$，(7，1)重复码，两个码字是(0000000)和(1111111)，可以纠正 3 个错，译码错误概率为

$$P = C_7^4 p^4 (1-p)^3 + C_7^5 p^5 (1-p)^2 + C_7^6 p^6 (1-p) + p^7 = 2.6 \times 10^{-3}$$

对于信源$[X, P]$，每个符号的平均信息量为 $H(X)$，则每个码符号的平均信息量为

$$H' = \frac{H(X)}{n} \tag{5.1}$$

其单位为信息单位/码符号。例如，当信息单位为比特时，每个码符号的平均信息量的单位为比特/码符号。由例 5.1 可知，每个信源符号所需要的码元越多，传输效率越低。我们把编码后的信息传输率定义为

$$R = \frac{\log M}{n} \quad \text{比特 / 码符号} \tag{5.2}$$

其中，M 为使用的码字的个数，信源符号的个数不会超过 M。由于在信源输出和信道之间增加了信道编码器，导致信源符号和信道输入符号集不一致，因此码元符号集和信道输入符号集必须保持一致，以保证码元符号在信道中能顺利传输。

在例 5.1 中，取 $n=3$ 的重复码，相当于在集合$\{000,001,010,011,100,101,110,111\}$中取出了两个串 000 和 111 作为与信源符号 0 和 1 对应的码字，故 $M=2$。

一般地，信道输入符号集有 r 个符号，编码长度为 n 时，可以组成的可用码字个数为 r^n 个。从中选取 $M(M<r^n)$个作为与信源符号对应的码字，则选择种类为 $C_{r^n}^{M}$。

同样的个数 M，选择不同的码字集合，对应的最小错误概率可能就不一样。若信源符号集的符号个数为 m，则应有 $m \leqslant M$。例 5.1 中，$m=2$，信源符号集为$\{0, 1\}$。若取该信源符号集的离散无记忆二次扩展信源：

$$X^2 = \{00, 01, 10, 11\}$$

$m=4$，仍在$\{000, 001, 010, 011, 100, 101, 110, 111\}$中选取 $M=m=4$ 个码字，则可以有 $C_8^4=70$ 种选择方法。例如$\{000,001,010,100\}$、$\{000,011,100,110\}$和$\{000,010,100,110\}$，每一种选择均可计算出相应的最小错误概率。在信道输入符号集分布未知的情况下，我们假设等概率分布是合理的。对应码字集合$\{000, 001, 010, 100\}$的最小错误概率约为 2.28×10^{-2}，对应码字集合$\{000, 011, 100, 110\}$的最小错误概率约为 2×10^{-2}。

若传输每个码元的平均时间为 t 秒，则编码后每秒钟传输的信息量为

$$R = \frac{\log M}{nt} \quad \text{比特 / 秒} \tag{5.3}$$

在简单重复编码方法中，有以下结论(设 $t=1$)：

$n=1$：

$$R = 1 \quad \text{比特 / 码符号}$$

$$R_t = 1 \quad \text{比特 / 秒}$$

$n=3$：

$$R = \frac{1}{3} \quad \text{比特 / 码符号}$$

$$R_t = \frac{1}{3} \quad \text{比特 / 秒}$$

$n=5$：

$$R = \frac{1}{5} \quad \text{比特 / 码符号}$$

$$R_t = \frac{1}{5} \quad \text{比特 / 秒}$$

$n=7$：

$$R=\frac{1}{7}\quad \text{比特/码符号}$$

$$R_t=\frac{1}{7}\quad \text{比特/秒}$$

可见，对于重复扩展编码方法而言，P_e 下降时，编码信息传输率 R 也随之下降。在实际的通信系统中，我们并不希望 R 下降太多，而是希望在保证 R 为一定值的前提下，使错误概率下降。显然，重复编码并不是最好的编码方法。

【例 5.2】 信源符号集为{0，1}，离散无记忆二次扩展信源为{00，01，10，11}，信道为前述的二进制离散无记忆对称信道，$P(0|1)=P(1|0)=p<1/2$。取码长为 5，且与输入符号集有以下对应关系。

输入信源符号：

$$ab\quad a,b\in\{0,1\}$$

对应码字：

$$abcde$$
$$c=a\oplus b$$
$$d=a$$
$$e=a\oplus b$$

故使用码字{00000，01101，10111，11010}分别与{00，01，10，11}对应。

信道输出端可以有 32 种可能输出结果，译码规则为

00000，00001，00010，00100，01000，10000，10001，00011　　译码成　00000

01101，01100，01111，01001，00101，11101，11100，01110　　译码成　01101

10111，10110，10101，10011，11111，00111，00110，10100　　译码成　10111

11010，11011，11000，11110，10010，01010，01011，11001　　译码成　11010

此译码规则具有很强的规律性和对称性（实际上，这个表称之为陪集译码表）：1 个全正确的码字，5 个错一个码元的码字，1 个第一和第五码元错误的码字，1 个第四和第五码元错误的码字。信道看成是离散无记忆的 5 次扩展信道，故正确译码概率为

$$P_D=(1-p)^5+5p(1-p)^4+2p^2(1-p)^3$$

错误概率为

$$P_e=1-P_D=1-\{(1-p)^5+5p(1-p)^4+2p^2(1-p)^3\}$$

当 $p=0.01$ 时，$P_e\approx 7.8\times 10^{-4}$，$R=\frac{2}{5}$比特/码符号。

此码的编码信息传输率与 $M=4$，$n=3$ 的编码相差不大，错误概率却低得多；与 $M=2$，$n=3$ 的编码相比，错误概率相当，但编码信息传输率要高。

从前面的论述中可以看出，在有噪信道中消息传输的错误概率与所采用的编译码方法有关，那么在有噪信道中，有没有存在一种最佳的编码方法使得错误概率尽可能小呢？在使平均错误概率尽可能小的情况下，可达的最大的信息传输速率是多少呢？Shannon 在 1948 年的论文中首先给出了肯定的回答，并指出这个可达的最大的信息传输速率是有噪信道的信道容量，这就是著名的 Shannon 第二编码定理，也称为有噪信道编码定理。

5.2　联合 ε 典型序列

Shannon 在有噪信道编码定理的证明中，采用的是基于联合 ε 典型序列的思想。到目前为止，它是一种较为简洁的证明。

我们分析的信道是 n 次无记忆扩展信道，如图 5－1 所示。

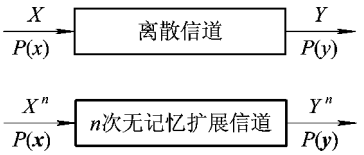

图 5－1　离散信道的 n 次无记忆扩展

图中：

$$\boldsymbol{x}=(x_1,\ x_2,\ \cdots,\ x_N)\quad x_i\in X$$

$$\boldsymbol{y}=(y_1,\ y_2,\ \cdots,\ y_N)\quad y_i\in Y$$

$P(x_i)$、$P(y_i)$是单符号离散信道输入和输出的概率分布，其信道转移概率为 $P(y_i|x_i)$，而 x 和 y 都是 n 次无记忆扩展信道的输入和输出的长为 n 的随机序列，已知其满足

$$P(\boldsymbol{y}\mid\boldsymbol{x})=\prod_{i=1}^{n}P(y_i\mid x_i)$$

它们对应的信息熵和联合熵为 $H(X)$、$H(Y)$、$H(XY)$。

定义 5.1　令 $H(X)$是信源$[X,\ P(\boldsymbol{x})]$的熵，ε 是正数，集合

$$T_X(N,\ \varepsilon)=\{\boldsymbol{x}:\ |\ H(X)-I_n\ |\leqslant\varepsilon,\ \boldsymbol{x}\in X^n\}\tag{5.4}$$

称为信源$[X,\ P(\boldsymbol{x})]$输出的长为 n 的典型序列集。

其中：

$$I_n=\frac{I(\boldsymbol{x})}{n}\qquad \boldsymbol{x}\in X^n$$

X^n 为 X 的 n 次离散无记忆扩展信源。

上述定义又称 ε 弱典型序列集。其实质是由长为 n、平均信息熵和 $H(X)$相差不超过 ε 的所有矢量组成的集合。

定义 5.2　令 $\boldsymbol{x}$ 是信源$[X,\ P(\boldsymbol{x})]$发出的长为 n 的序列，ε 是正数，集合：

$$T_X(n,\ \varepsilon)=\left\{\boldsymbol{x}:\left|\frac{n_k}{n}-P(a_k)\right|\leqslant\varepsilon,\ \boldsymbol{x}\in X^n\right\}\tag{5.5}$$

称为信源$[X,\ P(\boldsymbol{x})]$输出的长为 n 的ε 典型序列集。式(5.5)中，n_k 是在长为 n 的输出序列中符号 a_k 出现的次数。

上述定义又称做强 ε 典型序列集。其实质是由长为 n、每个符号的出现频率和概率相差不超过 ε 的所有矢量组成的集合。

定义 5.3　$T_X(n,\ \varepsilon)$的补集 $\overline{T}_X(n,\ \varepsilon)$称为非典型序列集。

定理 5.1(信源划分定理)　给定信源$[X,\ P(\boldsymbol{x})]$，ε 是正数，当 $n\to\infty$时，$P\{T_X(n,\ \varepsilon)\}\to1$，即对 $\forall\varepsilon>0$，$\exists n_0$ 使得当 $n>n_0$ 时，有

$$P\{\boldsymbol{x}\in T_X(n,\ \varepsilon)\}\geqslant1-\varepsilon\tag{5.6}$$

推论 1(特定典型序列出现的概率) 若 $\boldsymbol{x}\in T_X(n,\varepsilon)$，则

$$2^{-n[H(x)+\varepsilon]}\leqslant P(\boldsymbol{x})\leqslant 2^{-n[H(x)-\varepsilon]} \tag{5.7}$$

证明：由定义式(5.4)有：

$$-[H(X)+\varepsilon]\leqslant\frac{1}{n}\log P(\boldsymbol{x})\leqslant-[H(X)-\varepsilon] \tag{5.8}$$

不等式各项取指数即可得证。

该推论说明，所有典型序列出现的概率都在同一范围内，当 ε 足够小时，概率几乎为一常量，且当 n 增大时，该概率随 $nH(X)$ 指数减少。

推论 2(典型序列的数目) 当 n 足够大时，对于给定的信源和 $\varepsilon>0$，典型序列的数目满足：

$$(1-\varepsilon)2^{n[H(X)-\varepsilon]}\leqslant\|T_X(n,\varepsilon)\|\leqslant 2^{n[H(X)+\varepsilon]} \tag{5.9}$$

证明：

$$1=\sum_{\boldsymbol{x}\in X^n}P(\boldsymbol{x})\geqslant\sum_{\boldsymbol{x}\in T_X(n,\varepsilon)}P(\boldsymbol{x})\geqslant\sum_{\boldsymbol{x}\in T_X(n,\varepsilon)}2^{-n(H(X)+\varepsilon)}=\|T_X(n,\varepsilon)\|2^{-n(H(X)+\varepsilon)}$$

故

$$\|T_X(n,\varepsilon)\|\leqslant 2^{n[H(X)+\varepsilon]}$$

据式(5.5)和式(5.6)有：

$$1-\varepsilon\leqslant\sum_{T_X(n,\varepsilon)}P(\boldsymbol{x})\leqslant\sum_{T_X(n,\varepsilon)}2^{-n(H(X)-\varepsilon)}=\|T_X(n,\varepsilon)\|2^{-n(H(X)-\varepsilon)}$$

故

$$\|T_X(n,\varepsilon)\|\geqslant(1-\varepsilon)2^{n[H(X)-\varepsilon]}$$

证毕。

由式(5.9)可知：

$$\|T_X(n,\varepsilon)\|\approx 2^{nH(X)} \tag{5.10}$$

信源划分定理及其推论说明，一个离散无记忆信源的输出符号序列可分成 $T_X(n,\varepsilon)$ 和 $\overline{T}_X(n,\varepsilon)$ 两组，有 $T_X(n,\varepsilon)\bigcup\overline{T}_X(n,\varepsilon)=X^n$。$T_X(n,\varepsilon)$ 中各序列出现的概率几乎相等，且每个序列中一个信源符号的平均信息量接近于信源的熵 $H(X)$，所有典型序列的概率和趋向于 1。因此，典型序列集又可称为高概率集，非典型序列集称为低概率集，信源的这种划分性质渐近等同于分割性。典型序列又称做渐近等概序列。

由于典型序列集在整个信源输出序列中占绝对优势，所以在信息论的研究中可以忽略非典型序列集而只考察典型序列集。

尽管 $T_X(n,\varepsilon)$ 中元素个数不少，约为 $2^{nH(X)}$ 个，但与 $\|X^n\|$ 相比，仍是非常少的一部分。

事实上，若 $\|X\|=r$，则 X^n 中的元素个数为 $r^n=2^{n\log r}$ 个，有：

$$\alpha=\frac{\|T_X(n,\varepsilon)\|}{\|X^n\|}\leqslant 2^{-n[\log r-H(X)-\varepsilon]}$$

若 $-n[\log r-H(X)-\varepsilon]>0$，即信源非等概分布，则有 $n\to\infty$ 时 $\alpha\to 0$。

若信源等概，$H(X)=\log r$，则由式(5.10)得：$\|T_X(n,\varepsilon)\|\approx 2^{nH(X)}$，即信源等概分布时几乎所有序列都是典型序列。

定义 5.4　令 X、Y 是两个概率空间，$\boldsymbol{x}=(x_1,x_2,\cdots,x_N)\in X^n$，$\boldsymbol{y}=(y_1,y_2,\cdots,y_n)\in Y^n$。若序列对 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 满足：

(1) $\boldsymbol{x}$ 是 ε 典型序列，即对任意小的正数 ε，存在 n，使得：

$$\left|-\frac{1}{n}\log P(\boldsymbol{x})-H(X)\right|\leqslant\varepsilon \tag{5.11}$$

(2) $\boldsymbol{y}$ 是 ε 典型序列，即对任意小的正数 ε，存在 n，使得：

$$\left|-\frac{1}{n}\log P(\boldsymbol{y})-H(Y)\right|\leqslant\varepsilon \tag{5.12}$$

(3) $(\boldsymbol{x},\boldsymbol{y})$ 是 ε 典型序列，即对任意小的正数 ε，存在 n，使得：

$$\left|-\frac{1}{n}\log P(\boldsymbol{xy})-H(XY)\right|\leqslant\varepsilon \tag{5.13}$$

则称序列对 $(\boldsymbol{x},\boldsymbol{y})$ 是联合 ε 典型序列，表示为 $T_{XY}(n,\varepsilon)$。

给定 $\boldsymbol{y}$，称集合：

$$T_{X|Y}(n,\varepsilon)=\{\boldsymbol{x}:\boldsymbol{xy}\in T_{XY}(n,\varepsilon)\} \tag{5.14}$$

为条件 $\boldsymbol{y}$ 下 X 的 ε 典型序列。

定理 5.2　当 n 足够大时，有：

$$\|T_{X|Y}(n,\varepsilon)\|\leqslant 2^{n[H(X|Y)+2\varepsilon]} \tag{5.15}$$

证明：$\forall\,\boldsymbol{y}\notin T_Y(n,\varepsilon)$，$|T_{X|Y}(n,\varepsilon)|=0$，引理显然成立。

$\forall\,\boldsymbol{y}\in T_Y(n,\varepsilon)$，有：

$$1=\sum_{X^n}P(\boldsymbol{x}\mid\boldsymbol{y})=\sum_{X^n}\frac{P(\boldsymbol{x},\boldsymbol{y})}{P(\boldsymbol{y})}\geqslant\sum_{\boldsymbol{x}\in T_{X|Y}(n,\varepsilon)}\frac{P(\boldsymbol{x},\boldsymbol{y})}{P(\boldsymbol{y})}\geqslant\sum_{\boldsymbol{x}\in T_{X|Y}(n,\varepsilon)}\frac{2^{-n[H(XY)+\varepsilon]}}{P(\boldsymbol{y})}$$

$$\geqslant\sum_{\boldsymbol{x}\in T_{X|Y}(n,\varepsilon)}\frac{2^{-n[H(XY)+\varepsilon]}}{2^{-n[H(Y)-\varepsilon]}}=\|T_{X|Y}(n,\varepsilon)\|\,2^{-n[H(X|Y)+2\varepsilon]}$$

$$\|T_{X|Y}(n,\varepsilon)\|\leqslant 2^{N[H(X|Y)+2\varepsilon]}$$

同理可证：

$$\|T_{Y|X}(n,\varepsilon)\|\leqslant 2^{n[H(Y|X)+2\varepsilon]} \tag{5.16}$$

证毕。

定理 5.3　若对于 $\boldsymbol{x}\in X^n$，$\boldsymbol{y}\in Y^n$，$\boldsymbol{x}$、$\boldsymbol{y}$ 统计独立，$(\boldsymbol{x},\boldsymbol{y})$ 是联合 ε 典型序列，则当 n 足够大时，有：

$$(1-\varepsilon)2^{-n[I(X;Y)+3\varepsilon]}\leqslant\sum_{\boldsymbol{x}}\sum_{(\boldsymbol{x},\boldsymbol{y})\in T_{XY}(n,\varepsilon)}P(\boldsymbol{x})P(\boldsymbol{y})\leqslant 2^{-n[I(X;Y)-3\varepsilon]} \tag{5.17}$$

证明：由于 $\boldsymbol{x}$、$\boldsymbol{y}$ 均为典型序列，故有：

$$2^{-n[H(X)+\varepsilon]}\leqslant P(\boldsymbol{x})\leqslant 2^{-n[H(X)-\varepsilon]} \tag{5.18}$$

$$2^{-n[H(Y)+\varepsilon]}\leqslant P(\boldsymbol{y})\leqslant 2^{-n[H(Y)-\varepsilon]} \tag{5.19}$$

$$(1-\varepsilon)2^{n[H(XY)-\varepsilon]}\leqslant\|T_{XY}(n,\varepsilon)\|\leqslant 2^{n[H(XY)+\varepsilon]} \tag{5.20}$$

将上述三个不等式相应项相乘，考虑到：

$$-I(X;Y)=H(XY)-H(X)-H(Y) \tag{5.21}$$

证毕。

如前所述，联合 ε 典型序列要求 $(\boldsymbol{x},\boldsymbol{y})$ 必须是 ε 典型序列，令：

$$L_\varepsilon=\|T_X(n,\varepsilon)\|\cdot\|T_Y(n,\varepsilon)\| \tag{5.22}$$

由信源划分定理推论 2 可得出 L_ε 的上、下限为

$$(1-\varepsilon)^2 2^{n[H(X)+H(Y)-2\varepsilon]} \leqslant L_\varepsilon \leqslant 2^{n[H(X)+H(Y)+2\varepsilon]} \tag{5.23}$$

由式(5.20)和式(5.23)可得：

$$(1-\varepsilon)2^{-n[I(X;Y)+3\varepsilon]} \leqslant \frac{T_{XY}(n,\varepsilon)}{L_\varepsilon} \leqslant (1-\varepsilon)^{-2}2^{n[I(X;Y)+3\varepsilon]} \tag{5.24}$$

由于 ε 为任意小的正数，因此 $(1-\varepsilon)^{-2}\approx 1$，从而 $T_{XY}(n,\varepsilon)/L_\varepsilon$ 和式(5.17)中的量 $\sum_{x}\sum_{(x,y)\in T_{XY}(n,\varepsilon)}{}_{y} P(x)P(y)$ 的上、下限保持一致。$T_{XY}(n,\varepsilon)/L_\varepsilon$ 为联合 ε 典型序列的由 ε 典型序列 x、y 构成的空间中的平均密度。

5.3 有噪信道编码与 Shannon 第二编码定理

在前面的讨论中已经提到，在通信系统中，信道是非常重要的组成部分，且由于信道中随机噪声的存在，信道在传输符号时会产生畸变，从而使接收到的符号序列与发送端的符号序列不一致。为此，需要寻找编码方法，在信道噪声存在的情况下能够检测进而纠正发生错误的符号，以提高通信系统的可靠性。

从概率的角度研究噪声通信的问题就是要研究是否存在一种编码方法使得编码信息传输率保持较大，这就是著名的 Shannon 第二编码定理，也称为有噪信道编码定理。

任意给定希望达到的编码信息速率 R 和错误概率限度 ε，只要 R 小于信道容量，就可以找到一种编码方法，使得码长足够大时，错误概率达到小于 ε 的要求。

定义 5.5 对给定离散无记忆信道和任意 $\varepsilon>0$，若有一种编码信息速率为 R 的码，在码长 n 足够大时，能使 $P_e<\varepsilon$，则称 R 是可达的。

定理 5.4(Shannon 第二编码定理，有噪信道编码定理) 给定信道容量为 C 的离散无记忆信道$[P(y|x)]$，若编码信息传输率 $R<C$，则 R 是可达的。

等价描述：给定信道容量为 C 的离散无记忆信道$[P(y|x)]$，若编码信息传输率$R<C$，则只要码长 n 足够长，总可以在输入集合中找到 M 个码字对应 M 种可能的信源消息符号，组成一个码集合，$M\leqslant 2^{n(C-\varepsilon)}$，$\varepsilon$ 为任意小的正数，在一定的译码规则下，可使信道输出的错误概率任意小。

解释：设信道容量为 C 的离散无记忆信道有 r 个输入符号(从而码符号也有 r 个)、s 个输出符号，则码长为 N 时，可将信道看成是离散无记忆的 n 次扩展信道，扩展信道的输出符号集有 s^n 种选择，输入符号集有 r^n 种选择，从 r^n 种选择中选出 M 种作为码字，对应 M 种可能的消息符号。为了正确译码，应有 $M\leqslant s^n$。

下面我们来证明 Shannon 第二编码定理。

设信道输入序列为 $\boldsymbol{x}$，$\boldsymbol{x}\in X^n$，$\boldsymbol{x}=(x_1, x_2, \cdots, x_n)$，输出为 $\boldsymbol{y}\in Y^n$，$\boldsymbol{y}=(y_1, y_2, \cdots, y_n)$，转移概率为

$$P(\boldsymbol{y}\mid\boldsymbol{x}) = \prod_{i=1}^{n} P(y_i \mid x_i)$$

若发送消息为 $\boldsymbol{x}_m$，接收序列为 $\boldsymbol{y}$，则在发送消息 $\boldsymbol{x}_m$ 的条件下的译码错误概率为

$$P_{em} = P(\boldsymbol{x}_k \neq \boldsymbol{x}_m \mid \boldsymbol{y})$$

消息独立、等概时的平均错误率(针对每个 X^N)为

$$P_e = \frac{1}{M}\sum_m P_{em} \tag{5.25}$$

其中，M 为消息数目。

从 X^n 中独立随机地选择 $M=2^{nR}$ 个序列作为码字，这相当于每个码字出现的概率为

$$P(\boldsymbol{x}) = \prod_{i=1}^{n} P(x_i)$$

其中，R 是要求的编码信息传输率；$\boldsymbol{x}\in X^n$，$\boldsymbol{x}=(x_1, x_2, \cdots, x_n)$，并设 X 中的所有元素独立、等概地出现。这种编码方法称为随机编码。

取译码规则：对给定的接收序列 $\boldsymbol{y}$，若存在唯一的 $k\in[1, 2^{nR}]$，使

$$(\boldsymbol{x}_k, \boldsymbol{y}) \in T_{XY}(n, \varepsilon)$$

则将 $\boldsymbol{y}$ 译为 $\boldsymbol{x}_k$，即

$$F(\boldsymbol{y}) = \boldsymbol{x}_k$$

若实际发送的为 $\boldsymbol{x}_m$，则当 $k\neq m$ 或两个以上(含两个)的 k 与 $\boldsymbol{y}$ 构成联合 ε 典型序列时就认为出现译码错误。由于随机码具有对称性，因此译码错误概率与送出消息的标号无关，即随机码集合的平均错误概率就是任一特定消息被错误译码的概率。不失一般性，设发送的是第一个消息，令事件：

$$E_m = \{(\boldsymbol{x}_m, \boldsymbol{y}) \in T_{XY}(n, \varepsilon)\} \quad m \in [1, 2^{nR}]$$

则发送消息 $\boldsymbol{x}_1$ 时，译码错误概率为

$$P_e = P_{e1} = P\{\boldsymbol{x}_k \neq \boldsymbol{x}_1 \mid \boldsymbol{x}_1\} = P\left[\left(\bigcup_{k\neq 1} E_k\right) \mid \boldsymbol{x}_1\right] = \sum_{k\neq 1} P(E_k \mid \boldsymbol{x}_1)$$

而

$$P(E_k \mid \boldsymbol{x}_1) = \sum_{j=1}^{|Y^n|} P\{(\boldsymbol{x}_m, \boldsymbol{y}) \in T_{XY}(n, \varepsilon)\} P(\boldsymbol{y}_j \mid \boldsymbol{x}_1) \quad k \neq 1$$

又

$$P\{(\boldsymbol{x}_m, \boldsymbol{y}) \in T_{XY}(n, \varepsilon)\} = P\{(\boldsymbol{x}_k, \boldsymbol{y}) \in T_{XY}(n, \varepsilon)\}$$

与 j 无关，故

$$P(E_k \mid \boldsymbol{x}_1) = P\{(\boldsymbol{x}_m, \boldsymbol{y}) \in T_{XY}(n, \varepsilon)\} \quad k \neq 1$$

由定理 5.2 得：

$$P(E_k) \leqslant 2^{-n[I(X;Y)-3\varepsilon]} \tag{5.26}$$

所以

$$\sum_{k\neq 1} P(E_k) \leqslant 2^{nR} \cdot 2^{-n[I(X;Y)-3\varepsilon]} = 2^{n[R-I(X;Y)+3\varepsilon]}$$

为了使 R 大，故使 $I(X;Y)$ 极大化，即取 $I(X;Y)=C$。若 $R<C-3\varepsilon$，则 $n\to\infty$ 时，上式趋向于 0。由于错误概率非负，因此错误概率趋向于 0。

由于当 $R<C-3\varepsilon$ 时，随机码集合的平均译码错误概率趋向于 0，所以必然存在一种码，当 n 足够大时，其译码错误概率任意小。

证毕。

码的个数为 2^{nR}，由于 ε 为任意小的正数，因此 $\eta=3\varepsilon$ 仍为任意小的正数，故码字数目满足 $M=2^{nR}\leqslant 2^{n(C-\eta)}$，$\eta$ 为任意小的正数，从而等价描述成立。

借助于典型序列的概念，我们证明了离散无记忆信道编码定理，指出只要 $R<C$，必然存在编码方法使译码错误概率任意小，却未给出具体的编码方法。那么当 $R>C$ 时情况如何呢？信道编码逆定理给出了答案。

定理 5.5（信道编码逆定理） 设离散无记忆信道的信道容量为 C，当信息传输速率 $R>C$（选用码字个数 $M=2^{n(C+\varepsilon)}$）时，无论 n 多大也不能找到一种编码，使译码错误概率任意小。

证明：假设选用 $M=2^{n(C+\varepsilon)}$ 码字组成的一个码，每个码字的概率是 $1/M$，则离散无记忆信道的 n 次扩展信道的平均互信息为

$$H(X^n)-H(X^n\mid Y^n)\leqslant nC \tag{5.27}$$

则

$$\log 2^{n(C+\varepsilon)}-H(X^n\mid Y^n)\leqslant nC \tag{5.28}$$

即

$$n\varepsilon\leqslant H(X^n\mid Y^n)$$

由 Fano 不等式有：

$$n\varepsilon\leqslant H(X^n\mid Y^n)\leqslant H(P_e)+P_e\log(M-1)\leqslant 1+P_e\log(M-1)$$

所以

$$n\varepsilon\leqslant 1+P_e\log(M-1)<1+P_e\log M=1+P_e(nC+n\varepsilon)$$

求得：

$$P_e>\frac{n\varepsilon-1}{nC+n\varepsilon}$$

当 $n\to\infty$ 时，$\dfrac{n\varepsilon-1}{nC+n\varepsilon}\to\dfrac{\varepsilon}{C+\varepsilon}>0$，从而 P_e 为非零值。

证毕。

以上只是在离散无记忆信道的情况下对定理做了证明，事实上，Shannon 第二编码定理和逆定理对连续信道、有记忆信道同样成立。现已证明离散无记忆信道中译码错误概率 P_e 趋于零的速度与 n（码长）成指数关系，即当 $R\to C$ 时，译码错误概率为

$$P_e\leqslant e^{-nE_r(R)} \tag{5.29}$$

式中，$E_r(R)$ 为随机编码指数，又称做可靠性函数，其表达式为

$$E_r(R)=\max_{0\leqslant\rho\leqslant 1}\max_{P(x)}\{E_0[\rho,P(x)]-\rho R\} \tag{5.30}$$

其中，ρ 为人为定义的修正系数，$0\leqslant\rho\leqslant 1$。可见，可靠性函数又与输入概率分布有关。一般 $E_r(R)$ 与 R 的关系曲线如图 5-2 所示，是一条下凸函数。从图 5-2 中可以看出，$R<C$ 时，$E_r(R)>0$，所以式(5.29)表明 P_e 随着 n 的增大以指数趋于零。

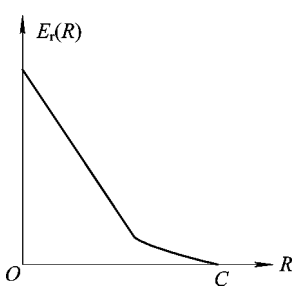

图 5-2 $E_r(R)$ 与 R 的关系曲线

Shannon 第二编码定理只是一个存在性定理，它说明错误概率趋于零的好码是存在的，但是并没有给出怎么构造，尽管如此，它对信道编码的发展仍具有重大的指导意义。

5.4　Shannon 理论对信道编码的指导意义

信息传输的可靠性是所有通信系统努力追求的首要目标。要实现高可靠性的传输，可采取诸如增大发射功率、增加信道带宽、提高天线增益等传统方法，但这些方法往往难度比较大，有些场合甚至无法实现。Shannon 信息论指出：对信息序列进行适当的编码后同样可以提高信道的传输可靠性，这种编码即为信道编码。信道编码是在著名的信道编码定理的指导下发展起来的，几十年来已取得了丰硕的成果。

5.4.1　二进制离散无记忆信道下的极限

二进制对称信道的信道容量为

$$C = 1 + p\log p + (1-p)\log(1-p) \quad \text{比特 / 码元} \tag{5.31}$$

采用相干 PSK 信号发送和接收码字比特，有

$$p = Q\left(\sqrt{\frac{E_s}{N_0}}\right) = Q\left(\sqrt{\frac{2E_bR_c}{N_0}}\right) \tag{5.32}$$

令式(5.31)中的 $C=R_c$，求出满足式(5.32)的 E_b/N_0 值。假如采用码率 $R_c=1/2$ 的编码方式，则为了获得硬判决译码时该码的容量，要求每比特最小信噪比约为 1.6 dB。

当码率渐近于零时，我们关心的是最小信噪比极限。在 R_c 较小时，概率 p 近似为

$$p \approx \frac{1}{2} - \sqrt{\frac{E_bR_c}{N_0\pi}} \tag{5.33}$$

把 p 的表达式代入式(5.31)，并利用式：

$$\log(1+x) \approx \frac{x - \frac{x^2}{2}}{\ln 2}$$

可把信道容量公式简化为

$$C = \frac{2}{\pi \ln 2}\frac{E_b}{N_0}R_c \tag{5.34}$$

再令 $C=R_c$，在 R_c 趋近于零的极限情况下，可得：

$$\frac{E_b}{N_0} = \frac{1}{2}\pi \ln 2 \quad (0.37\ \text{dB}) \tag{5.35}$$

也就是说，在二进制离散信道下，只要传输每比特信息的信噪比不低于 0.37 dB，就可能实现完全无差错的传输。

5.4.2　AWGN 信道下的理论极限

由信息论中对波形信道的分析可知，当信道中的噪声为加性高斯白噪声时，其信道容量为

$$C_t = W\log\left(1 + \frac{P_s}{N_0W}\right) \tag{5.36}$$

其中，P_s 是信号的平均功率，W 为信道带宽。

令 $P_s=R_t \cdot E_b$，R_t 为信息传输速率(比特/秒)，E_b 为传输每一比特信息所需的平均功率，则式(5.36)可化为

$$C_t = W \log\left(1+\frac{R_t}{W}\frac{E_b}{N_0}\right) \tag{5.37}$$

式中，E_b/N_0 可理解为每赫兹带宽传输每比特信息所需的信噪功率比。式(5.37)表明了通信系统中带宽 W、信道容量 C_t、信噪比 E_b/N_0 之间的定量关系。显然，增加带宽和提高信噪比都能使信道容量得到增加，从而使一定的信息传输率 R_t 下的错误概率减小。

在信道带宽不受限的情况下，信道容量仅取决于信噪比，进而此时传输可靠性主要由信噪比决定。由式(5.37)可得：

$$\frac{E_b}{N_0} = \frac{2^{\frac{C_t}{W}}-1}{\frac{R_t}{W}} \tag{5.38}$$

令式(5.38)中 $W\to\infty$，且 $R_t\to C$，则可得到可靠传输时所需的最小信噪比为

$$\left(\frac{E_b}{N_0}\right)_{\min} = \lim_{W\to\infty}\frac{2^{\frac{C_t}{W}}-1}{\frac{C}{W}} = \ln 2 = -1.6\ \text{dB} \tag{5.39}$$

式(5.39)表示在带宽不受限的高斯白噪声信道中，只要传输每比特信息的信噪比不低于 -1.6 dB，就可能实现完全无差错的传输。这是高斯信道中传送信息的极限能力，这一信噪比的极限值称为 Shannon 限。

通过前面的分析可以知道，在二进制对称信道下，达到信道容量所需要的信噪比为 0.37 dB，所以在实际的译码过程中建议采用波形信道模型，可获得大约 2 dB 的增益。

当信道带宽有限，且信道输入为二元输入时，需将式(5.37)中的 C_t 和 R_t 按通带归一化，令

$$C_n = \frac{C_t}{2W},\ R = \frac{R_t}{2W}$$

即得：

$$C_n = \frac{1}{2}\log\left(1+2R\frac{E_b}{N_0}\right) \tag{5.40}$$

或

$$\frac{E_b}{N_0} = \frac{2^{2C_n}-1}{2R} \tag{5.41}$$

令 $R\to C_n$，则可得带限信道中二元信息可靠传输时所需的信噪比为

$$\frac{E_b}{N_0} = \frac{1}{2R}[2^{2R}-1] \tag{5.42}$$

令码率 $R=\frac{1}{2}$，则得：

$$\frac{E_b}{N_0} = 1 \tag{5.43}$$

这是带限高斯信道以 1/2 的码率可靠传输每比特二元信息所需的最小信噪比，是 Shannon 限的另一种表述。

实际上不采取任何编码措施的传输系统是不可能以这么低的信噪比实现可靠传输的。采取信道编码后，在某一误码率指标下，经过仿真计算，所需信噪比可比不编码时有明显减少。

编码的分析设计有两条基本途径。一是不涉及具体的编码，运用概率统计的方法，在特定信道条件下对编码信号的性能作出统计分析，求出差错概率的上下边界，这就是本章前半部分的内容。这种方法不能得知最优码是如何编出来的，却能得出最优码可以好到什么程度，对指导编码技术具有特别重要的理论价值。另一条途径是数学解析途径，将代数、集合、数论等理论运用到编译码中来设计某种码，比如分组码、卷积码、Turbo 码等。本章后面几节将简单介绍这几种编码方式，主要目的是了解 Shannon 定理对编码的指导意义。信道编码更深、更系统的知识请参照相关的专业书籍。

5.5　线性分组码

本节讨论纠错码中最基本的一类码，即线性分组码。这类码有明显的数学结构，它虽然简单，但却是各类码的基础，且许多已知的好码都属于线性分组码。

5.5.1　线性分组码的编码

为了引入线性分组码的概念，首先看下面的例子。

【例 5.3】　任给一个由 $k=3$ 位信息组成的信息组 $\boldsymbol{m}=(m_2, m_1, m_0)$，由它生成的 $(6, 3)$线性分组码的码字 $\boldsymbol{v}=(v_5, v_4, v_3, v_2, v_1, v_0)$由下列关系式确定：

$$v_{i+3}=m_i \quad i=0, 1, 2$$

$$v_2=m_2+m_1,\ v_1=m_2+m_0,\ v_0=m_1+m_0$$

由于每个信息组共有 $k=3$ 位信息码元，信息组集合共由 $2^3=8$ 个不同的信息组构成，因此上述关系生成的$(6, 3)$线性分组码共有 8 个码字。对于任意信息组 $\boldsymbol{m}=(m_2, m_1, m_0)$，生成的码字：

$$\boldsymbol{v}=(m_2, m_1, m_0, m_2+m_1, m_2+m_0, m_1+m_0) \tag{5.44}$$

若用向量与矩阵的乘积表示式(5.44)，则可写成：

$$v=(m_2, m_1, m_0)\boldsymbol{G}$$

其中：

$$\boldsymbol{G}=\begin{bmatrix}1 & 0 & 0 & 1 & 1 & 0\\ 0 & 1 & 0 & 1 & 0 & 1\\ 0 & 0 & 1 & 0 & 1 & 1\end{bmatrix}$$

称为该$(6, 3)$码的生成矩阵。有了生成矩阵 $\boldsymbol{G}$，不难将 $2^3=8$ 个信息组变换成表 5－1 所示的$(6, 3)$码的 8 个码字。

表 5-1 (6，3)码的码字

信息组	码字	信息组	码字
000	000000	100	100110
001	001011	101	101101
010	010101	110	110011
011	011110	111	111000

从生成的码字可以看出，前 $k=3$ 位是原信息组，而后 $n-k=3$ 位是由式(5.44)确定的监督元，因此我们称(6，3)码为系统码。

定义 5.6 信息组以不变的形式，在码字的任意 k 位中出现的码称为系统码，否则，称为非系统码。

一般地，我们把信息组排在码的前 k 位，分析$(n，k)$系统码的生成矩阵可以看出，它的左边必然是一个 $k\times k$ 阶单位矩阵，因此矩阵 $\boldsymbol{G}$ 的秩为 k。若分别用 $\boldsymbol{I}_k$ 和 $\boldsymbol{P}$ 表示 $\boldsymbol{G}$ 的两个矩阵块，则生成矩阵 $\boldsymbol{G}$ 可表示成

$$\boldsymbol{G}=[\boldsymbol{I}_k \quad \boldsymbol{P}] \tag{5.45}$$

通常把上式表示的生成矩阵 $\boldsymbol{G}$ 称为标准型生成矩阵，$\boldsymbol{G}$ 的 k 行是线性无关的。

事实上，$\boldsymbol{G}$ 的每一行都是码集合中的一个码字，展开来写就是：

$$\boldsymbol{G}=\begin{bmatrix} 1 & 0 & \cdots & 0 & v_{1,r-1} & v_{1,r-2} & \cdots & v_{1,0} \\ 0 & 1 & \cdots & 0 & v_{2,r-1} & v_{2,r-2} & \cdots & v_{2,0} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & v_{k,r-1} & v_{k,r-2} & \cdots & v_{k,0} \end{bmatrix}$$

因此可归纳出一般$(n，k)$系统线性分组码的结构形式如下：

k 位信息位	$n-k$ 位校验位

一般地，如果码元 v_{n-1}，v_{n-2}，…，v_{n-k} 也是 m_{k-1}，m_{k-2}，…，m_0 的线性组合，则可得到非系统线性分组码，即

$$\boldsymbol{C}=\boldsymbol{m}\cdot\boldsymbol{G}=(m_k \quad m_{k-1} \quad \cdots \quad m_0)\begin{bmatrix} g_{1,n-1} & g_{1,n-2} & \cdots & g_{1,0} \\ g_{2,n-1} & g_{2,n-2} & \cdots & g_{2,0} \\ \vdots & \vdots & & \vdots \\ g_{k,n-1} & g_{k,n-2} & \cdots & g_{k,0} \end{bmatrix}$$

其中：

$$\boldsymbol{G}=\begin{bmatrix} g_{1,n-1} & g_{1,n-2} & \cdots & g_{1,0} \\ g_{2,n-1} & g_{2,n-2} & \cdots & g_{2,0} \\ \vdots & \vdots & & \vdots \\ g_{k,n-1} & g_{k,n-2} & \cdots & g_{k,0} \end{bmatrix}$$

是一个秩为 k 的 $k\times n$ 阶矩阵。

当 $\boldsymbol{G}$ 是式(5.45)表示的标准型生成矩阵时，它才生成系统码，这时它生成的码字前 k

位是原信息组，后 $r=n-k$ 位是监督元。

对于一般情况，如果不给出信息组和码字的对应码表，而只给出一个码集合，如

000000，001011，010101，011110，100110，101101，110011，111000

此时我们无法知道该码字对应哪个信息组。

实际上，k 个线性无关的码向量的一切线性组合共有 2^k 个可能，它恰好是 V 的 2^k 个码字，因此任意两个码字的和仍是码字，而全零 n 维向量 $\mathbf{0}=(00\cdots0)$ 是由信息组 $\boldsymbol{m}=(00\cdots0)$ 生成的码字，对于任意 $l\in F_2$，$v\in V$，则 lV 仍是 V 中的码字，即 (n,k) 线性分组码 V 是 n 维线性空间 $V_n(2^n)$ 的一个 k 维子空间。既然 (n,k) 分组码的 2^k 个码字组成了一个 k 维子空间，那么这 2^k 个码字完全可由 k 个独立向量所组成的基底而生成。设基底为

$$\begin{aligned}\boldsymbol{C}_1&=(g_{1,n-1},g_{1,n-2},\cdots,g_{1,0})\\ \boldsymbol{C}_2&=(g_{2,n-1},g_{2,n-2},\cdots,g_{2,0})\\ &\vdots\\ \boldsymbol{C}_k&=(g_{k,n-1},g_{k,n-2},\cdots,g_{k,0})\end{aligned}$$

若把这组基底写成矩阵形式，则有

$$\boldsymbol{G}=\begin{bmatrix}g_{1,n-1}&g_{1,n-2}&\cdots&g_{1,0}\\ g_{2,n-1}&g_{2,n-2}&\cdots&g_{2,0}\\ \vdots&\vdots&&\vdots\\ g_{k,n-1}&g_{k,n-2}&\cdots&g_{k,0}\end{bmatrix}$$

(n,k) 码的任何码字都可由这组基底的线性组合而生成，即

$$\boldsymbol{C}=\boldsymbol{m}\cdot\boldsymbol{G}=(m_k\quad m_{k-1}\quad\cdots\quad m_0)\begin{bmatrix}g_{1,n-1}&g_{1,n-2}&\cdots&g_{1,0}\\ g_{2,n-1}&g_{2,n-2}&\cdots&g_{2,0}\\ \vdots&\vdots&&\vdots\\ g_{k,n-1}&g_{k,n-2}&\cdots&g_{k,0}\end{bmatrix}$$

若已知信息组 $\boldsymbol{m}$，通过上式可求得相应的码字，称上式中的 $\boldsymbol{G}$ 为 (n,k) 码的生成矩阵。

例如表 5-1 中的 $(6,3)$ 码，可从它的 8 个码字中任意挑选出 $k=3$ 个线性无关的码字，作为码的生成矩阵的行，则

$$\boldsymbol{G}_1=\begin{bmatrix}1&0&0&1&1&0\\ 0&1&0&1&0&1\\ 0&0&1&0&1&1\end{bmatrix}$$

也可以是

$$\boldsymbol{G}_2=\begin{bmatrix}1&0&1&1&0&1\\ 1&1&0&0&1&1\\ 1&1&1&0&0&0\end{bmatrix}$$

下面介绍一下描述分组码的几个参数。

定义 5.7　码率 $R=k/n$ 表示 (n,k) 分组码中，信息位在码字中所占的比重。

定义 5.8　汉明距离 $d(\boldsymbol{a},\boldsymbol{b})$ 表示向量 $\boldsymbol{a}$ 与 $\boldsymbol{b}$ 不相等的对应分量的个数，简称距离。

定义 5.9　n 重向量 $\boldsymbol{v}$ 中非零码元的个数称为汉明重量，用 $W(\boldsymbol{v})$ 来表示。

(n,k) 分组码中，任两个码字之间距离的最小值称为该分组码的最小汉明距离 $d_{\min}$，

简称最小距离，即

$$d_{\min} = \min\{d(v_i, v_j) \mid v_i, v_j \in \boldsymbol{v}, i \neq j\}$$

R 和 $d_{\min}$ 是 (n, k) 分组码的两个重要的参数。纠错编码的基本任务之一就是构造出 R 一定、$d_{\min}$ 尽可能大的码字，或 $d_{\min}$ 一定、R 尽可能高的码字。

5.5.2 最小距离译码

译码器接收到一个二进制序列，而译码器输出的是信息序列 $\boldsymbol{m}$ 的估值序列 $\hat{\boldsymbol{m}}$。译码器的基本任务就是根据某种译码规则，由接收序列 R 给出与发送的信息序列 $\boldsymbol{m}$ 最接近（最好是相同）的估值序列 $\hat{\boldsymbol{m}}$。由于 $\boldsymbol{m}$ 与码字 $\boldsymbol{v}$ 之间存在一一对应的关系，所以这等价于译码器根据 R 产生一个 $\boldsymbol{v}$ 的估值序列 $\hat{\boldsymbol{v}}$。显然，当且仅当 $\hat{\boldsymbol{v}}=\boldsymbol{v}$ 时，$\hat{\boldsymbol{m}}=\boldsymbol{m}$，这时译码器正确译码。

当给定接收序列 R 时，译码器的条件译码错误概率定义为

$$P(E \mid R) = P(\hat{\boldsymbol{v}} \neq \boldsymbol{v} \mid R) \tag{5.46}$$

所以，译码器的错误译码概率：

$$P(E) = \sum_{R} P(E \mid R)P(R) \tag{5.47}$$

$P(R)$ 是接收 R 的概率，与译码方法无关，所以译码错误概率最小的最佳译码规则是使

$$\begin{aligned} &\min P(E) = \min_{R} P(E \mid R) = \min_{R} P(\boldsymbol{v} \neq \boldsymbol{v} \mid R) \\ &\min P(\hat{\boldsymbol{v}} \neq \boldsymbol{v} \mid R) \Rightarrow \max P(\hat{\boldsymbol{v}} = \boldsymbol{v} \mid R) \end{aligned} \tag{5.48}$$

因此，如果译码器对输入的 R，能在 2^k 个码字中选择一个使

$$P(\hat{\boldsymbol{v}}_i = \boldsymbol{v} \mid R) \quad i = 1, 2, \cdots, 2^k$$

最大的码字 $\boldsymbol{v}_i$ 作为 $\boldsymbol{v}$ 的估值序列 $\hat{\boldsymbol{v}}$，则这种译码规则一定使译码器输出错误概率最小，称这种译码规则为最大后验概率译码。

由贝叶斯公式得到：

$$P(\boldsymbol{v} \mid R) = \frac{P(\boldsymbol{v})P(R \mid \boldsymbol{v})}{P(R)}$$

若发端 V 中每个码字等概率发送，即假设先验概率 $P(\boldsymbol{v})$ 相等，则 $P(\boldsymbol{v})=1/2^k$，且由于 $P(R)$ 与译码方法无关，所以后验概率 $P(\boldsymbol{v}|R)$ 最大等价于条件概率 $P(R|\boldsymbol{v})$ 最大，$P(R|\boldsymbol{v})$ 称为似然函数，这种译码准则称为最大似然译码，记为 MLD。显然，最大似然译码等价于最大后验概率译码。

当 r 与 v_j 的距离最小时，认为发方发送 v_j、收到 r 的概率最大，可以据此译码。其依据是下面的定理。

定理 5.6 如果 BSC 信道的转移概率 p 是一个很小的数，则当接收到的 r 与 V 中的码字 v_j 的距离最小，即

$$d(r, v_j) < d(r, v_i) \quad i = 1, 2, \cdots, q^k; \ i \neq j$$

时，发端发送 v_j、收到 r 的条件概率最大，即

$$P(r \mid v_j) > P(r \mid v_i) \quad i = 1, 2, \cdots, q^k; \ i \neq j$$

证明：设 $d(r, v_j)=d$，对于 V 中任意一个码字 $v_i \neq v_j$，令 $d(r, v_i)=d'$，由定理条件知 $d<d'$，于是发送 v_j、收到 r 的条件概率为

$$P(r \mid v_j) = p^d(1-p)^{n-d} \approx p^d$$

发送 v_i、收到 r 的条件概率为

$$P(r \mid v_i) = p^{d'}(1-p)^{n-d'} \approx p^{d'}$$

由于 $d<d'$，且 p 是一个很小的数，因此有

$$p^d > p^{d'},\ P(r|v_j) > P(r|v_i)$$

由此可见，对于硬判决译码，在先验等概的情况下，最大似然译码等价于最小距离译码。

【例 5.4】 一个数字通信系统的误码率为 $P=10^{-4}$，因而收到一个 7 位码元的码字有一位码元有错的概率为

$$P(1) = C_7^1 10^{-4}(1-10^{-4})^6 \approx 7\times 10^{-4}$$

同理，有 2～7 位码元有错的概率分别为

$$P(2) = C_7^2 p^2(1-p)^5 \approx 2.1\times 10^{-7}$$
$$P(3) = C_7^3 p^3(1-p)^4 \approx 3.5\times 10^{-11}$$
$$P(4) = C_7^4 p^4(1-p)^3 \approx 3.5\times 10^{-15}$$
$$P(5) = C_7^5 p^5(1-p)^2 \approx 2.1\times 10^{-19}$$
$$P(6) = C_7^6 p^6(1-p) \approx 7\times 10^{-24}$$
$$P(7) = C_7^7 p^7 \approx 10^{-28}$$

如果不发生错误的概率为 $P(0)$，则

$$P(0) = (1-p)^7 \approx 0.9993$$

由以上结果可以看出：

$$P(0) \gg P(1) \gg \cdots \gg P(7)$$

即发生 t 个错误的概率远远大于发生 $t+1$ 个错误的概率。因此，当我们收到一个 r 时，若它不属于码集 V，则可将 r 与 V 中所有码字作比较。当 r 与 v_j 的距离最小时，发送 v_j、收到 r 的条件概率最大，因而可将 r 判决为 v_j。

实现最小距离译码有一种直观的方法，即计算接收到的码字与所有可能发送的码字之间的距离，取距离最小的码字为译码结果，也就是将收到的码字与码集合当中所有可能的码字进行比较，与哪个码字的距离最小，就判决输出。这种方法固然实现了最小距离译码的思想，但事实上，当码长较长时，计算量相当大，无法实现。

下面介绍一种最简单的译码方法，称为陪集译码。

构造译码表的方法如下：

(1) 表的第一行为 2^k 个码字，全 0 码字放在最左边。

(2) 将所有可能的 n 维向量中重量最小的，且在前面行中没有出现过的向量放在此行的第一列中，将此向量与码向量相加，放到此行相应的列上。

(3) 重复第(2)步，直到 2^n 个向量完全分到表中。

例如，例 5.3 中的(6，3)码的陪集译码表如表 5－2 所示。

表 5－2　例 5.3 中的(6，3)码的陪集译码表

码字	000000 (陪集首)	001011	010101	011110	100110	101101	110011	111000
禁用码	100000	101011	110101	111110	000110	001101	010011	011000
	010000	011011	000101	001110	110110	111101	100011	101000
	001000	000011	011101	010110	101110	100101	111011	110000
	000100	001111	010001	011010	100010	101001	110111	111100
	000010	001001	010111	011100	100100	101111	110001	111010
	000001	001010	010100	011111	100111	101100	110010	111001
	100001	101010	110100	111111	000111	001100	010010	011001

表 5－2 中第一列称为陪集首。通过译码表的构造过程可以知道，每个码字所在列上其他向量都是距离正确码字距离最小的向量(陪集首的重量最小)。如果接收到向量出现了错误，则将它译成所在列的码字(即第一行上的向量)。若接收到向量 011101，则译成 010101。

5.5.3　伴随式译码

下面首先引入校验矩阵的概念。从例 5.3 中很容易得到：

$$\begin{cases} v_5 + v_4 + v_2 = 0 \\ v_5 + v_3 + v_1 = 0 \\ v_4 + v_3 + v_0 = 0 \end{cases} \tag{5.49}$$

若 $V=(v_5, v_4, \cdots, v_0)$是该(6，3)码的一个码字，则 V 的 6 个码元一定满足式(5.49)的三个方程。反之，若有一个 $\boldsymbol{r}=(r_5, r_4, \cdots, r_0)$不完全满足式(5.49)中的三个方程，则它一定不是该(6，3)码的码字，因此式(5.49)中的三个方程对码字起监督作用，我们称这三个方程为监督方程。若用矩阵表示，则有

$$\begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} v_5 \\ v_4 \\ \vdots \\ v_0 \end{bmatrix} = 0$$

若令

$$\boldsymbol{H} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

则 $\boldsymbol{H}$ 称为码的一致监督矩阵，简称监督矩阵。该码的任一个码字 $\boldsymbol{v}$ 都应满足 $\boldsymbol{H}\boldsymbol{v}^{\mathrm{T}}=0$。

事实上，监督方程也可以用更一般的形式表示，即监督方程可以写成如下形式：

$$\begin{aligned} h_{11}v_{n-1} + h_{21}v_{n-2} + \cdots + h_{n1}v_0 &= 0 \\ h_{12}v_{n-1} + h_{22}v_{n-2} + \cdots + h_{n2}v_0 &= 0 \\ &\vdots \end{aligned}$$

$$h_{1r}v_{n-1}+h_{2r}v_{n-2}+\cdots+h_{nr}v_0=0$$

若用矩阵表示，即得：

$$\begin{bmatrix}h_{11} & h_{21} & \cdots & h_{n1}\\ h_{12} & h_{22} & \cdots & h_{n2}\\ \vdots & \vdots & & \vdots\\ h_{1r} & h_{2r} & \cdots & h_{nr}\end{bmatrix}\begin{bmatrix}v_{n-1}\\ v_{n-2}\\ \vdots\\ v_0\end{bmatrix}=0$$

或者 $\boldsymbol{H}\boldsymbol{v}^{\mathrm{T}}=0$，这里

$$\boldsymbol{H}=\begin{bmatrix}h_{11} & h_{21} & \cdots & h_{n1}\\ h_{12} & h_{22} & \cdots & h_{n2}\\ \vdots & \vdots & & \vdots\\ h_{1r} & h_{2r} & \cdots & h_{nr}\end{bmatrix}$$

仍旧称为该码的监督矩阵，它的 r 行是线性无关的。

从监督矩阵的推导过程知道，它的每一行表示求一个监督元的监督方程，因为每个码字共有 $r=n-k$ 个监督元，所以任何一个(n,k)线性分组码的监督矩阵 $\boldsymbol{H}$ 至少有 $r=n-k$ 行，且有 r 行是线性无关的。

我们知道，线性分组码的纠错能力与它的最小距离 $d_{\min}$有着密切的关系，$d_{\min}$越大，它的纠错能力越强，因此讨论 $\boldsymbol{H}$ 与 $d_{\min}$的关系就是讨论 $\boldsymbol{H}$ 与码的纠错能力的关系。

定理 5.7　一个(n,k)线性分组码的最小距离为 $d_{\min}$的充分必要条件是它的监督矩阵 $\boldsymbol{H}$ 的任意 $d_{\min}-1$ 列线性无关，而有 $d_{\min}$列线性相关。

证明：先证必要性。如果 (n,k)码的最小距离为 $d_{\min}$，则由分组码的性质可知，最小码重 $W_{\min}=d_{\min}$，于是存在一个码字 $\boldsymbol{v}=(v_{n-1},v_{n-2},\cdots,v_0)$，它的重量为 $d_{\min}$，即 $\boldsymbol{v}$ 有 $d=d_{\min}$个码元不为 0，不妨设 v_{n-i_1}、v_{n-i_2}、…、v_{n-i_d}不为 0，其余码元为 0，设码的监督矩阵为

$$\boldsymbol{H}=[\alpha_{n-1}\alpha_{n-2}\cdots\alpha_0]$$

其中 α_{n-i}表示 $\boldsymbol{H}$ 矩阵的第 i 个列向量，于是由 $\boldsymbol{H}\boldsymbol{v}^{\mathrm{T}}=0$ 得到

$$\boldsymbol{H}\boldsymbol{v}^{\mathrm{T}}=\alpha_{n-i_1}v_{n-i_1}+\alpha_{n-i_2}v_{n-i_2}+\cdots+\alpha_{n-i_d}v_{n-i_d}=0$$

即 $\boldsymbol{H}$ 中有 $d=d_{\min}$列线性相关。

如果 $\boldsymbol{H}$ 中有 $d_{\min}-1=d$ 列也线性相关，不妨设

$$\alpha_{n-p_1}v_{n-p_1}+\alpha_{n-p_2}v_{n-p_2}+\cdots+\alpha_{n-p_d}v_{n-p_d}=0$$

于是存在一个码字 $\boldsymbol{v}$，它的码元 v_{n-p_1}、v_{n-p_2}、…、v_{n-p_d}不为 0，其余码元为 0，使

$$\boldsymbol{H}\boldsymbol{v}'^{\mathrm{T}}=\alpha_{n-1}v_{n-1}+\alpha_{n-2}v_{n-2}+\cdots+\alpha_0v_0=0$$

即存在一个码字 $\boldsymbol{v}$，$W(\boldsymbol{v})=d=d_{\min}-1$，这与 $W_{\min}=d_{\min}$矛盾，因此 $\boldsymbol{H}$ 中任意 $d_{\min}-1$ 列都线性无关。

下面证明充分性。设 $\boldsymbol{H}$ 中有 $d_{\min}$列线性相关，不妨设 v_{n-i_1}、v_{n-i_2}、…、$v_{n-i_{d_{\min}}}$不为 0，使

$$\alpha_{n-i_1}v_{n-i_1}+\alpha_{n-i_2}v_{n-i_2}+\cdots+\alpha_{n-i_{d_{\min}}}v_{n-i_{d_{\min}}}=0 \tag{5.50}$$

我们令 $\boldsymbol{v}=(v_{n-1},v_{n-2},\cdots,v_0)$，它的码元 v_{n-i_1}、v_{n-i_2}、…、$v_{n-i_{d_{\min}}}$不为 0，其余码元为 0。由式(5.50)知 $\boldsymbol{H}\boldsymbol{v}^{\mathrm{T}}=0$，即该$(n,k)$码有一个码字 $\boldsymbol{v}$，重量 $W(\boldsymbol{v})=d_{\min}$。

最后证明码中没有重量小于等于 $d_{\min}-1$ 的码字。为此设有一个码字 $\boldsymbol{v}$，它的重量为 $W(\boldsymbol{v})=d=d_{\min}-1$，不妨设 v_{n-p_1}、v_{n-p_2}、…、v_{n-p_d} 不为 0，其余码元为 0，于是

$$\boldsymbol{H}\boldsymbol{v}'^{\mathrm{T}}=\alpha_{n-p_1}v_{n-p_1}+\alpha_{n-p_2}v_{n-p_2}+\cdots+\alpha_{n-p_d}v_{n-p_d}=0$$

这与 $\boldsymbol{H}$ 中任意 $d_{\min}-1$ 列线性无关矛盾。

推论 3 当 (n,k) 码的最小距离 $d_{\min}\geqslant 2t+1$ 时，它的 $\boldsymbol{H}$ 矩阵任意 t 列的线性组合与其他任意 t 列的线性组合均不相同。

设 V 是 (n,k) 线性分组码，$\boldsymbol{v}$ 是 V 中任一个码字，当发端发送 $\boldsymbol{v}$ 时，收端收到一个 $\boldsymbol{r}$，错误图样为

$$\boldsymbol{e}=\boldsymbol{r}-\boldsymbol{v}=(e_{n-1},e_{n-2},\cdots,e_0)$$

于是定义

$$\boldsymbol{s}^{\mathrm{T}}=\boldsymbol{H}\boldsymbol{r}^{\mathrm{T}}=\boldsymbol{H}\boldsymbol{e}^{\mathrm{T}} \tag{5.51}$$

从式(5.51)可以看出，$\boldsymbol{s}^{\mathrm{T}}$ 仅与错误图样 $\boldsymbol{e}$ 有关，而与发送的码字无关。我们称 $\boldsymbol{s}$ 为 $\boldsymbol{r}$ 或 $\boldsymbol{e}$ 的伴随式。$\boldsymbol{s}$ 完全由 $\boldsymbol{e}$ 确定，它在一定程度上反映了信道的干扰情况。译码器的主要任务就是如何从 $\boldsymbol{s}$ 求得错误图样，从而译出发送的码字 $\boldsymbol{v}$。

设一个最小距离 $d_{\min}=3$、纠错能力 $t=1$ 的 (n,k) 码的监督矩阵为

$$\boldsymbol{H}=\begin{bmatrix}h_{11} & h_{21} & \cdots & h_{i1} & \cdots & h_{n1}\\ h_{12} & h_{22} & \cdots & h_{i2} & \cdots & h_{n2}\\ \vdots & \vdots & & \vdots & & \vdots\\ h_{1r} & h_{2r} & \cdots & h_{ir} & \cdots & h_{nr}\end{bmatrix}$$

当收到的 $\boldsymbol{r}$ 无错，即 $\boldsymbol{r}=\boldsymbol{v}$ 时，$\boldsymbol{e}=\boldsymbol{0}$，一定有 $\boldsymbol{s}^{\mathrm{T}}=\boldsymbol{H}\boldsymbol{r}^{\mathrm{T}}=\boldsymbol{H}\boldsymbol{e}^{\mathrm{T}}=\boldsymbol{0}$。当接收的 $\boldsymbol{r}$ 有 1 位错时，设错误图样为 $\boldsymbol{e}=(00\cdots0e_i0\cdots00)$，则伴随式为

$$\boldsymbol{s}^{\mathrm{T}}=\begin{bmatrix}h_{11} & h_{21} & \cdots & h_{i1} & \cdots & h_{n1}\\ h_{12} & h_{22} & \cdots & h_{i2} & \cdots & h_{n2}\\ \vdots & \vdots & & \vdots & & \vdots\\ h_{1r} & h_{2r} & \cdots & h_{ir} & \cdots & h_{nr}\end{bmatrix}\begin{bmatrix}\vdots\\ 0\\ e_i\\ 0\\ \vdots\end{bmatrix}=\begin{bmatrix}h_{i1}\\ h_{i2}\\ \vdots\\ h_{ir}\end{bmatrix}$$

可见，伴随式正好是 $\boldsymbol{H}$ 矩阵的第 i 列，它是 r 维列向量。

又因为码的最小距离 $d_{\min}=3$，$\boldsymbol{H}$ 矩阵任何两列都不相同，所以伴随式与单个错误的错误图样一一对应。

如果发生了两个错误，设错误图样为 $\boldsymbol{e}=(00\cdots0e_i0\cdots0e_j0\cdots0)$，则伴随式为 $\boldsymbol{H}$ 矩阵的第 i 列和第 j 列之和。但是，此时的伴随式也可能是另外两列之和(请考虑 $\boldsymbol{H}$ 矩阵列相关性和最小距离之间的关系)，所以此时只能发现错误，但是不能纠正错误。

纠正单个错的伴随式译码方法如下：

(1) 如果 $\boldsymbol{s}^{\mathrm{T}}=\boldsymbol{0}$，则判 $\boldsymbol{r}$ 无错，此时 $\boldsymbol{v}=\boldsymbol{r}$；

(2) 如果 $\boldsymbol{s}^{\mathrm{T}}\neq\boldsymbol{0}$，且 $\boldsymbol{s}^{\mathrm{T}}$ 与 $\boldsymbol{H}$ 矩阵的第 i 列相同，则判第 i 位有错；

(3) 如果 $\boldsymbol{s}^{\mathrm{T}}\neq\boldsymbol{0}$，但不是 $\boldsymbol{H}$ 的列，则只判 $\boldsymbol{r}$ 有错，而不纠错。

纠正多个错误的伴随式译码的原理跟单个错误情况相同，这里不再赘述。

5.5.4　分组码最小距离的边界

在离散二进制对称信道上，人们在研究分组码的误码率边界时得出了以下结论：

$$P_e \leqslant (M-1)[4p(1-p)]^{d_{\min}/2} \tag{5.52}$$

$$P_e \geqslant \sum_{m=[d_{\min}/2]+1}^{d_{\min}} \binom{d_{\min}}{m} p^m (1-p)^{d_{\min}-m} \tag{5.53}$$

因此，$d_{\min}$的上、下边界在评价码的性能时十分重要。

由 5.5.3 节可知，分组码的校验矩阵 $\boldsymbol{H}$ 中有 $d_{\min}$ 列线性相关，而 $d_{\min}-1$ 列是线性无关的，由于 $\boldsymbol{H}$ 的秩最多为 $n-k$，则必有 $n-k \geqslant d_{\min}-1$，所以 $d_{\min}$的上界是：

$$d_{\min} \leqslant n-k+1 \tag{5.54}$$

式(5.54)是 Singleton 给出的最简单的上边界，称为 Singleton 限。

用分组长度 n 可将此式归一化，即

$$\frac{d_{\min}}{n} \leqslant (1-R_c) + \frac{1}{n}$$

式中，R_c 是码率。当 n 很大时，因子 $1/n$ 可以忽略。

如果某一编码具有最大的可能距离，即 $d_{\min}=n-k+1$，则这种码叫做极大最小距离可分码。除了普通重复型码外，不存在二进制的极大最小距离可分码。事实上，式(5.54)规定的上边界对二进制码来说是非常松散的。另一方面，具有 $d_{\min}=n-k+1$ 的非二进制码却是存在的。例如里德-索罗门(Reed-Solomon)码就是极大最小距离可分码，它属于 BCH 码的子类。

除了上述上边界之外，还有几个线性分组码的最小距离较紧的边界。下面我们将简述 4 种重要的边界，其中三个是上边界，一个是下边界。这些边界的推导冗长，这里不进行详述，有兴趣的读者请参阅彼得森和韦尔登(1972 年)著作的第 4 章。

第一个最小距离的上边界为

$$1-R_c \geqslant \frac{1}{n} \log \sum_{i=0}^{t} \binom{n}{i} \tag{5.55}$$

由于码的纠错能力(用 t 度量)与最小距离有关，所以上述关系式是最小距离的上边界，叫做 Hamming 上边界或 Hamming 限。

令 $n \to \infty$，可得到式(5.55)的渐近形式。对于任意 n，令 t_0 是使式(5.55)成立的 t 的最大整数，可以证明，当 $n \to \infty$时，任何(n, k)分组码 t/n 决不会超过 t_0/n。这里，t_0/n 满足等式：

$$1-R_c = H\left(\frac{t_0}{n}\right) \tag{5.56}$$

式中，$H(x)$是的二进制熵函数 $H(x)=-p\log p-(1-p)\log(1-p)$，其中 p 为 0 的概率。

将 Hamming 界推广到非二进制码，即

$$1-R_c \geqslant \frac{1}{n} \log_q \left[\sum_{i=0}^{t} \binom{n}{i} (q-1)^i \right] \tag{5.57}$$

另一个最小距离的上边界是普洛特金(Plotkin)1960 年推出的，后人称之为 Plotkin 限。说明如下：在一个(n, k)线性分组码中，为了达到最小距离 $d_{\min}$，所需的校验位数目必

须满足不等式：

$$n-k \geqslant \frac{qd_{\min}-1}{q-1}-1-\log_q d_{\min} \tag{5.58}$$

其中，q 是字符集大小。对二进制编码来说，式(5.58)可表示为

$$\frac{d_{\min}}{n}\left(1-\frac{1}{2d_{\min}}\log d_{\min}\right) \leqslant \frac{1}{2}\left(1-R_c+\frac{2}{n}\right)$$

在 $d_{\min}/n \leqslant 1/2$ 而 $n \to \infty$ 的极限情况下，式(5.58)简化为

$$\frac{d_{\min}}{n} \leqslant \frac{1}{2}(1-R_c) \tag{5.59}$$

最后一个紧的最小距离上边界是埃利斯(Elias，1968 年)发现的，以渐近线形式表示为

$$\frac{d_{\min}}{n} \leqslant 2A(1-A) \tag{5.60}$$

式中，参数 A 与码率有关，两者之间的关系式为

$$R_c = 1+A\log A+(1-A)\log(1-A) \qquad 0 \leqslant A \leqslant \frac{1}{2} \tag{5.61}$$

(n, k)分组码最小距离的下边界同样存在，特别是具有归一化最小距离的二进制分组码，它们渐近地满足不等式：

$$\frac{d_{\min}}{n} \geqslant \alpha \tag{5.62}$$

式中，α 通过下列等式与码率建立联系：

$$\begin{aligned} R_c &= 1-H(\alpha) \\ &= 1+\alpha\log\alpha+(1-\alpha)\log(1-\alpha) \qquad 0 \leqslant \alpha \leqslant \frac{1}{2} \end{aligned} \tag{5.63}$$

这个下边界是吉尔伯特(Gilbert，1952 年)和乌沙莫夫(Varsharmov，1957 年)提出的下边界的一个特例，适用于非二进制和二进制分组码，简称 V－G 限。

上述三个性能限中，前两个是必要条件，即想让 $d_{\min}$ 达到某值，$n-k$ 必须大于某值，或者说对于一定冗余度 $n-k$，$d_{\min}$ 最大能达到多少，所以 Hamming 限和 Plotkin 限给出的是上边界。V－G 限是充分条件，它说明只要满足这个条件，就一定存在最小距离为 $d_{\min}$ 的码，所以 V－G 限给出的是下边界。然而，V－G 限指出的“存在”并不一定能被找到。如图 5－3 所示，斜线区是可能实现的区域，双重斜线区是必能实现的区域。

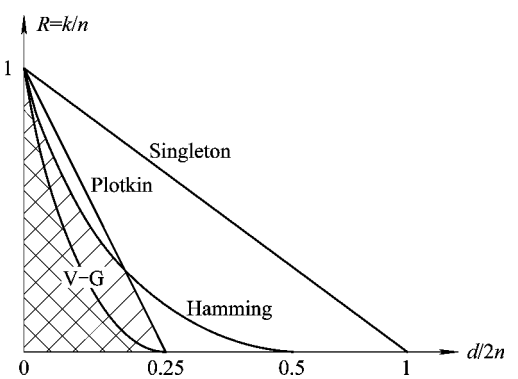

图 5－3 几种码限的比较

5.6 卷 积 码

卷积码(convolutional code)是 1955 年由 P. Elias 提出的一种记忆码，在二十世纪六七十年代开始广泛使用。它与分组码不一样，不仅是输入 k_0 位信息码元的函数，还是前面 m 个 k_0 位信息码元的函数。卷积码由 n_0、k_0、m 描述，n_0 为每组输出码元数，k_0 为每组输入码元数，k_0/n_0 表示卷积码的编码效率，m 称为编码约束度，表示 k_0 组移位寄存器中最大的级数。由于卷积码各组之间相互有关，因此在卷积码的分析过程中，未找到像分组码那样有效的数学工具，性能分析比较困难，从分析上得到的成果不像分组码那样多，往往要借助计算机的搜索来寻找好码。

【例 5.5】 一个简单的卷积码如图 5－4 所示。初始状态为 00。

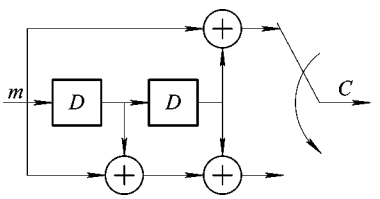

图 5－4　卷积码生成器

设输入 $m=101100\cdots$，则输出与输入的关系为

输入：	1	0	1	1	0	0	0
状态：	00	10	01	10	11	01	00
输出：	11	01	00	10	10	11	00

卷积码的特点如下：

(1) 当前码分组输出不仅与当前信息分组输入有关，还与前面 m 个信息分组有关。

(2) 尚没有完善的数学工具有效地分析其结构和性能，必须借助计算机搜索来寻找好码。

(3) 研究表明：卷积码是渐近好码，相同码率、相同译码复杂性条件下，卷积码的性能要好于分组码。

(4) 卷积码仍是线性码，满足线性叠加关系。

例 5.5 中的编码器一共有 4 个状态：00、10、01、11，分别记为：S0、S1、S2、S3，该编码器的状态转移图如图 5－5 所示。

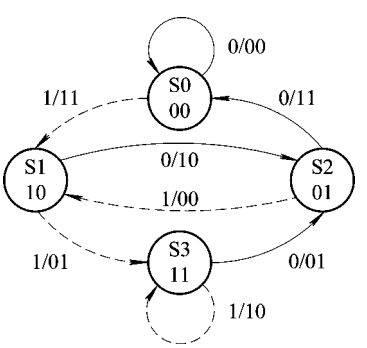

图 5－5　状态转移图

卷积码的状态图只表示编码状态之间的转移关系，无法表示状态转移与时间节拍的关系。为了表示状态转移与时间节拍的关系，引入卷积码的格图(Trellis Diagram)表示。卷积码的格图也称为篱笆图。

从初始状态出发，格图上的每一条路径都对应着一个输入信息序列所对应的编码序列。给定信息序列，可在格图上找到一条路径，进而得到所对应的编码序列。

反过来，给定编码序列，也可在格图上找到一条路径，进而得到所对应的信息序列。

例 5.5 中，(2，1，2)卷积码将状态转移图按时间节拍展开，如图 5－6 所示。

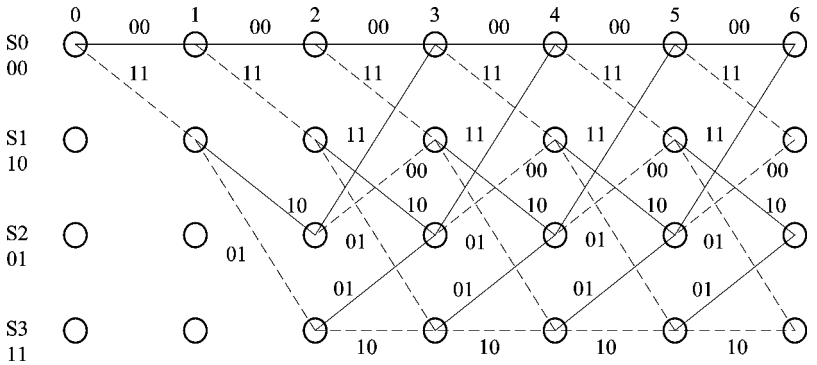

图 5－6 (2，1，2)卷积码的网格图

Shannon 在证明 Shannon 第二编码定理时引用了如下 3 个基本条件：

(1) 采用随机性编译码。

(2) 编码长度趋于无穷，即分组的码组长度无限。

(3) 译码过程采用最佳的最大似然译码(ML)方案。

在信道编码的研究与发展过程中，基本上以后两个条件为主要方向，如卷积码使得码长变得很长，并采用次最优的最大似然译码(viterbi 算法)。对于条件(1)，虽然随机选择编码码字可以使获得好码的概率增大，但是最大似然译码器的复杂度随码字数目的增大而加大，当编码长度很大时，译码几乎不可能实现。总体而言，目前各种单一的构造性很强的编译码方法其性能都很有限，与信道容量之间的差距是很大的，这也就是为什么信息论发展半个多世纪以来人们关心的容量仍不是信息论意义上的容量。因此，多年来随机编码理论一直作为分析和证明编码定理的主要方法，而如何在构造码上发挥作用却并未引起人们的足够重视。直到 1993 年，Turbo 码的发现才较好地解决了这一问题，为 Shannon 随机码理论的应用研究奠定了基础。

5.7 Turbo 码

Turbo 码又称为并行级联卷积码(PCCC)，它通过在编码器中引入随机交织器，使码字具有近似随机的特性；通过分量码的并行级联，实现通过短码(分量码)构造长码(Turbo 码)；在接收端虽然采用了次最优的迭代算法，但分量码采用的是最优的最大后验概率译码算法，同时通过迭代过程可使译码接近最大似然译码。综合上述分析可见，Turbo 码充分考虑了 Shannon 信道编码定理证明时所假设的条件，从而获得了接近 Shannon 理论极限的性能。

Turbo 码的最大特点在于它通过在编译码器中使用交织器和解交织器，有效地实现了随机性编译码的思想，通过短码的有效结合实现长码，从而达到了接近 Shannon 理论极限的性能。

Turbo 码编码器是由两个递归系统卷积码(RSC)编码器通过一个随机交织器并行连接

而成的。编码后的校验位经过删余矩阵，从而产生不同码率的码字，见图 5－7。

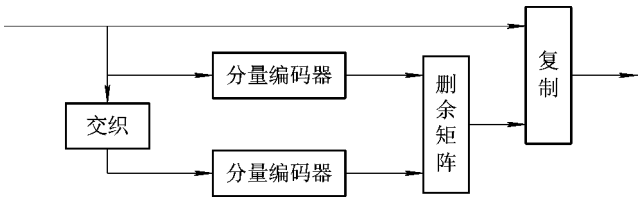

图 5－7　Turbo 码编码器的结构框图

在 Turbo 码编码过程中，信息序列 $\boldsymbol{u}=(u_1, u_2, \cdots, u_N)$ 经过一个 N 位交织器，形成一个新序列 $\boldsymbol{u}_1=(u_1', u_2', \cdots, u_N')$（长度与内容没变，但比特位置经过重新排列）。$\boldsymbol{u}$ 与 $\boldsymbol{u}_1$ 分别送到两个分量编码器，同时 $\boldsymbol{u}$ 作为系统输出 $\boldsymbol{X}^s$ 直接送至复接器。一般情况下，这两个编码器结构相同，生成序列 $\boldsymbol{X}^{p1}$ 与 $\boldsymbol{X}^{p2}$。为了提高码率，序列 $\boldsymbol{X}^{p1}$ 与 $\boldsymbol{X}^{p2}$ 需要经过删余矩阵，采用删余技术从这两个校验序列中周期地删除一些校验位，形成校验位序列 $\boldsymbol{X}^p$。$\boldsymbol{X}^p$ 与未编码序列 $\boldsymbol{X}^s$ 经过复用调制后，生成 Turbo 码序列 $\boldsymbol{X}$。

编码器中交织器的使用是实现 Turbo 码近似随机编码的关键。交织器实际上是一个一一映射函数，其作用是将输入信息序列中的比特位置进行重置，以减小分量编码器输出校验序列的相关性和提高码重。

删余矩阵的作用是提高编码码率，其元素取自集合{0，1}。矩阵中每一行分别与两个分量编码器相对应。其中，“0”表示相应位置上的校验比特被删除，而“1”则表示保留相应位置的校验比特。

下面通过一个具体实例来说明 Turbo 码的编码过程。

图 5－8 是一个码率为 1/3 的 Turbo 码编码器的组成框图。

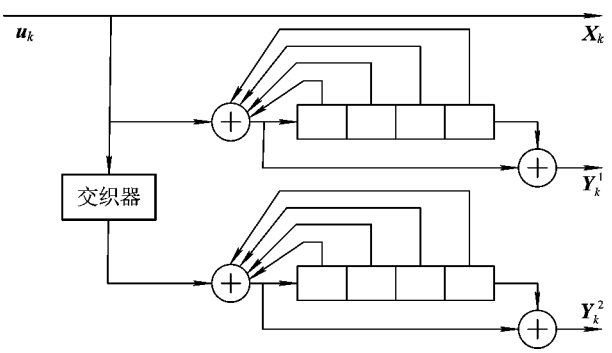

图 5－8　一个码率为 1/3 的 Turbo 码编码器的组成框图

这个编码器是基于(2，1，4)RSC(递归系统卷积码)的 Turbo 码编码器，分量码是码率为 1/2、寄存器级数为 4 的（2，1，4）RSC 码，生成的多项式为 $(1+D+D^2+D^3+D^4, 1+D^4)$。

假设输入序列为

$$\boldsymbol{u}_k = (1011001) \tag{5.64}$$

则第一个分量码的输出序列为

$$\begin{cases} \boldsymbol{X}_k = (1011001) \\ \boldsymbol{Y}_k^1 = (1110001) \end{cases} \tag{5.65}$$

假设经过交织器后信息序列变为

$$\tilde{\boldsymbol{u}}_k = (1101010) \tag{5.66}$$

第二个分量码编码器所输出的校验位序列为

$$\boldsymbol{Y}_k^2 = (1000000) \tag{5.67}$$

则得到 Turbo 码序列为

$$\boldsymbol{v} = (111, 010, 110, 100, 000, 000, 110) \tag{5.68}$$

迄今为止，Turbo 码以及低密度奇偶校验码(LDPC)已经距 Shannon 限非常近。模拟结果表明，如果采用大小为 65 535 的随机交织器，并进行 18 次迭代，则在 $E_b/N_0 \geqslant 0.7$ dB 时，码率为 1/2 的 Turbo 码在 AWGN 信道上的误码率(BER)$\leqslant 10^{-5}$，接近 Shannon 限(1/2 码率的 Shannon 限为 0 dB)。

编码理论经过了近 60 年的发展，早已成长为信息论发展的一个庞大的自成体系的学科。本书的重点内容是信息论基础，所以无法展开叙述。本章涉及到信道编码的内容只是期望能起到抛砖引玉的作用，了解信息论的发展对信道编码的指导意义，有兴趣的读者可以参考其他专业书籍，进一步学习信道编码的内容。

习 题 5

5.1 设有码 $\boldsymbol{C}$，在用此码传输时能使 $H(\boldsymbol{U}|\boldsymbol{Y})=0$(式中 $\boldsymbol{U}$ 表示码字矢量，$\boldsymbol{Y}$ 代表接收矢量)，试证此时的码率 R 必小于该信道的信道容量 C。(提示：$H(\boldsymbol{U})=H(\boldsymbol{U}|\boldsymbol{Y})+I(\boldsymbol{U};\boldsymbol{Y})$)

5.2 设有码表示如下：

信息	码字
00	00000
01	01101
10	10111
11	11010

(1) 找出生成矩阵 $\boldsymbol{G}$ 与监督矩阵 $\boldsymbol{H}$；

(2) 在二进制对称信道(BSC)下给出最大似然译码的译码表；

(3) 求正确译码的概率。

5.3 试证：

(1) 二元线性码中码字重必全为偶数或奇偶各半；

(2) (n, k)码的平均码字重不超过 $n/2$。

5.4 设 X 和 Y 为信道的输入和输出，两者均取值于集合 $A=\{a_1, a_2, \cdots, a_K\}$。已知 $P(x=a_k)=p_k$，$P(y=a_j \mid x=a_k)=p_{kj}$，其中 $k=1, 2, \cdots, K$，定义 $P_e = \sum\limits_k p_k \sum\limits_{j\neq k} p_{kj}$。求证：

$$H(X \mid Y) \leqslant P_e \log(K-1) + H(P_e)$$

其中，$H(P_e)=-P_e \log P_e-(1-P_e)\log(1-P_e)$。

第 6 章　离散信源及其信息冗余

6.1　信源的描述与分类

信源是发出信息的某种设备，可以是人、生物、机器或其他任何向外发出信息的事物。信源的输出称做消息。在人类的社会活动中，发出信息的信息源多种多样，其输出可以是离散的符号，如书信中的文字和字母，也可以是连续的信号，如人发出的语声。

在信息理论的研究中，人们感兴趣的仅仅是载荷着信息的信源输出的符号或信号，而对于信源内部的结构和产生输出符号的控制过程一般不需要作深入了解。例如，人用语言方式向外界发出信息，便是由大脑指挥发声器官发出声音，以“音”的排列构成符号(消息)序列输出某种信息。研究这样的输出语音的信源(人)时，我们并不需要考虑人脑控制发声器官发出声音并构成某种“音”的排列的复杂过程，而只需要分析它的输出，即载荷着信息的符号(消息)。因此，在下面的讨论中，我们并不研究信源的内部结构和产生输出符号的复杂关系，只是将信源看做一个发出某种符号(序列)的设备(见图 6－1)。

图 6－1　信源模型

由于信息的基本属性是随机性，因此信源输出的符号或符号序列具有随机性，信源输出的符号需要使用随机变量、随机矢量或随机过程表示。所以，信源的描述与分类也仅仅考虑信源的外部特性，即依据信源的输出——符号(序列)所具有的特点和统计关系加以描述和分类。

6.1.1　信源的描述

由于信源输出的消息载荷着信息，这种消息所具有的一个基本属性便是随机性，因此信源输出的符号或符号序列可以使用随机变量、随机矢量或随机过程表示。由第 2 章的讨论我们知道，如果已知信源的消息集合(即样本空间或值域)和消息发生的概率分布，则可以使用由样本空间和它的概率分布所构成的概率空间来描述信源。

设信源输出随机变量 X 的样本空间(值域)为 R。若在此值域上随机变量 X 的概率分布为 P，则此信源的概率空间为$[R, P]$或$[X, P]$。

6.1.2　信源的分类

信源的输出是随机变量，信源的描述是一个统计模型，那么对信源进行分类也应从信源的统计特性出发，即依据信源输出的消息在取值范围、概率分布及统计关系等方面的不

同特点和性质，将信源划分为不同的类型。

1. 连续信源与离散信源

根据信源输出消息 X 的取值特点，可将信源划分为连续信源和离散信源。

1）离散信源

信源输出符号为离散随机变量的信源称为离散信源。

设离散信源输出随机变量 X 的值域 R 为一离散集合 $R=\{a_1, a_2, \cdots, a_n\}$，其中，$n$ 可以是有限正数，也可以是可数的无限大正数。若已知 R 上每一消息发生的概率分布为

$$P(a_1), P(a_2), \cdots, P(a_n)$$

则离散信源 X 的概率空间为

$$[R, P]=[X, P]=\begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ P(a_1) & P(a_2) & \cdots & P(a_n) \end{bmatrix} \tag{6.1}$$

其中，信源输出消息的概率 $P(a_i)(i=1, 2, \cdots, n)$ 满足：

$$\begin{cases} P(a_i) \geqslant 0 \quad i=1, 2, \cdots, n \\ \sum\limits_{i=1}^{n} P(a_i)=1 \end{cases}$$

2）连续信源

如果信源输出符号的种类是不可数的无限值，即信源输出随机变量 X 的取值为一个连续区间，那么这样的信源称为连续信源。

假设随机变量 X 的取值是在一个连续区间 (a, b) 内，即 $X \in \mathbf{R}$，而 $\mathbf{R}=(a, b)$（当 $a \to -\infty$，$b \to +\infty$ 时，这个区间成为一个包含所有实数的集合）。如果 $\mathbf{R}$ 内 X 的概率密度分布为 $p(x)$，则连续信源的概率空间为

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

其中：

$$p(x) \geqslant 0 \quad x \in \mathbf{R}$$

且

$$\int_{\mathbf{R}} p(x)\mathrm{d}x=1$$

或

$$\int_a^b p(x)\mathrm{d}x=1$$

3）混合信源

若信源的一些符号取值于一个连续区间，而另一些符号取值于离散集合，则这样的信源称为混合信源。

2. 有记忆信源与无记忆信源

实际信源的输出往往是由信源输出的一系列符号所组成的一个符号序列：

$$\cdots X_{-1} X_0 X_1 \cdots X_N X_{N+1} \cdots$$

信源输出的符号序列为一个随机变量序列，可以用随机矢量来表示。例如，若将英文文字信源的取值集合看做 26 个英文字母和空格及标点符号，则使用一段英文文字表达具有某

种含义的信息时，该信源的输出将是一个由不同字母、空格和标点符号构成的消息序列。

由于信源输出的符号序列是一个随机变量序列，因此可以用随机矢量来表示。设信源输出为一个 N 维随机矢量

$$\boldsymbol{X}=(X_1, X_2, \cdots, X_N)$$

其中：

$$X_i \in \mathbf{R}(i=1, 2, \cdots, N)。$$

N 维随机矢量 $\boldsymbol{X}$ 的统计特性需要使用联合概率密度函数：

$$P(X_1=x_1, X_2=x_2, \cdots, X_N=x_N)=P(x_1, x_2, \cdots, x_N)=P(\boldsymbol{x})$$

来描述。这里随机变量的样值为

$$x_i \in \mathbf{R} \quad i=1, 2, \cdots, N$$

于是，根据信源输出的随机矢量中各分量之间是否存在相关性，可将信源划分为无记忆信源和有记忆信源。

1）无记忆信源

如果信源输出的随机矢量中，各元素 X_i 相互独立，则信源是无记忆的。

对于输出 N 维随机矢量的离散无记忆信源，如果各元素 X_i 具有相同的概率密度，则其联合概率密度可表示为

$$P(\boldsymbol{x})=P(x_1, x_2, \cdots, x_N)=\prod_{i=1}^{N}P(x_i) \tag{6.2}$$

2）有记忆信源

若随机矢量 $\boldsymbol{X}=(X_1, X_2, \cdots, X_N)$ 中各分量之间存在某种关联，则信源是有记忆信源。有记忆信源需要使用条件概率 $P(x_N|x_1, x_2, \cdots, x_{N-1})$ 描述其统计关系。

显然，与无记忆信源相比，有记忆信源的描述要困难得多。当信源输出符号之间的记忆长度较大时，有记忆信源的描述与分析将更加困难。

3. 平稳信源与非平稳信源

对于更一般的信源，其输出实际上是时间 t 的一个连续函数 $X(t)$，或者是一个任意的消息序列 $X_1X_2\cdots X_i\cdots$，即信源的输出是随机过程或随机序列。对这种信源的分析需要应用随机过程理论。对于一个随机过程或随机序列，根据其统计特性是否随着时间的平移而改变可将其划分为平稳随机过程和非平稳随机过程。相应地，信源也可划分为平稳信源或非平稳信源。

如果信源输出的随机变量序列 $X_1X_2\cdots X_i\cdots$ 是平稳的随机序列，则该信源是平稳信源，否则该信源是非平稳信源。

6.2　离散无记忆信源的扩展信源

实际信源的输出往往是一个相当长的符号序列。例如，电报系统中的二进制信源的输出是一个由二进制符号 0、1(点、画)构成的数据序列。灰度图像信源输出有 256 种取值可能，图像信源输出的一幅灰度图像需要由若干行、若干列的像素组成，每一像素的取值可能为 0～255。显然，实际信源输出的任意长度符号序列的统计关系复杂，其输出信息特性的分析非常困难。此处，我们首先从最简单的信源入手，针对离散无记忆信源所输出的离散随机变量序列，分析其统计关系、信息度量及基本关系。

设离散无记忆信源输出离散随机变量序列为$\cdots X_1X_2\cdots X_i\cdots$，为了便于分析此信源输出信息的特性，将信源输出的随机变量分组(如取 N 个随机变量为一组)，并且将每组随机变量表示为一个随机矢量。于是，输出随机变量序列的信源可以等效为一个输出 N 维随机矢量的新信源，并且将这样的新信源称做原信源的 N 次扩展信源。

例如，将输出 0、1 符号的电报系统看做一个二进制离散无记忆信源 X。若将该信源 X 输出的数据序列…0100011000010100 0…中的每两个二进制符号划分为一组，则可等效为输出二维随机矢量的新信源。信源输出二维随机矢量 $\boldsymbol{X}\in\{00,\ 01,\ 10,\ 11\}$。这个输出二维随机矢量 $\boldsymbol{X}$ 的新信源 X^2 称为二进制离散无记忆信源 X 的二次扩展信源。

对于一般的离散无记忆信源 X，定义其 N 次扩展信源如下：

设有一离散无记忆信源：

$$[X,\ P]=\begin{bmatrix} a_1 & a_2 & \cdots & a_r \\ P(a_1) & P(a_2) & \cdots & P(a_r) \end{bmatrix}$$

若将其输出的符号序列用一组长度为 N 的矢量序列表示，则可等效为一个新信源。新信源输出的符号是长度为 N 的原符号序列，用 N 维随机矢量表示 $\boldsymbol{X}$，其样矢量为

$$\boldsymbol{x}=x_1x_2\cdots x_N \qquad x_i\in X;\ i=1,\ 2,\ \cdots,\ N$$

显然，离散无记忆信源 X 输出的长度为 N 的符号序列(N 维随机矢量 $\boldsymbol{X}$)可以组成 r^N 种不同的样矢量，并且样矢量中的各符号相互独立。

定义输出 N 维随机矢量 $\boldsymbol{X}$ 的新信源为离散无记忆信源 X 的 N 次扩展信源 X^N。X^N 的概率空间为

$$[X^N,\ P(\boldsymbol{x})]=\begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_{r^N} \\ P(\boldsymbol{x}_1) & P(\boldsymbol{x}_2) & \cdots & P(\boldsymbol{x}_{r^N}) \end{bmatrix} \tag{6.3}$$

其中：

$$\begin{cases} 0\leqslant P(\boldsymbol{x}_i)\leqslant 1 & i=1,\ 2,\ \cdots,\ r^N \\ \sum\limits_{i=1}^{r^N}P(\boldsymbol{x}_i)=1 \end{cases}$$

由于信源 X 是离散无记忆的，扩展信源 X^N 输出的样矢量 $\boldsymbol{x}=x_1x_2\cdots x_N$ 中各分量统计独立，因此有：

$$P(\boldsymbol{x})=P(x_1,\ x_2,\ \cdots,\ x_N)=P(x_1)P(x_2)\cdots P(x_N)=\prod_{j=1}^{N}P(x_j)$$

于是，N 次扩展信源 X^N 的熵为

$$H(\boldsymbol{x})=H(X^N)=-\sum_{X^N}P(\boldsymbol{x})\log P(\boldsymbol{x}) \tag{6.4}$$

性质　离散无记忆信源 X 的 N 次扩展信源 X^N 的熵等于离散信源 X 的熵的 N 倍，即

$$H(X^N)=N\cdot H(X) \tag{6.5}$$

证明：

$$\begin{aligned} H(X^N)&=\sum_{X^N}P(\boldsymbol{x})\log\frac{1}{P(\boldsymbol{x})} \\ &=\sum_{j=1}^{N}\left[\sum_{X^N}P(\boldsymbol{x})\cdot\log\frac{1}{P(x_j)}\right] \end{aligned}$$

其中，内和式为

$$\begin{aligned}\sum_{X^N} P(\boldsymbol{x}) \cdot \log \frac{1}{P(x_j)} &= \sum_{X^N} P(x_1)P(x_2)\cdots P(x_N)\log \frac{1}{P(x_j)} \\ &= \sum_{x_1 \in X}\sum_{x_2 \in X}\cdots \sum_{x_N \in X} P(x_1)P(x_2)\cdots P(x_N)\log \frac{1}{P(x_j)} \\ &= \sum_{x_1 \in X} P(x_1) \sum_{x_2 \in X} P(x_2)\cdots \sum_{x_{j-1} \in X} P(x_{j-1}) \sum_{x_j \in X} P(x_j)\log \frac{1}{P(x_j)} \\ &\quad \cdot \sum_{x_{j+1} \in X} P(x_{j+1})\cdots \sum_{x_N \in X} P(x_N)\log \frac{1}{P(x_j)}\end{aligned}$$

由于 $l \neq j$ 时：

$$\sum_{x_l \in X} P(x_l) = 1$$

则

$$\begin{aligned}\sum_{X^N} P(\boldsymbol{x}) \cdot \log \frac{1}{P(x_j)} &= \sum_{x_j \in X} P(x_j)\log \frac{1}{P(x_j)} \\ &= \sum_{l=1}^{r} P(a_l)\log \frac{1}{P(a_l)} = H(X)\end{aligned}$$

因此

$$H(X^N) = \sum_{j=1}^{N} H(X) = N \cdot H(X)$$

证毕。

【例 6.1】　有一离散无记忆信源$[X, P] = \begin{bmatrix} a_1 & a_2 \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix}$，求出其三次扩展信源 X^3 的概率空间，计算 $H(X^3)$并与 $H(X)$比较。

解：离散无记忆信源 X 的三次扩展信源 X^3 共包含 $2^3=8$ 种不同的样矢量。

因为信源 X 无记忆，所以有

$$P(\boldsymbol{x}) = P(x_1)P(x_2)P(x_3) \quad x_l \in X;\ l = 1, 2, 3$$

三次扩展信源 X^3 的样矢量和相应的概率分布为

X^3：	$\boldsymbol{x}_1$	$\boldsymbol{x}_2$	$\boldsymbol{x}_3$	$\boldsymbol{x}_4$	$\boldsymbol{x}_5$	$\boldsymbol{x}_6$	$\boldsymbol{x}_7$	$\boldsymbol{x}_8$
符号取值：	$a_1a_1a_1$	$a_1a_1a_2$	$a_1a_2a_1$	$a_1a_2a_2$	$a_2a_1a_1$	$a_2a_1a_2$	$a_2a_2a_1$	$a_2a_2a_2$
概率分布：	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{9}{64}$	$\frac{3}{64}$	$\frac{9}{64}$	$\frac{3}{64}$	$\frac{3}{64}$	$\frac{1}{64}$

由此可以写出三次扩展信源的概率空间为

$$[X^3, P] = \begin{bmatrix} a_1a_1a_1 & a_1a_1a_2 & a_1a_2a_1 & a_1a_2a_2 & a_2a_1a_1 & a_2a_1a_2 & a_2a_2a_1 & a_2a_2a_2 \\ \frac{27}{64} & \frac{9}{64} & \frac{9}{64} & \frac{3}{64} & \frac{9}{64} & \frac{3}{64} & \frac{3}{64} & \frac{1}{64} \end{bmatrix}$$

分别计算信源 X 及其三次扩展信源 X^3 的熵为

$$H(X) = H\left(\frac{3}{4}, \frac{1}{4}\right) = 0.811\ 25 \quad \text{比特 / 符号}$$

$$H(X^3) = H\left(\frac{27}{64}, \frac{9}{64}, \frac{9}{64}, \frac{3}{64}, \frac{9}{64}, \frac{3}{64}, \frac{3}{64}, \frac{1}{64}\right)$$
$$= 2.433\ 75 \quad \text{比特 / 符号}$$

因此

$$H(X^3) = 3 \cdot H(X)$$

以上结论也可以用熵的可加性解释。由于矢量的各分量间统计独立同分布，因此可以将扩展信源的熵看成是三个统计独立信源(信源熵均为 $H(X)$)的联合熵。由熵的可加性有：

$$H(X^3) = H(X) + H(X) + H(X) = 3H(X)$$

对于 X^N 而言，使用 $N-1$ 次熵的可加性即有：

$$H(X^N) = N \cdot H(X)$$

6.3 离散平稳信源

实际信源的输出往往是一个具有任意长度的随机变量序列。因此，对于实际信源输出的信息量及其关系的讨论是非常困难的。此处，我们限定随机矢量的长度为 N，通过分析 N 维随机矢量的统计关系，讨论信源输出信息的特性。

设信源连续输出的长度为 N 的符号序列可以表示成 N 维随机矢量：

$$\boldsymbol{x}_j = (x_{j+1}, x_{j+2}, \cdots, x_{j+N})$$

其中：

$$x_{j+i} \in X \quad X = \{a_1, a_2, \cdots, a_r\};\ i = 1, 2, \cdots, N(N \text{ 为任意正整数})$$

于是，N 维随机矢量的统计关系可以使用联合概率分布描述，即：

$$P(\boldsymbol{x}_j) = P(x_{j+1}, x_{j+2}, \cdots, x_{j+N})$$

由前几章的讨论可知，信源输出信息量的大小依赖于符号的统计特性。在实际信源发出的符号序列中，信源的统计特性不仅与符号序列以及符号之间的统计关系有关，而且一般情况下这种统计关系也是时间或位置的函数，即有：

$$P(\boldsymbol{x}_i) \neq P(\boldsymbol{x}_j)$$
$$P(x_i \mid x_{i-1}x_{i-2}\cdots x_{i-N}\cdots) \neq P(x_j \mid x_{j-1}x_{j-2}\cdots x_{j-N}\cdots) \qquad i \neq j$$

因此，一般信源输出的随机变量序列应当是一个统计关系与时间、位置有关的非平稳的随机过程。显然，对于这种输出非平稳随机变量序列的非平稳信源，其信息特性的分析是十分困难的。

如果信源符号序列的统计关系与其起点和位置的起点无关，则由随机过程理论可知，该符号序列具有平稳性。输出随机变量序列满足平稳性的信源为离散平稳信源。

设有离散平稳信源，其输出为 N 维随机矢量 $\boldsymbol{x}=(x_1, x_2, \cdots, x_N)$，由于 N 维随机矢量 x 的联合概率分布 $P(x)$与时间和位置的起点无关，即

$$P(\boldsymbol{x}_i) = P(x_{i+1}, x_{i+2}, \cdots, x_{i+N}) = P(x_{j+1}, x_{j+2}, \cdots, x_{j+N}) = P(\boldsymbol{x}_j) \quad i \neq j \qquad (6.6)$$

于是，N 维随机矢量 x 的联合概率分布 $P(x)$可以表示为

$$P(\boldsymbol{x}) = P(x_1, x_2, \cdots, x_N)$$

平稳信源输出的符号之间可能存在某种程度的关联，这种关联需要使用条件概率加以表示。依据联合概率与条件概率的关系：

$$P(\boldsymbol{x}_j) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1x_2)\cdots P(x_N \mid x_1x_2\cdots x_{N-1})$$

可知，平稳信源输出符号之间的条件概率也与时间、位置的起点无关，即满足：

$$\begin{cases} P(x_i) = P(x_j) = P(x) \\ P(x_{i+1} \mid x_i) = P(x_{j+1} \mid x_j) = P(x_2 \mid x_1) \\ P(x_{i+2} \mid x_ix_{i+1}) = P(x_{j+2} \mid x_jx_{j+1}) = P(x_3 \mid x_1x_2) \\ \vdots \\ P(x_{i+N-1} \mid x_ix_{i+1}\cdots x_{i+N-2}) = P(x_{j+N-1} \mid x_jx_{j+1}\cdots x_{j+N-2}) = P(x_N \mid x_1x_2\cdots x_{N-1}) \end{cases} \tag{6.7}$$

如果信源输出的某一符号只与该符号前面的 $N-1$ 个符号有统计依赖关系，则认为该信源符号序列的关联长度为 N。

由于离散平稳信源的熵及其特性的讨论仍然比较复杂，因此此处我们仅给出反映离散平稳信源信息特性的几个定义和重要结论。

设有离散平稳有记忆信源$[X, P]=\begin{bmatrix} a_1 & a_2 & \cdots & a_r \\ p_1 & p_2 & \cdots & p_r \end{bmatrix}$，其中：

$$0 \leqslant p_i \leqslant 1 \quad i = 1, 2, \cdots, r$$

$$\sum_{i=1}^{r} p_i = 1$$

该信源发出符号序列$\cdots X_1X_2\cdots X_NX_{N+1}\cdots$，设符号间关联长度为 N，则该平稳信源输出的长度为 N 的样矢量：

$$\boldsymbol{x} = x_1x_2\cdots x_N \qquad x_i \in X;\ i = 1, 2, \cdots, N$$

具有联合概率分布 $P(x_1, x_2, \cdots, x_N)$。

下面首先给出度量平稳信源输出信息量的三个定义。

定义 6.1　平稳信源输出的长度为 N 的随机矢量的联合熵为

$$H(X_1X_2\cdots X_N) = -\sum_{X_1X_2\cdots X_N} P(x_1x_2\cdots x_N)\log P(x_1x_2\cdots x_N) \tag{6.8}$$

定义 6.2　关联长度为 N 信源符号中平均每个符号所携带的信息量为平均符号熵，即

$$H_N(\boldsymbol{X}) = \frac{1}{N}H(X_1, X_2, \cdots, X_N) \tag{6.9}$$

定义 6.3　设信源符号序列之间的依赖关系长度为 N，若已知前面 $N-1$ 个符号，则第 N 个符号所携带的平均信息量为条件熵，即

$$H(X_N \mid X_1\cdots X_{N-2}X_{N-1}) = -\sum_{X_1\cdots X_N} P(x_1\cdots x_N)\log P(x_N \mid x_1\cdots x_{N-1}) \tag{6.10}$$

对于离散平稳信源，当 $H_1(X)=H(X)<\infty$ 时，具有以下性质：

性质 1　条件熵 $H(X_N|X_1X_2\cdots X_{N-2}X_{N-1})$ 随 N 的增加是非递增的，即

$$\begin{aligned} H(X_1) &\geqslant H(X_2 \mid X_1) \geqslant H(X_3 \mid X_2X_1) \geqslant \cdots \\ &\geqslant H(X_{N-1} \mid X_1X_2\cdots X_{N-2}) \geqslant H(X_N \mid X_1X_2\cdots X_{N-1}) \end{aligned} \tag{6.11}$$

证明：第 2 章中式(2.33)所给出的条件熵的性质 $H(X|Y)\leqslant H(X)$ 表明，对于任意的 X 和 Y，在已知 Y 的条件下关于 X 的平均不确定性不会超过关于 X 的先验不确定性，即条件熵不大于无条件熵。因此，当 $N=2$，即平稳信源输出二维随机矢量(X_1, X_2)时，有

$$H(X_2) \geqslant H(X_2 \mid X_1)$$

由于 X_1，$X_2 \in X$ **且具有相同的概率分布，则**

$$H(X_1) = H(X_2) = H(X)$$

因此

$$H(X_1) \geqslant H(X_2 \mid X_1)$$

当 $N \geqslant 3$，**即平稳信源输出** N **维随机矢量**$(X_1, X_2, \cdots, X_N)$**时，考察：**

$$\begin{aligned} &H(X_N \mid X_1X_2\cdots X_{N-1}) - H(X_{N-1} \mid X_1X_2\cdots X_{N-2}) \\ &= -\sum_{X_1\cdots X_{N-1}X_N} P(x_1x_2\cdots x_{N-1}x_N)\log P(x_N \mid x_1x_2\cdots x_{N-2}x_{N-1}) \\ &\quad + \sum_{X_1\cdots X_{N-2}X_{N-1}} P(x_1x_2\cdots x_{N-2}x_{N-1})\log P(x_{N-1} \mid x_1x_2\cdots x_{N-2}) \end{aligned}$$

由于离散平稳信源的联合概率分布和条件概率分布与时间或位置的起点无关，有：

$$\begin{aligned} &\sum_{X_1\cdots X_{N-2}X_{N-1}} P(x_1x_2\cdots x_{N-2}x_{N-1})\log P(x_{N-1} \mid x_1x_2\cdots x_{N-2}) \\ &= \sum_{X_2\cdots X_{N-1}X_N} P(x_2\cdots x_{N-1}x_N)\log P(x_N \mid x_2x_3\cdots x_{N-1}) \\ &= \sum_{X_1\cdots X_{N-1}X_N} P(x_1x_2\cdots x_{N-1}x_N)\log P(x_N \mid x_1x_2\cdots x_{N-2}) \end{aligned}$$

因此

$$\begin{aligned} &H(X_N \mid X_1X_2\cdots X_{N-1}) - H(X_{N-1} \mid X_1X_2\cdots X_{N-2}) \\ &= \sum_{X_1\cdots X_{N-1}X_N} P(x_1x_2\cdots x_{N-1}x_N)\log\left[\frac{P(x_N \mid x_2\cdots x_{N-1})}{P(x_N \mid x_1x_2\cdots x_{N-1})} \cdot \frac{P(x_1x_2\cdots x_{N-1})}{P(x_1x_2\cdots x_{N-1})}\right] \\ &= \sum_{X_1\cdots X_N} P(x_1x_2\cdots x_{N-1}x_N)\log\frac{P(x_N \mid x_2\cdots x_{N-1}) \cdot P(x_1x_2\cdots x_{N-1})}{P(x_1x_2\cdots x_{N-1}x_N)} \end{aligned}$$

已知对数函数是上凸函数，于是：

$$\begin{aligned} \text{**上式**} &\leqslant \log\Big(\sum_{X_1\cdots X_N} P(x_1x_2\cdots x_{N-1}x_N)\,\frac{P(x_N \mid x_2\cdots x_{N-1}) \cdot P(x_1x_2\cdots x_{N-1})}{P(x_1x_2\cdots x_{N-1}x_N)}\Big) \\ &= \log\Big(\sum_{X_1\cdots X_N} P(x_N \mid x_2\cdots x_{N-1})P(x_1x_2\cdots x_{N-1})\Big) \\ &= \log\Big(\sum_{X_1\cdots X_{N-1}} P(x_1x_2\cdots x_{N-1})\sum_{X_N} P(x_N \mid x_2\cdots x_{N-1})\Big) \\ &= \log\sum_{X_1\cdots X_{N-1}} P(x_1x_2\cdots x_{N-1}) \\ &= \log 1 = 0 \end{aligned}$$

因此有：

$$H(X_N \mid X_1X_2\cdots X_{N-1}) \leqslant H(X_{N-1} \mid X_1X_2\cdots X_{N-2})$$

证毕。

在观察平稳信源输出的符号序列时，如果已知 X_1、X_2、…、X_{N-1}**已发生，则关于将要发生的符号** X_N **的平均不确定性为条件熵** $H(X_N|X_1X_2\cdots X_{N-1})$**。性质 1 表明，当条件增加时，关于信源输出符号** X_N **的平均不确定性不会增大。**

性质 2

$$H_N(\boldsymbol{X}) \geqslant H(X_N \mid X_1X_2\cdots X_{N-1}) \tag{6.12}$$

证明：

由平均符号熵的定义知：

$$\begin{aligned} NH_N(\boldsymbol{X}) &= H(X_1X_2\cdots X_N) \\ &= -\sum_{X_1\cdots X_N} P(x_1\cdots x_N)\log P(x_1\cdots x_N) \end{aligned}$$

因为

$$P(x_1\cdots x_N) = P(x_1)P(x_2 \mid x_1)P(x_3 \mid x_1x_2)\cdots P(x_N \mid x_1x_2\cdots x_{N-1})$$

所以

$$\begin{aligned} NH_N(\boldsymbol{X}) = -\Big[&\sum_{X_1\cdots X_N} P(x_1x_2\cdots x_N)\log P(x_1) + \sum_{X_1\cdots X_N} P(x_1x_2\cdots x_N)\log P(x_2 \mid x_1) + \cdots \\ &+ \sum_{X_1\cdots X_N} P(x_1x_2\cdots x_N)\log P(x_N \mid x_1x_2\cdots x_{N-1})\Big] \end{aligned}$$

其中：

$$\begin{aligned} \sum_{X_1\cdots X_N} P(x_1x_2\cdots x_N)\log P(x_1) &= \sum_{X_1}\Big[\sum_{X_2\cdots X_N} P(x_1x_2\cdots x_N)\Big]\log P(x_1) \\ &= \sum_{X_1} P(x_1)\log P(x_1) = H(X_1) \\ \sum_{X_1\cdots X_N} P(x_1x_2\cdots x_N)\log P(x_2 \mid x_1) &= \sum_{X_1X_2}\Big[\sum_{X_3\cdots X_N} P(x_1x_2\cdots x_N)\Big]\log P(x_2 \mid x_1) \\ &= \sum_{X_1X_2} P(x_1x_2)\log P(x_2 \mid x_1) = H(X_2 \mid X_1) \end{aligned}$$

$$\sum_{X_1\cdots X_N} P(x_1x_2\cdots x_N)\log P(x_N \mid x_1x_2\cdots x_{N-1}) = H(X_N \mid X_1X_2\cdots X_{N-1})$$

于是有

$$H(X_1X_2\cdots X_N) = H(X_1) + H(X_2 \mid X_1) + \cdots + H(X_N \mid X_1X_2\cdots X_{N-1}) \tag{6.13}$$

性质 1 表明：

$$H(X_1) \geqslant H(X_2 \mid X_1) \geqslant \cdots \geqslant H(X_N \mid X_1X_2\cdots X_{N-1})$$

故必有

$$H_N(\boldsymbol{X}) = \frac{1}{N}H(X_1X_2\cdots X_N) \geqslant H(X_N \mid X_1X_2\cdots X_{N-1})$$

即

$$H_N(\boldsymbol{X}) \geqslant H(X_N \mid X_1X_2\cdots X_{N-1})$$

证毕。

由定义 6.2、6.3 可以看出，对于平稳信源输出 N 个信源符号 $X_1X_2\cdots X_{N-1}X_N$，平均符号熵 $H_N(\boldsymbol{X})$ 以简单的算术平均表示每一符号所载荷的平均信息量，而条件熵 $H(X_N|X_{N-1}\cdots X_1)$ 描述的是已知前面 $N-1$ 个符号时关于符号 X_N 的信息量。可见，平均符号熵 $H_N(\boldsymbol{X})$ 和条件熵 $H(X_N|X_1X_2\cdots X_{N-1})$ 均为单个符号平均信息量的度量方法。由于当已知条件的数量 N 增加时，符号 X_N 所携带的信息量将单调减小，因此，以 N 个符号的算术平均度量单个符号的信息量，一定不小于 $N-1$ 个符号已发生的条件下，关于第 N 个符号的信息量，即

$$H(X_N \mid X_1X_2\cdots X_{N-1}) \leqslant H_N(\boldsymbol{X})$$

性质 3

$$H_N(\boldsymbol{X}) \leqslant H_{N-1}(\boldsymbol{X}) \tag{6.14}$$

证明：

$$\begin{aligned} NH_N(\boldsymbol{X}) &= H(X_1X_2\cdots X_N) \\ &= H(X_N \mid X_1X_2\cdots X_{N-1}) + H(X_1X_2\cdots X_{N-1}) \\ &= H(X_N \mid X_1X_2\cdots X_{N-1}) + (N-1)H_{N-1}(\boldsymbol{X}) \\ &\leqslant H_N(\boldsymbol{X}) + (N-1)H_{N-1}(\boldsymbol{X}) \end{aligned}$$

于是：

$$\begin{aligned} (N-1)H_N(\boldsymbol{X}) &\leqslant (N-1)H_{N-1}(\boldsymbol{X}) \\ H_N(\boldsymbol{X}) &\leqslant H_{N-1}(\boldsymbol{X}) \end{aligned}$$

证毕。

性质 3 表明，平均符号熵 $H_N(\boldsymbol{X})$也是随着 N 的增加而非递增的。

进一步分析：

对于信息熵 $H(X)<\infty$的平稳信源 X，由于熵具有非负性，即 $H_N(\boldsymbol{X})\geqslant 0$，因此由性质 3 可以看出，信源符号序列长度不同时，平稳信源输出的每一个符号所载荷的信息量，即平均符号熵满足：

$$\infty > H(\boldsymbol{X}) = H_1(\boldsymbol{X}) \geqslant H_2(\boldsymbol{X}) \geqslant \cdots \geqslant H_{N-1}(\boldsymbol{X}) \geqslant H_N(\boldsymbol{X}) \geqslant 0 \tag{6.15}$$

可见，随着信源符号序列长度 N 的增加，$H_N(\boldsymbol{X})$将单调非递增地收敛于某有限值。

对于平稳信源 X，当其输出的符号序列足够长($N\to\infty$)时，每一个信源符号所载荷的信息有效地反映出了信源输出的实际信息量。

因此，定义 $H_\infty=\lim\limits_{N\to\infty}H_N(\boldsymbol{X})$为离散平稳信源 X 的极限信息量或极限熵。

性质 4 $H_\infty=\lim\limits_{N\to\infty}H_N(\boldsymbol{X})$**存在，且**

$$H_\infty = \lim_{N\to\infty}H_N(\boldsymbol{X}) = \lim_{N\to\infty}H(X_N \mid X_1X_2\cdots X_{N-1}) \tag{6.16}$$

证明：对于一个任意的正整数 k，有

$$\begin{aligned} H_{N+k}(\boldsymbol{X}) &= \frac{1}{N+k}H(X_1X_2\cdots X_NX_{N+1}\cdots X_{N+k}) \\ &= \frac{1}{N+k}[H(X_1X_2\cdots X_{N-1}) + H(X_N \mid X_1X_2\cdots X_{N-1}) \\ &\qquad + H(X_{N+1} \mid X_1X_2\cdots X_N) + \cdots + H(X_{N+k} \mid X_1X_2\cdots X_{N+k-1})] \\ &\leqslant \frac{1}{N+k}[H(X_1X_2\cdots X_{N-1}) + (k+1)H(X_N \mid X_1X_2\cdots X_{N-1})] \\ &= \frac{1}{N+k}H(X_1X_2\cdots X_{N-1}) + \frac{k+1}{N+k}H(X_N \mid X_1X_2\cdots X_{N-1}) \end{aligned}$$

固定 N，令 $k\to\infty$，得

$$\lim_{k\to\infty}H_{N+k}(\boldsymbol{X}) \leqslant 0 + H(X_N \mid X_1X_2\cdots X_{N-1})$$

已知 $N\to\infty$时，$H_N(\boldsymbol{X})$的极限存在且为 H_∞，故有

$$\lim_{k\to\infty}H_{N+k}(\boldsymbol{X}) = H_\infty$$

结合性质 2 指出的关系可知，条件熵 $H(X_N|X_1X_2\cdots X_{N-1})$满足：

$$H_\infty(\boldsymbol{X}) \leqslant H(X_N \mid X_1X_2\cdots X_{N-1}) \leqslant H_N(X)$$

令 $N\to\infty$，因为 $\lim\limits_{N\to\infty}H_N(\boldsymbol{X})=H_\infty$，所以

$$\lim_{N\to\infty} H(X_N \mid X_1X_2\cdots X_{N-1}) = H_\infty$$

于是：

$$H_\infty = \lim_{N\to\infty} H_N(\boldsymbol{X}) = \lim_{N\to\infty} H(X_N \mid X_1X_2\cdots X_{N-1})$$

证毕。

性质 2、3、4 说明：当平稳信源输出的符号序列长度达到无限大时，平均符号熵 $H_N(X)$和条件熵$H(X_N \mid X_1X_2\cdots X_{N-1})$都非递增地一致收敛于平稳信源的信息极限熵。因此，对于一个实际的平稳信源，当经过足够长的时间后，信源输出每一符号所载荷的信息量，即信源的极限熵可以用平均符号熵或条件熵加以度量。

6.4　马尔可夫信源

实际信源往往是有记忆的，其输出符号序列中相邻符号间存在某种程度的相关性。

例如，若将英文文字信源的取值集合看做 26 个英文字母、空格和标点符号，则英文信源输出的一段语句不仅是一个由字母、空格和标点符号组成的随机变量序列，并且由于该语句具有某种含义，英文文字信源输出的符号组合必须满足英文的词法、语法结构要求，随机变量序列中的符号相互依赖。同样，由于模拟信号的变化相对于采样速率很慢，因此离散图像信源和离散语音信源输出的数据序列中，数值的变化大多是缓慢的。在图像信源输出的灰度点阵中，相邻像元点的灰度相似，即像元点的灰度取值与其邻近像素点的灰度取值相近。语音信源输出的离散数据序列中，同一帧内相邻样点的幅度值差异小。

可见，有记忆信源输出的符号序列中，符号之间存在关联并且这种关联的程度与符号之间的距离有关。因此，有记忆信源的统计模型需要由一组信源符号和一组条件概率 $P(x_i \mid x_{i-1}x_{i-2}\cdots x_{i-L-1}x_{i-L}\cdots)$来描述。显然，相对于无记忆信源，有记忆信源信息特性的分析要复杂得多，特别是相关距离很大甚至是无限大时。

在实际信源输出的随机变量序列中，信源在某时刻的输出符号往往只与前面邻近的若干个信源符号有较强的依赖关系，而相距较远时，随机变量之间的相关性将变得较弱，甚至可以忽略不计。因此可以依实际信源的统计关系和应用要求，将有记忆信源的统计模型适当地简化。于是，通常限定有记忆信源的记忆长度，认为某时刻信源符号 X_i 发生的概率只与该时刻前面邻近的若干个信源输出符号相关联，而与更前面的符号是无关的。由于这类信源的输出符号序列近似为一个马尔可夫(Markov)过程，因此这类有记忆信源称为马尔可夫信源。

马尔可夫信源是一类特殊的离散有记忆信源，其特殊性在于任何时刻信源符号发生的概率只与前面已经发生的 m 个符号有关，而与更前面发生的符号无关。

对符号序列：

$$\cdots,\ x_{-1},\ x_0,\ x_1,\ \cdots,\ x_{m-1},\ x_m,\ x_{m+1},\ \cdots$$
$$x_i \in X,\ X = \{a_1,\ a_2,\ \cdots,\ a_r\}$$

有：

$$P(x_{t+m} \mid x_{t+m-1}\cdots x_{t+1}x_tx_{t-1}\cdots) = P(x_{t+m} \mid x_{t+m-1}\cdots x_{t+1}x_t) \tag{6.17}$$

因为这种信源在任何时刻符号发生的概率只与前 m 个符号有关，所以，可以把这前 m

个符号看做信源在此时刻所处的状态，它对应于 r^m 个不同的长为 m 的符号序列。

定义状态集：

$$E=\{E_1,\ E_2,\ \cdots,\ E_J\} \qquad J=r^m$$

若信源所处状态为 $s\in E$，在每一状态下可能输出的符号 $x\in X$，则当发出一个符号后，所处的状态就变了，即从状态 s 变到了另一状态，新状态由 s 的后 $m-1$ 个符号和发出的一个符号唯一确定。可见，状态的转移依赖于发出的信源符号，因此任何时刻信源处在什么状态完全由前一时刻的状态和发出的符号决定，对应信源的符号输出序列：

$$\cdots,\ x_1,\ x_2,\ \cdots,\ x_{l-1},\ x_l$$

有唯一的状态序列：

$$\cdots,\ s_1,\ s_2,\ \cdots,\ s_{l-1},\ s_l$$

对于 m 阶有记忆离散信源，它在 $t=1$ 时刻输出符号 x_1 的概率只与 $t=1$ 时刻前面 m 个符号所对应的状态 s_{l-1} 有关。在输出符号后，进入由最新的 m 个符号组成的状态 s_l，如图 6-2 所示。

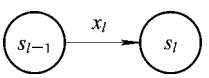

图 6-2 状态转移图

其状态转移概率分布为

$$P(x_l=a_k\mid s_{l-1}=E_i) \qquad k=1,\ 2,\ \cdots,\ r;\ i=1,\ 2,\ \cdots,\ r^m$$

一般情况下，状态转移概率及已知状态下发出符号的概率均与时刻 l 有关。若与 l 无关，即

$$P(x_l=a_k\mid s_l=E_i)=P(a_k\mid E_i) \tag{6.18}$$

则记 $p_{ij}=P(E_j\mid E_i)$，称此状态序列是时齐的。

定义 6.4 一个随机状态序列：$s_1,\ s_2,\ \cdots,\ s_l,\ \cdots$，若满足以下条件：

(1) 有限性：可能状态数目 $J<\infty$；

(2) 时齐性：状态转移概率 $p_{ij}=P(s_l=E_j\mid s_{l-1}=E_i)$ 只由状态 j 和状态 i 确定；

(3) 马氏性：进入某状态的概率只与前一时刻的状态有关，与更早的状态无关，即

$$P(s_l\mid s_{l-1},\ s_{l-2},\ \cdots)=P(s_l\mid s_{l-1})$$

则称此随机状态序列为有限状态齐次马尔可夫链。

定义 6.5 若信源输出的符号序列与信源的状态满足下述条件：

(1) 某一时刻信源的输出只与当前的信源状态有关，与以前的状态无关，即

$$\begin{aligned}P(a_k)&=P(x_l=a_k\mid s_l=E_j,\ x_{l-1}=a_h,\ s_{l-1}=E_i,\ \cdots)\\&=P(x_l=a_k\mid s_l=E_j)=P(a_k\mid E_j)\end{aligned}$$

(2) 信源状态只由当前输出符号和前一时刻信源状态唯一确定，即

$$\begin{cases}E_i=(a_{i1},\ a_{i2},\ \cdots,\ a_{im})\\E_j=(a_{j1},\ a_{j2},\ \cdots,\ a_{jm})\\P(s_l=E_j\mid x_l=a_k,\ s_{l-1}=E_i)=\begin{cases}1 & \text{当 } E_j=(a_{i2},\ \cdots,\ a_{im},\ a_k)\text{ 时}\\0 & \text{当 } E_j\neq(a_{i2},\ \cdots,\ a_{im},\ a_k)\text{ 时}\end{cases}\end{cases} \tag{6.19}$$

则称此信源为 m 阶马尔可夫信源。

m 阶马尔可夫信源的关联长度为 $m+1$。当 $m=1$，即符号发生的概率只与前面一个符号有关时，称为一阶马尔可夫信源。

m 阶马尔可夫信源的数学模型由一组信源符号和一组条件概率确定：

$$\begin{pmatrix} X \\ P \end{pmatrix} = \begin{bmatrix} a_1, a_2, \cdots, a_r \\ P(a_{k_{m+1}} \mid a_{k_1} a_{k_2} \cdots a_{k_m}) \end{bmatrix} \tag{6.20}$$

其中：

$$X = \{a_1, a_2, \cdots, a_r\}$$

$$k_1, k_2, \cdots, k_{m+1} \in \{1, 2, \cdots, r\}$$

满足：

$$\sum_{k_{m+1}=1}^{r} P(a_{k_{m+1}} \mid a_{k_1} a_{k_2} \cdots a_{k_m}) = 1$$

$$k_i = 1, 2, \cdots, r;\ i = 1, 2, \cdots, m+1$$

m 阶马尔可夫信源输出的随机变量序列 $X_1 X_2 \cdots X_{i-m+1} \cdots X_{i-2} X_{i-1} X_i X_{i+1} \cdots$ 为一个 m 阶马尔可夫过程。可以将 m 个相邻的随机变量划分为一组，将该 m 个随机变量的取值组合：

$$a_{i_1} a_{i_2} \cdots a_{i_m} \qquad i_1, i_2, \cdots, i_m = 1, 2, \cdots, r \tag{6.21}$$

看做一种状态 E。

已知信源 X 有 r 种不同的输出符号，故信源 X 输出的长度为 m 的随机变量序列有 r^m 种不同的取值组合。于是，长度为 m 的随机变量序列可以表示为 r^m 种不同的状态，并构成状态集合 $E=\{E_1, E_2, \cdots, E_{r^m}\}$。此时，马尔可夫信源的数学模型又可表示为

$$\begin{pmatrix} E \\ P \end{pmatrix} = \begin{pmatrix} E_1, E_2, \cdots, E_{r^m} \\ P(E_j \mid E_i) \end{pmatrix}$$

如果 m 阶马尔可夫信源 X 在时刻 i 的状态为 $s_i = a_{i_1} a_{i_2} \cdots a_{i_m} \in E$，则由于某时刻信源的输出仅与该时刻前面的 m 个符号有关，因此信源在时刻 i 输出的符号与信源所处的状态 s_i 有关，即依条件概率 $P(a_{i_{m+1}} \mid s_i)$ 输出的符号 $a_{i_{m+1}} \in X$，长度为 m 的符号序列便变为 $a_{i_2} a_{i_3} \cdots a_{i_m} a_{i_{m+1}}$。由马尔可夫信源的统计关系可知，新的长度为 m 的符号序列由原状态 s_i 对应的符号序列 $a_{i_1} a_{i_2} \cdots a_{i_m}$ 中的后 $m-1$ 个信源符号与该时刻信源发出的新符号 $a_{i_{m+1}}$ 组成，并构成了一个新的状态 $s_{i+1} = a_{i_2} a_{i_3} \cdots a_{i_m} a_{i_{m+1}} \in E$，即由状态 s_i 转移为状态 s_{i+1}（s_i，$s_{i+1} \in E$），并且依输出符号之间的条件概率关系可知，状态转移的概率为

$$P(a_{i_{m+1}} \mid s_i = a_{i_1} a_{i_2} \cdots a_{i_m}) = P(s_{i+1} \mid s_i)$$

因此，m 阶马尔可夫信源的输出符号序列 $X_1 X_2 \cdots X_{i-m+1} \cdots X_{i-2} X_{i-1} X_i X_{i+1} \cdots$ 可以转换为一个状态序列 $s_1 s_2 \cdots s_i s_{i+1} \cdots$。

这样的状态序列可以用马尔可夫链的状态图描述（见图 6-3），也可以由状态转移矩阵描述。

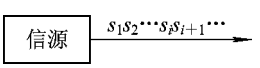

图 6-3　马尔可夫信源状态图

已知条件概率可变换成：

$$P(a_{k_{m+1}} \mid a_{k_1} a_{k_2} \cdots a_{k_m}) = P(a_{k_{m+1}} \mid E_i) = P(a_k \mid E_i)$$

其中：E_i 为 m 阶马尔可夫信源可能有的状态，$i=1, \cdots, r^m$；$P(a_k \mid E_i)$ 为任何时刻信源处在

状态 E_i 时发出信源符号 a_k 的概率，a_k 可任取 a_1，a_2，…，a_r 之一。若新状态为 E_j，则状态 E_i 到 E_j 的一步转移概率为

$$P(E_j \mid E_i) = P(a_k \mid E_i)$$

马尔可夫信源的状态转移图如图 6-4 所示。

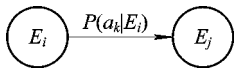

图 6-4 马尔可夫信源的状态转移图

此时，马尔可夫信源的数学模型又可表示为

$$\begin{pmatrix} E \\ P \end{pmatrix} = \begin{bmatrix} E_1, E_2, \cdots, E_{r^m} \\ P(E_j \mid E_i) \end{bmatrix} \quad i = 1, 2, \cdots, r^m \tag{6.22}$$

该数学模型表明，状态一步转移概率是描述 m 阶马尔可夫信源有依赖地发出消息(状态)的统计特性的最本质参数。

一般地，马尔可夫信源输出的状态序列中，状态之间的转移关系与系统的初始状态有关。但是，由于实际马尔可夫信源输出的状态种类 r^m 有限，并且近似满足时齐性和遍历性条件，因此我们主要考虑有限、时齐、遍历马尔可夫信源的统计关系。

对于状态集合为 $E=\{E_1, E_2, \cdots, E_{r^m}\}$ 的时齐马尔可夫信源，由该信源的 m 个符号组成的状态 E_i 和 $E_j(i \neq j)$，存在一个正整数 $n_0>0$，且经过 n_0 步，从状态 E_i 转移到状态 E_j 的转移概率 $p_{ij}^{(n)}>0$，则在经过足够长的时间后，信源可以由一种状态转移到任意一种其他状态，并且满足：

(1) 新状态 s_{i+1} 的概率仅与前一时刻的状态有关，而与更前面的状态无关，即

$$P(s_{i+1} \mid s_i s_{i-1} s_{i-2} \cdots) = P(s_{i+1} \mid s_i)$$

(2) 状态之间的转移概率 $P(E_l \mid E_k)$ 与状态发生转移的时间或位置无关。

(3) 无论信源的初始状态 E_i 为何，各种状态发生的概率都稳定在某极限概率：

$$\lim_{n \to \infty} p_{ij}^{(n)} = P(E_j) > 0$$

即每一种由各态历经过程产生的符号序列具有同样的统计特性。

可见，在经过足够长的时间后，有限、时齐、遍历马尔可夫信源成为一种平稳信源，且存在不依赖初始状态的状态极限概率(即平稳分布)。

需要强调的是，只有在转移一定步数后各状态之间均可相通的条件下，当转移步数足够大，m 阶马尔可夫信源达到稳定时，各状态出现的概率才能稳定在某一极限值，存在极限概率。

对于 m 阶马尔可夫信源，状态一步转移概率可由状态一步转移矩阵表示，也可以用状态一步转移图表示。对于各态遍历的马尔可夫信源，相应的状态一步转移图具有以下两个特点：

(1) 不可约性。在状态空间集合中，不存在一个各状态之间都能相互到达，而又与集合以外的状态不相通的集合，即闭集中不能再分出闭集。

(2) 非周期性。对于各态遍历的 m 阶马尔可夫信源来说，存在一个正整数 $n_0>0$，经过 n_0 步转移后，各态均能相通，各态均能回到出发状态。在回到出发状态的可能步数中，必定包括 n_0 和 n_0+1 或 n_0-1。所以，在可能回到出发状态的步数中，不存在大于 1 的公因

子，具有非周期性。

【例 6.2】 有一个二进制二阶马尔可夫信源，$X=\{0,1\}$，条件概率为

$$P(0\mid 00)=P(1\mid 11)=0.8$$
$$P(1\mid 00)=P(0\mid 11)=0.2$$
$$P(0\mid 01)=P(0\mid 10)=P(1\mid 01)=P(1\mid 10)=0.5$$

该信源的转移特性只与符号序列有关，有 4 个状态：

E_1　　对应序列 00
E_2　　对应序列 01
E_3　　对应序列 10
E_4　　对应序列 11

由概率分布可知，状态一步转移概率为

$$P(E_1\mid E_1)=P(E_4\mid E_4)=0.8$$
$$P(E_2\mid E_1)=P(E_3\mid E_4)=0.2$$
$$P(E_3\mid E_2)=P(E_2\mid E_3)=P(E_4\mid E_2)=P(E_1\mid E_3)=0.5$$

其他状态转移概率等于 0。

信源是有限、既约、齐次的马尔可夫链，且可以进一步判定是各态历经源，故有：

$$\begin{cases}P(E_1)=0.8P(E_1)+0.5P(E_3)\\P(E_2)=0.2P(E_1)+0.5P(E_3)\\P(E_3)=0.5P(E_2)+0.2P(E_4)\\P(E_4)=0.8P(E_4)+0.5P(E_2)\\P(E_1)+P(E_2)+P(E_3)+P(E_4)=1\end{cases}$$

可求出：

$$P(E_2)=P(E_3)=\frac{1}{7},\ P(E_1)=P(E_4)=\frac{5}{14}$$

为马尔可夫信源的平稳分布，且等于初始分布。

由上述讨论可以看出，由于有记忆信源输出符号之间存在关联，因此其信息特性分析的难度较大，特别是相关距离很大甚至是无限大时。工程应用中，图像、语音等实际信源输出的符号序列中，相邻符号之间的相关性较大，而距离较远的符号之间的相关性很小，往往可以忽略。因此，对于这类实际信源信息特性的分析，可以从两个方面作适当假设，以在掌握此类信源本质特性的前提下简化分析过程。

首先，由于实际信源的相关性主要表现在相邻符号之间，因此可以对有记忆信源的记忆长度加以限制。假设信源在某时刻输出的符号仅与该时刻前面输出的 m 个符号存在依赖关系，而与更前面的符号无关，则将信源的相关长度限定为 m，看做 m 阶的马尔可夫信源。

其次，假设实际信源输出的随机变量序列满足时齐性和遍历性，那么，在经过足够长的时间后，时齐、遍历的 m 阶马尔可夫信源输出的符号序列为一平稳随机过程。

因此，可以将工程应用中的实际信源看做一个平稳的 m 阶马尔可夫信源，依据平稳信源和马尔可夫信源的信息测度所满足的基本关系和性质，分析实际信源的信息特性。

假设时齐、遍历的 m 阶马尔可夫信源的输出是一个具有任意长度的符号序列：

$$X_1X_2\cdots X_{i-1}X_i\cdots$$

由 6.3 节的讨论可知，当其输出的符号序列足够长($N\to\infty$)时，平均符号熵 $H_N(\boldsymbol{X})$的极限值反映了信源输出的实际信息量。已知此信源的极限信息量或极限熵为

$$H_\infty = \lim_{N\to\infty} H_N(\boldsymbol{X}) = \lim_{N\to\infty} H_N(X_1X_2\cdots X_{N-1}X_N)$$

因为此处的 $N\to\infty$表示信源输出的符号序列足够长，这表明信源的输出已经经历了足够长的时间，所以信源符号序列可以看做一个平稳的随机过程，该信源成为一种平稳信源。对于平稳信源，极限熵 H_∞也可以表示为序列长度 N 足够长时的条件熵，即有：

$$H_\infty = \lim_{N\to\infty} H(X_N \mid X_1X_2\cdots X_{N-1}) \tag{6.23}$$

由于此信源是一个 m 阶的马尔可夫信源，输出符号 X_N 仅与该符号前面的 m 个符号相关，而与更前面的符号无依赖关系，因此，式(6.23)反映的无限长的条件依赖关系可以转化为有限长的依赖关系，满足：

$$H_\infty = \lim_{N\to\infty} H(X_N \mid X_{N-m}X_{N-m+1}\cdots X_{N-2}X_{N-1})$$

进一步地，考虑到信源已进入一个平稳状态，而平稳信源输出随机变量序列的统计特性与时间或位置的起点无关，故有：

$$H_\infty = H(X_{m+1} \mid X_1X_2\cdots X_m)$$

于是，得到 m 阶马尔可夫信源的极限信息量：

$$H_\infty = H(X_{m+1} \mid X_1X_2\cdots X_m) = H_{m+1}$$

可见，时齐、遍历的 m 阶马尔可夫信源所提供的平均信息量为已知前 m 个信源符号的条件下，对信源将要输出的符号的不确定性。因此，m 阶马尔可夫信源的极限熵的计算可以转化为条件熵的计算，即

$$H_\infty = \lim_{N\to\infty} H(X_N \mid X_1\cdots X_{N-1}) = H(X_{m+1} \mid X_1\cdots X_m) \tag{6.24}$$

而

$$H_{m+1} = -\sum_{k_{m+1}}\cdots\sum_{k_1} P(a_{k_1}\cdots a_{k_{m+1}})\cdot\log P(a_{k_{m+1}} \mid a_{k_1}\cdots a_{k_m})$$

令

$$E_i = (a_{k_1}a_{k_2}\cdots a_{k_m})$$

其中：

$$i = 1, 2, \cdots, r^m;\ k_j = 1, 2, \cdots, r;\ j = 1, 2, \cdots, m$$

则

$$P(a_{k_1}\cdots a_{k_{m+1}}) = P(E_i)P(a_{k_{m+1}} \mid E_i)$$

因此

$$\begin{aligned} H_{m+1} &= -\sum_{E_i\in E}\sum_{k} P(E_i)P(a_k \mid E_i)\cdot\log P(a_k \mid E_i) \\ &= \sum_{E} P(E_i)H(X \mid E_i) \end{aligned} \tag{6.25}$$

上例中：

$$\begin{aligned} H_\infty &= H_{m+1} = H_3 \\ &= \sum_E P(E_i)H(X \mid E_i) \\ &= P(E_1)H(X \mid E_1) + P(E_2)H(X \mid E_2) + P(E_3)H(X \mid E_3) + P(E_4)H(X \mid E_4) \end{aligned}$$

则

$$H(X \mid E_1) = H(0.8, 0.2) = 0.7219 \text{ 比特 / 符号}$$
$$H(X \mid E_4) = H(0.8, 0.2) = 0.7219 \text{ 比特 / 符号}$$
$$H(X \mid E_2) = H(0.5, 0.5) = 1.0 \text{ 比特 / 符号}$$
$$H(X \mid E_3) = H(0.5, 0.5) = 1.0 \text{ 比特 / 符号}$$
$$H_\infty = \frac{5}{7} \times 0.7219 + \frac{2}{7} \times 1.0 \approx 0.80 \text{ 比特 / 符号}$$

我们在 6.3 节讨论 N 维离散平稳信源 $\boldsymbol{X} = X_1X_2\cdots X_N$ 时，把一个实际上关联长度为无穷的信源近似地当做一个关联长度有限的信源来考虑，即每 N 个符号组成的符号序列之间看做统计独立，互不相关，只在由 N 个符号组成的符号序列内考虑符号的统计依赖关系，然后用 N 维联合熵 $H(\boldsymbol{X}) = H(X_1X_2\cdots X_N)$ 表示信源每个符号序列的平均信息量，用平均符号熵 $H_N = \frac{1}{N}H(X_1X_2\cdots X_N)$ 表示离散平稳信源有记忆信源中每一个符号提供的平均信息量。然而，对于一般的有记忆信源，考虑到关联长度为无穷这一因素，计算 $N \to \infty$ 时的极限值 H_∞ 非常困难，实际上几乎不可操作。但是由于各态遍历的 m 阶马尔可夫信源的记忆长度有限，再加上它具有独特的马氏特性，所以各态遍历的 m 阶马尔可夫信源的极限熵是可求的。因此，对于离散平稳有记忆信源的极限熵计算来说，各态遍历的 m 阶马尔可夫信源的极限熵更加贴近实际，具有广泛的应用性。

6.5　信源的信息冗余

前面我们已经讨论了各类离散信源，包括离散无记忆信源及其 N 次扩展信源、离散平稳信源、马尔可夫信源。由于这些信源的统计特性各不相同，它们输出信息的能力也各不相同。

对于有 r 种输出符号的离散信源 X，如果信源 X 的输出符号之间无记忆且等概分布，则由最大熵定理可知，其输出的平均信息量达到其最大值，即该信源的熵为

$$H_0 = \log r$$

如果信源 X 输出的符号之间无记忆，但是信源符号为非等概率分布，则信源 X 输出的平均信息量为

$$H_1 = H(X)$$

如果信源 X 是有记忆的，其输出的符号之间存在某种程度的相关性，则其输出的信息量将进一步减小。对于 m 阶的马尔可夫信源：

当 $m=1$ 时，信源的熵为

$$H_2 = H(X_2 \mid X_1)$$

当 $m=2$ 时，信源的熵为

$$H_3 = H(X_3 \mid X_1X_2)$$

$$\vdots$$

为 m 时，信源的熵为

$$H_{m+1} = H(X_{m+1} \mid X_1X_2\cdots X_{m-1}X_m)$$

当信源满足平稳性要求，即为平稳信源时，输出的平均信息量为极限熵 H_∞。

条件熵的性质：

$$H(X_1) \geqslant H(X_2 \mid X_1) \geqslant H(X_3 \mid X_2X_1) \geqslant \cdots$$
$$\geqslant H(X_{N-1} \mid X_1X_2\cdots X_{N-2}) \geqslant H(X_N \mid X_1X_2\cdots X_{N-1})$$

表明，上述具有不同统计特性的信源输出的信息量满足：

$$\log r = H_0(X) \geqslant H_1(X) \geqslant H_2(X) \geqslant \cdots \geqslant H_{m+1}(X) \geqslant \cdots \geqslant H_\infty(X) \quad (6.26)$$

由此可以看出，对于一个有 r 种不同输出符号的离散平稳信源 X，按照最大熵定理该信源输出信息的能力应当是 $\log r$。然而式(6.26)表明，若信源 X 输出符号为非等概率分布，则信源输出的平均信息量将减少；若信源输出符号间有关联，则信源输出的平均信息量将进一步减少，且符号间相关长度越长，信源输出的信息量就越小。因此，非等概率分布的无记忆信源和有记忆信源所输出的信源符号中，每一位信源符号所载荷的平均信息量并没有达到其应当具有的最大输出信息能力，这表明信源输出符号中含有一定程度的不含有信息的多余部分。

为了衡量信源输出的符号和符号序列中不含有信息的多余部分的大小，引入了信源的信息冗余度(亦称剩余度、多余度)。

定义 6.6 信源输出的实际信息熵与具有同样符号集的最大熵之比称为相对熵 η，即

$$\eta = \frac{H_\infty}{H_0} \quad (6.27)$$

定义 6.7 信源的信息冗余度 γ：

$$\gamma = 1 - \eta \quad (6.28)$$

信源的信息冗余度 γ 度量了信源输出的符号不含有信息的多余部分的大小。对于有记忆信源，由于这部分多余成分的大小取决于信源符号之间的相关性，因此，信源的信息冗余度本质上是反映信源统计特性的一个物理量。若信源符号之间的依赖关系较强，即符号间的相关距离长，则 H_∞ 较小，于是该信源的信息冗余度较大。反之，若信源输出符号的相关性较弱，则该信源的信息冗余度较小。若信源符号相互独立且等概率分布，则信源输出的平均信息量达到该类信源输出平均信息量的最大值 H_0，信源输出的符号中不包含任何多余成分，于是信息冗余度 $\gamma=0$。

在各类信息传输系统中，图像、语音、文字等信源输出的符号序列中的相邻符号之间往往存在明显的关联。例如一幅图像中，由于同一物体具有相近的电磁波反射特性，因此某像素点的灰度取值与其相邻像素点的灰度取值接近。

【例 6.3】 考察英文信源输出的符号序列，计算其信息冗余度 γ。

设英文信源的输出符号为 26 个英文字母和空格，故此英文信源的输出符号集合中包括了 27 个符号。如果认为此英文信源是一个无记忆且等概率分布，则该信源的最大熵为

$$H_0 = \log 27 = 4.76 \quad \text{比特 / 符号}$$

如果认为英文信源是无记忆的(即不考虑相邻字母、符号之间的相关性)，则在实际的

英文单词拼写和语句组成中，各符号并非等概率发生。经过对英文信源符号的统计计算，可以得到空格和 26 个英文字母出现的概率(如表 6－1 所示)。于是，若假定英文信源是一个非等概率分布的无记忆信源，则可以求出信源的熵：

$$H_1 = \sum_{i=1}^{27} p_i \log p_i = 4.03 \quad \text{比特 / 符号}$$

表 6－1　英文信源符号发生的概率

字　母	概　率	字　母	概　率
空格	0.1859	N	0.0574
A	0.0642	O	0.0632
B	0.0127	P	0.0152
C	0.0218	Q	0.0008
D	0.0317	R	0.0484
E	0.1031	S	0.0514
F	0.0208	T	0.0796
G	0.0152	U	0.0228
H	0.0467	V	0.0083
I	0.0575	W	0.0175
J	0.0008	X	0.0013
K	0.0049	Y	0.0164
L	0.0321	Z	0.0005
M	0.0198		

显然，在有意义的实际英文单词和语句中，英文信源输出的符号之间存在着明显的相关性。因此，由 26 个英文字母和空格组成的英文信源是一个有记忆信源，在英文信源输出信息特性的分析中，我们可以将其近似为 m 阶马尔可夫信源。考虑的相关距离 m 不同，该英文信源输出的平均信息量不同。

若取 $m=1$，则 $H_2=3.32$ 比特/符号。

若取 $m=2$，则 $H_3=3.1$ 比特/符号。

……

一般认为，实际英文信源的熵为

$$H_\infty = 1.4 \quad \text{比特 / 符号}$$

于是，可以求出实际英文信源的相对熵为

$$\eta = \frac{H_\infty}{H_0} = 0.29$$

信息剩余度为

$$\gamma = 1 - 0.29 = 0.71$$

计算结果表明，由英文字母和空格组成的文章中，其中 71％的符号取值由英文的词

法、句法和语言结构所决定，只有29%的符号由写文字的人自由选择。显然，从信息传输系统的有效性考虑，这些不包含信息的多余部分应当予以消除。因此，对于一本100页的英文图书，如果将大约71%的信息冗余消除掉，则可以只存储或传输大约29页篇幅的内容，便可以在保持原图书载荷信息量的前提下，大大提高信息系统的有效性。可见，信源的信息冗余度表示了有记忆信源可以压缩的程度。

由于文字、图像、语音等各类实际信源往往都存在较强的相关性，因此实际信源输出的符号序列中存在较大的信息冗余。在信息传输系统设计中，需要通过某种信源数据的压缩编码，减少或去除冗余，便可以在保持信源信息的前提下传输或存储较少的符号，大大提高信息系统的有效性。因此，以提高系统有效性为目的的信源编码，成为当前信息技术中的关键技术。

但是应当看到，信源的信息冗余度在信息的传输和存储中也有其有利的一面。从信息传输和存储的可靠性来看，信源输出的数据之间所存在的相关性可以提高系统的抗干扰能力。例如，由于存在信道干扰，可能会使得在接收端接收到的符号产生错误。利用信源输出符号之间的相关性，可以将传输过程中产生的错误加以纠正，提高信息传输系统的可靠性。

例如，将“信息工程学院”简称为“工院”，显然该表示方法简捷，文字传输效率高。但是，当在传输过程中，字符“工”受到噪声的影响变得模糊不清、难以辨认时，由“X院”难以判断信源所输出的正确符号。然而，如果传输的是“信息工程学院”，当字符“工”受到噪声的影响变得模糊不清，成为“信息X程学院”时，则可利用符号之间的关联，在接收端正确恢复字符“工”，实现信源字符的可靠传输。

由此可知，信源符号序列的相关性愈强(信源信息冗余度愈大)，则抗干扰的能力愈强，信息系统的可靠性就愈高。但是，从信息传输有效性角度来看，减小信源符号之间的相关性，可以消除或去除信息冗余，提高系统的有效性。针对有效性和可靠性两种不同的技术目标，形成了两类不同的编码技术，即信源编码和信道编码。

信源编码以提高信息系统的有效性为目的，通过消除信源的信息冗余减小每一信源符号的平均比特数，从而实现信源输出信息的有效表示和传输。信道编码则以提高信息传输的可靠性为目的，在充分利用信道信息传输能力的条件下，通过加入特定形式的冗余数据，使得编码序列具有发现和纠正传输误码的能力，从而实现信源信息的可靠传输。

因此，提高系统的抗干扰能力往往是以降低系统的信息传输效率为代价的；反之，提高系统的传输效率又常常使抗干扰能力减弱。可见，在信息系统的研究和设计中，有效性和可靠性是一对相互矛盾的技术指标。信息理论和编码技术的研究将从信源、信道的统计关系出发，研究提高信息系统有效性和可靠性的基本理论和方法，使有效性和可靠性两者达到辩证的统一，从而实现信息系统的最优化。

习 题 6

6.1 已知离散无记忆信源$[X, P]=\begin{bmatrix}0 & 1\\ \frac{3}{4} & \frac{1}{4}\end{bmatrix}$。

(1) 试求 X 的二次扩展信源$[X^2, P]$和 $H(X^2)$；

(2) 试求 X 的三次扩展信源$[X^3, P]$和 $H(X^3)$；

(3) 验证 $H(X^2)=2H(X)$，$H(X^3)=3H(X)$。

6.2　已知离散无记忆信源的概率空间$[X, P]=\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \\ 0.7 & 0.05 & 0.1 & 0.15 \end{bmatrix}$，计算此信源的相对熵和信源冗余度。

6.3　某平稳信源$[X, P]=\begin{bmatrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$，设它发出的符号只与前一个有关，已知联合概率分布如表 6－2 所示。

(1) 若认为信源符号间无依赖性，计算 $H(X)$。

(2) 考虑题示符号间的相关性，计算 $H(X_2|X_1)$、$H(X_1X_2)$和平均符号熵 $H_2(\boldsymbol{X})$。

(3) 验证 $H(X_2|X_1)<H(X)$、$H(X_1X_2)<2H(X)$和 $H_2(X)<H(X)$，并解释 $H_2(\boldsymbol{X})<H(X)$的物理含义。

表 6－2　题 6.3 表

$P(x_ix_j)$	0	1	2
0	$\frac{3}{16}$	$\frac{1}{16}$	0
1	$\frac{1}{16}$	$\frac{3}{8}$	$\frac{1}{16}$
2	0	$\frac{1}{16}$	$\frac{3}{16}$

6.4　已知平稳信源$[X, P]=\begin{bmatrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{bmatrix}$。

(1) 若认为信源是无记忆的，此信源的冗余度如何？

(2) 若认为此信源实际的熵近似为 $H_2(\boldsymbol{X})$，即 $H_x\approx H_2(\boldsymbol{X})$，且已知两个符号的联合概率分布如表 6－2 所示，计算此时的冗余度。

(3) 由(1)和(2)的计算结果可得出什么结论？

6.5　证明平稳信源有 $H(X_3|X_1X_2)\leqslant H(X_2|X_1)$，并说明等式成立的条件。

6.6　设有一个信源，它产生 0，1 序列的消息。它在任意时间而且不论以前发生过什么符号，均按 $P(0)=0.4$，$P(1)=0.6$ 的概率发出符号。

(1) 试问这个信源是否平稳？

(2) 试计算 $H(X^2)$、$H(X_3|X_1X_2)$及 $\lim\limits_{N\to\infty}H_N(\boldsymbol{X})$。

(3) 试计算 $H(X^4)$并写出 X^4 信源中可能有的所有符号。

6.7　有一信源，它在开始时以 $P(a)=0.6$，$P(b)=0.3$，$P(c)=0.1$ 的概率发出 X_1。当 X_1 为 a 时，X_2 为 a、b、c 的概率各为 1/3；X_1 为 b 时，X_2 为 a、b、c 的概率各为 1/3；

X_1 为 c 时，X_2 为 a、b 的概率各为 $1/2$，为 c 的概率为 0。后面发出 X_i 的概率只与 X_{i-1} 有关，$P(X_i|X_{i-1})=P(X_2|X_1)$，$i\geqslant 3$。利用马尔可夫信源的图示法画出状态转移图，并计算信源熵 H_∞。

6.8 一阶马尔可夫信源的状态图如图 6-5 所示，信源 X 的符号集为{0，1，2}，并定义 $\overline{p}=1-p$。

(1) 求信源平稳后的概率分布 $P(0)$、$P(1)$和 $P(2)$。

(2) 求此信源的熵。

(3) 近似认为此信源为无记忆时，符号的概率分布等于平稳分布。求近似信源的熵 $H(X)$，并与 H_∞ 进行比较。

(4) 对一阶马尔可夫信源，p 取何值时 H_∞ 取最大值？当 $p=0$ 和 $p=1$ 时，结果如何？

6.9 一阶马尔可夫信源的状态图如图 6-6 所示，信源 X 的符号集为{0，1，2}。

(1) 求平稳后信源的概率分布。

(2) 求信源的熵 H_∞。

(3) 求当 $p=0$ 和 $p=1$ 时信源的熵，并说明其理由。

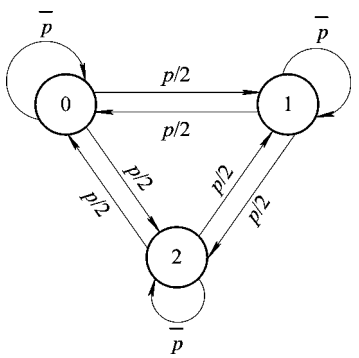

图 6-5 题 6.8 图

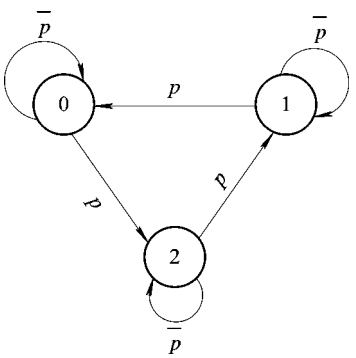

图 6-6 题 6.9 图

6.10 黑白气象传真图的消息只有黑色和白色两种，即信源 $X=${黑，白}，设黑色出现的概率 $P(\text{黑})=0.3$，白色出现的概率 $P(\text{白})=0.7$。

(1) 假设图上黑白消息出现前后没有关联，求熵 $H(X)$。

(2) 假设黑白消息前后有关联，其依赖关系为 $P(\text{白}|\text{白})=0.9$，$P(\text{黑}|\text{白})=0.1$，$P(\text{白}|\text{黑})=0.2$，$P(\text{黑}|\text{黑})=0.8$，求此一阶马尔可夫信源的熵 H_2。

(3) 分别求出上述两种信源的剩余度，并比较 $H(X)$和 H_2 的大小，说明其物理意义。

第 7 章　无失真信源编码

由第 3 章的讨论我们知道，信源输出的消息中可能存在不含有信息的多余成分。在各类信息传输和存储系统的设计中，应当消除由这些多余成分所产生的信息冗余，提高系统的有效性。因此，作为提高系统有效性的手段，高效的信源编码是各类信息传输和存储系统开发研制中的一项关键技术。

这一章我们将针对离散无记忆信源的统计特性，从无失真信源编码的要求出发，讨论无失真信源编码的原理、理论极限和基本方法。

7.1　信源编码的作用与构成

在一个信息传输系统中，信源以输出符号的形式输出信息。例如，离散无记忆信源的符号集 $U=\{u_1, u_2, \cdots, u_n\}$ 输出的符号或消息可以表示字母、数字、文字等不同的内容。

然而，某些信息系统对于传输、存储和处理的信息载体的形式有特定的要求。例如，电报通信系统传输的是电报码，计算机中可以运行的指令和可以被计算、处理的数据必须是二进制符号。为了使这些载荷信源信息的各种符号、消息能够通过某种信息系统传输、存储和处理，信源输出的符号、消息必须转换为信息系统所要求的信号或码字形式。

因此，从一般意义上讲，这种将各种信源输出的符号、消息转换成特定信息系统所要求的、易于处理的表示形式的过程称为信源编码。

设信源符号集为

$$U = \{u_1, u_2, \cdots, u_n\}$$

根据信息系统的要求，我们可以设计码符号集：

$$X = \{x_1, x_2, \cdots, x_i\}$$

其中，$x_i(i=1, 2, \cdots, m)$ 适合系统的要求，称为码符号或码元。

由若干位码元(如 K 位)可以组成一个码符号序列：

$$W_i = X_1 X_2 \cdots X_K$$

其中，$X_j \in X(j=1, 2, \cdots, K)$。这样的码符号序列 W_i 亦称做码字。由所有码字构成的集合称为码本：

$$C = \{W_1, W_2, \cdots, W_n\}$$

信源编码实际上是信源符号与码字之间按照一定规则进行的一种变换处理，即将信源符号集中的符号 $u_i(i=1, 2, \cdots, n)$ 变换成长度为 K 的码符号序列(码字)：

$$W_i = X_1 X_2 \cdots X_K$$

或者将信源输出的长为 L 的信源符号序列：

$$\boldsymbol{u}_i = U_1U_2\cdots U_L \quad i = 1, 2, \cdots, n^L$$
$$U_j \in U \qquad j = 1, 2, \cdots, L$$

与码本：

$$C = \{W_1, W_2, \cdots, W_{n^L}\}$$

中的码字 W_i 构成某种对应关系。

经过这样的变换处理，信源符号 u_i、信源符号序列 $U_1U_2\cdots U_L$ 所表示的信源信息就载荷在码字 $X_1X_2\cdots X_K$ 中了。

信源译码是信源编码的逆过程，即由码字 $X_1X_2\cdots X_K$ 通过反变换处理恢复信源符号 u_i、信源符号序列 $U_1U_2\cdots U_L$。

无失真信源编码要求译码结果与信源的输出相比不产生歧义，信源符号或者信源符号序列与码本中的码字之间必须构成一种一一对应的变换关系。这不但要求每一个信源符号 u_i 和信源符号序列 $U_1U_2\cdots U_L$ 与码本中的码字 $X_1X_2\cdots X_K$ 具有确定的变换关系，而且要求由码符号序列到信源符号序列的反变换即译码的结果也是唯一的，即一串有限长的码符号序列的译码结果只能是唯一的一串信源符号序列。

因此，在无失真信源编码中，信源符号 u_i、信源符号序列 $U_1U_2\cdots U_L$ 与码字 $X_1X_2\cdots X_K$ 之间必须具有确定的一一对应的变换关系。

图 7－1 示出了无失真信源编码器中信源符号与码符号序列之间的变换关系。

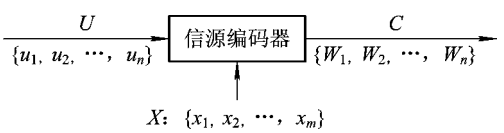

图 7－1 无失真信源编码器

在实际的通信系统和电子信息系统中，我们可以看到多种信源编码的应用实例。

电报通信系统中，信源符号集为 $U=\{$文字，符号$\}$，码符号集为 $X=\{$点，画，间隔$\}$，码本则是由不同长度的点、画的组合所构成的码字的集合。电报通信系统的信源编码通过文字、符号与码字的一一对应的变换处理来完成。

在计算机系统中，C 语言编写的程序由英文字母和符号组成，即 $U=\{A, B, \cdots, Z, +, -, \cdots\}$，而计算机机器码的码符号集为 $X=\{0, 1\}$，通过 ASCII 的映射关系将字母、符号组成的计算机程序编译连接为二进制符号序列。

由上述讨论可知，信源编码将信源输出的符号、消息转换为能够在信息系统中传输、存储和处理的信号，对于构成信息系统具有重要意义。信源编码器是各类通信系统和电子信息系统研究、开发中的一项不可缺少的组成部分。

对于离散无记忆信源 $[U, P]=\begin{bmatrix} u_1 & u_2 & \cdots & u_n \\ p_1 & p_2 & \cdots & p_n \end{bmatrix}$，信源符号等概率分布时，信源输出的信息量最大，即 $H_0=\log n$。然而一般信源是非等概分布的，其熵

$$H(U) = -\sum_{i=1}^{n} P(u_i)\log P(u_i) \leqslant H_0$$

因此，在信源输出的符号序列中存在一定的信息冗余，即

$$\gamma = 1 - \frac{H(U)}{H_0}$$

在信源编码处理中，信源符号 u_i 和信源符号序列 $U_1U_2\cdots U_L$ 与码字 $X_1X_2\cdots X_K$ 可以有多种对应关系，即可以有多种不同的编码方式。我们总可以在这些编码方式中选择一种最有效的编码，在将信源符号转换为信息系统要求的信号形式的同时，减小甚至消除信源的信息冗余。

由此可知，无失真信源编码的另一个重要作用就是选择适当的编码方法，在不产生信号畸变和失真的条件下，减小或去除信源符号中的冗余成分，提高通信系统和电子信息系统的有效性。

信源编码将信源符号 u_i 或信源符号序列 $U_1U_2\cdots U_L$ 变换成长度为 K 的码字 $X_1X_2\cdots X_K$（$X_j \in X$，$j=1$，2，…，K）。这样的编码在工程实现中可以分为两种形式。

第一种编码方法所使用的码字均由位数相同的码符号组合而成，即码本包含的码字 $X_1X_2\cdots X_K$ 中的 K 是一个不变的常数。这类码称为等长码。

另一种编码方法的码本中码字的长度可以不同，即对应于不同的信源符号 u_i 或信源符号序列 $U_1U_2\cdots U_L$，其编码的码字 $X_1X_2\cdots X_K$ 的长度 K 可能不同。这类编码称为变长码。

我们首先讨论等长信源编码中的基本问题和编码定理。

7.2　等长信源编码的相关概念和编码定理

7.2.1　等长码和编码效率

设有离散无记忆信源：

$$[U,\ P] = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \\ P(u_1) & P(u_2) & \cdots & P(u_n) \end{bmatrix}$$

根据系统的特性和应用要求，已确定码符号集为 $X=\{x_1, x_2, \cdots, x_m\}$。如果码字长度为 K，则码符号集中最多可以包含 m^K 个不同的码字。下面我们首先考虑对信源符号 $u_i(i=1, 2, \cdots, n)$ 进行无失真编码。

为了使信源符号与码字之间具有一一对应关系，显然必须满足：

$$m^K \geqslant n \tag{7.1}$$

即

$$K \geqslant \log_m n \tag{7.2}$$

例如，设某信源 U 中 $n=8$。由于信源含有 8 种符号，因此若进行二进制等长编码（$m=2$），则码字的长度 $K \geqslant \log 8 = 3$，即等长编码时码字的长度至少为 3 比特。

若离散无记忆信源 U 输出一个信源符号所含有的平均信息量为 $H(U)$，编码后一个信源符号使用 K 位码符号表示，则平均来讲，每一个码符号所携带的平均信息量（即编码的信息传输率，又称码率）为

$$R = \frac{H(U)}{K} \tag{7.3}$$

码率体现了信息传输有效性的高低。R 越大，表示每一个码元携带的平均信息量越多，即有效性越高；R 越小，表示每一个码元携带的平均信息量越少，即有效性越低。因为编码器的码符号集为 $X=\{x_1, x_2, \cdots, x_m\}$，所以编码器能够传输的最大信息率为 $\log m$。若编码码率达到了这一最大信息率，即

$$R = \log m$$

则表明通过编码去除了信源符号中的冗余，编码效率达到了 100%。一般情况下，编码效率可能达不到 100%，即 $R \leqslant \log m$。

为了衡量编码方法去除信源信息冗余的能力，定义等长码编码器的编码效率为

$$\eta = \frac{R}{\log m} = \frac{H(U)}{K \log m} \tag{7.4}$$

进一步可以求出编码后仍然存在的冗余度为

$$\gamma_{\mathrm{d}} = 1 - \eta = \frac{K \log m - H(U)}{K \log m} \tag{7.5}$$

如果对离散无记忆信源 U 的 L 次扩展信源 U^L 输出的样矢量 $\boldsymbol{u}_i = U_1U_2\cdots U_L$（$i=1, 2, \cdots, n^L$）进行无失真信源编码，则因为 L 次扩展信源 U^L 中包含的样矢量的个数为 n^L，所以为了实现无失真编码，使扩展信源 U^L 中每一个样矢量与码字之间具有一一对应关系，必须满足：

$$m^K \geqslant n^L \tag{7.6}$$

即

$$K \geqslant L \log_m n \tag{7.7}$$

例如，仍设某信源 U 中 $n=8$，若考虑其二次扩展信源（$L=2$）的二进制等长编码（$m=2$），则因为 U^2 中的样矢量个数为 $n^L = 8^2 = 64$，所以码符号集中码字的长度应当满足：

$$K \geqslant L \log_m n = 2 \log_2 8 = 6$$

即码字的长度至少为 6 比特。

在考虑离散无记忆信源的 L 次扩展信源 U^L 的无失真编码时，我们将信源输出的样矢量 $U_1U_2\cdots U_L$ 变换成码符号序列 $X_1X_2\cdots X_K$。在这样的处理中，我们不仅要求能够无失真地由 $X_1X_2\cdots X_K$ 恢复信源输出的样矢量 $U_1U_2\cdots U_L$，同时希望传送 $X_1X_2\cdots X_K$ 时需要的信息率达到最小。

因为编码器的码符号集仍为 $X=\{x_1, x_2, \cdots, x_m\}$，即 X 有 m 种取值，所以对于等长信源编码，码字 $X_1X_2\cdots X_K$ 的可能组合形式有 m^K 种。于是传送一个码字 $X_1X_2\cdots X_K$ 时，编码器输出的最大信息量为

$$\log m^K = K \log m \tag{7.8}$$

然而，传送一个码字 $X_1X_2\cdots X_K$ 对应于传送扩展信源输出的一个样矢量 $U_1U_2\cdots U_L$。由于一个样矢量由 L 个信源符号组成，因此当信源输出一个符号时，编码器需要输出的平均信息率为

$$R' = \frac{K \log m}{L} = \frac{K}{L} \log m \tag{7.9}$$

为了提高编码效率，显然应当在满足无失真的条件下寻找使编码器输出的信息率即 $\frac{K}{L} \log m$ 最小的编码方式。

然而，在无失真编、译码的要求下，信源编码处理所能够实现的最小信息率满足什么样的关系？其理论极限为何？下面讨论的等长无失真信源编码定理将回答这些问题。

7.2.2　信源符号序列的划分

为了证明等长无失真信源编码定理，我们首先分析离散无记忆信源 U 及其 L 次扩展信源 U^L 的统计特性和输出信源符号或样矢量所载荷的信息量所具有的统计关系，并依此将信源符号和样矢量进行分类划分，建立几个不等式。

设有离散无记忆信源：

$$[U,\ P]=\begin{bmatrix} u_1 & u_2 & \cdots & u_n \\ P(u_1) & P(u_2) & \cdots & P(u_n) \end{bmatrix}$$

其中，$\sum_{i=1}^{n} P(u_i)=1$，$0\leqslant P(u_i)\leqslant 1$，$i=1,2,\cdots,n$。

其 L 次扩展信源为

$$[U^L,\ P]=\begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_{n^L} \\ P(\boldsymbol{u}_1) & P(\boldsymbol{u}_2) & \cdots & P(\boldsymbol{u}_{n^L}) \end{bmatrix} \tag{7.10}$$

其中：

$$\boldsymbol{u}_i=(u_{i_1},\ u_{i_2},\ \cdots,\ u_{i_L})\quad i=1,2,\cdots,n^L;\ i_1,i_2,\cdots,i_L=1,2,\cdots,n \tag{7.11}$$

且

$$P(\boldsymbol{u}_i)=\prod_{l=1}^{L} P(u_{i_l})\qquad i=1,2,\cdots,n^L \tag{7.12}$$

根据系统的要求，设已选定编码符号集为 $X=\{x_1,x_2,\cdots,x_m\}$，于是，我们可以考虑对信源 U 和 L 次扩展信源 U^L 进行码字长度为 K 的等长无失真编码。

由 7.2.1 节的讨论我们知道，无失真信源编码要求信源符号或扩展信源的样矢量与其编码的码字之间必须具有一一对应的变换关系。不同的信息源及其扩展信源中信源符号的个数和样矢量的个数可能不同，然而由码符号集可以看出，能够组合使用的码字最多只有 m^K 个，即编码器可以使用的码字个数受到了限制。因此，在无失真信源编码中，我们只能够设法充分、有效地利用码本中的 m^K 个码字，使信源 U 和扩展信源 U^L 输出的信息量中贡献较大的信源符号和样矢量有确定的一一对应的码字，以保证包含信息量较多的这一部分信源符号和样矢量不产生译码差错。对于输出信息量很少的信源符号和样矢量，由于有 m^K 个可用码字的限制，因此不能够保证分配一一对应的码字。虽然这一部分信源符号和样矢量在译码时可能会产生失真，但是由于它们对信源和扩展信源输出的信息量贡献很小，因此仍可使信源编码系统的失真得到严格的控制。

下面我们将在编码器可用码字受到限制的实际情况下，按照这样的编码处理思想分析信源符号和样矢量的信息特性及它们与信源熵的关系，找出使译码失真最小，满足无失真信源编码要求的统计关系。

下面主要以信源符号编码为例进行讨论。由第 2 章的讨论我们知道，反映信源符号 u_i $(i=1,2,\cdots,n)$基本特性的参数是其载荷的信息量，即信源符号的自信息 $I(u_i)$。

对于信源符号 u_i，其发生的概率为 $P(u_i)$，当信源输出 u_i 时，u_i 载荷的自信息为

$$I(u_i)=-\log P(u_i)\qquad i=1,2,\cdots,n$$

信源输出大小为 $I(u_i)$的自信息是伴随着信源输出 u_i 而发生的。我们知道，u_i 是一个发生概率为 $P(u_i)$的随机变量，因此 $I(u_i)$也是一个随机变量，它与 u_i 具有相同的概率分布，即 $P(u_i)(i=1, 2, \cdots, n)$。

可以求出 $I(u_i)$的均值和方差：

$$E[I(u_i)]=-\sum_{i=1}^{n}P(u_i)\log P(u_i)=H(U) \tag{7.13}$$

$$\begin{aligned}D[I(u_i)]&=\sigma^2(U)=E\{[I(u_i)-H(U)]^2\}\\&=E[I(u_i)]^2-2H(U)E[I(u_i)]+[H(U)]^2\\&=E\{[I(u_i)]^2\}-[H(U)]^2\end{aligned} \tag{7.14}$$

当 n 为有限值时，$\sigma^2(U)<+\infty$是一个有限的定值。

信源符号的自信息的统计平均值即为信源的熵 $H(U)$，也就是信源 U 所输出的平均信息量。然而，式(7.13)表明，不同的信源符号 u_i 所携带的自信息量对信源输出的平均信息量的贡献 $-P(u_i)\log P(u_i)(i=1, 2, \cdots, n)$一般是不相同的。

结合式(7.14)可以进一步看出，如果信源符号 u_i 的自信息 $I(u_i)$与信源 U 的熵 $H(U)$接近(偏差较小)，则相应地 $-P(u_i)\log P(u_i)$在熵 $H(U)$的计算中将发挥较大的作用。这表明这一类信源符号 u_i 所载荷的自信息量在信源 U 输出的平均信息量中占有重要成分，对信源 U 输出信息的贡献较大。反之，如果 $I(u_i)$与信源 U 的熵 $H(U)$相差较远(偏差较大)，则相应地 $-P(u_i)\log P(u_i)$在熵 $H(U)$的计算中作用很小，该信源符号 u_i 对信源 U 输出信息的贡献也就很小。

于是，我们可以依据信源符号 u_i 输出的自信息 $I(u_i)$与信源输出的平均信息量 $H(U)$的偏差大小，将信源符号划分为不同的类型。

根据契比雪夫(Chebyshev)不等式，对于任意的 $\varepsilon>0$，有

$$P\{|I(u_i)-H(U)|\geqslant\varepsilon\}\leqslant\frac{\sigma^2(U)}{\varepsilon^2} \tag{7.15}$$

式(7.15)反映出了信源符号的自信息与信源的熵的偏差(两者之差的绝对值)不小于 ε 的事件发生的概率与给定的偏差范围 ε、$I(u_i)$的方差 $\sigma^2(U)$的关系。

下面我们考虑离散无记忆信源 U 的 L 次扩展信源 U^L(见式(7.10)～式(7.12))中的关系。

L 次扩展信源 U^L 输出的样矢量 $\boldsymbol{u}_i=(u_{i_1}, u_{i_2}, \cdots, u_{i_L})(i=1, 2, \cdots, n^L)$所载荷的自信息为

$$\begin{aligned}I(\boldsymbol{u}_i)&=-\log P(\boldsymbol{u}_i)=-\log\prod_{l=1}^{L}P(u_{i_l})\\&=-\sum_{l=1}^{L}\log P(u_{i_l})=\sum_{l=1}^{L}I(u_{i_l})\end{aligned} \tag{7.16}$$

样矢量的自信息 $I(\boldsymbol{u}_i)$是伴随着随机样矢量 $\boldsymbol{u}_i$ 的发生而发生的随机变量，与样矢量 $\boldsymbol{u}_i$ 有相同的概率分布。于是，我们可以求出 $I(\boldsymbol{u}_i)$的均值和方差：

$$E[I(\boldsymbol{u}_i)]=\sum_{l=1}^{L}E[I(u_{i_l})]=\sum_{l=1}^{L}H(U)=LH(U)=H(U^L) \tag{7.17}$$

$$\sigma^2[I(\boldsymbol{u}_i)]=D[I(\boldsymbol{u}_i)]=\sum_{l=1}^{L}D[I(u_{i_l})]=\sum_{l=1}^{L}\sigma^2(U)=L\sigma^2(U) \tag{7.18}$$

显然，样矢量的自信息 $I(\boldsymbol{u}_i)$ 的统计平均值即为 L 次扩展信源 U^L 所输出的平均信息量，它是离散无记忆信源 U 的熵的 L 倍；方差 $\sigma^2[I(\boldsymbol{u}_i)]$ 反映出了样矢量的自信息与 U^L 的熵之间的平均偏差。与式(7.14)相比可以看出，样矢量自信息 $I(\boldsymbol{u}_i)$ 的方差增大为了信源符号自信息 $I(\boldsymbol{u}_i)$ 的方差的 L 倍。

为了分析 L 次扩展信源 U^L 所输出的样矢量 $\boldsymbol{u}_i(i=1,2,\cdots,n^L)$ 对 U^L 输出信息的贡献大小，以区分不同的样矢量对信源输出信息的重要程度，我们仍然从样矢量的自信息 $I(\boldsymbol{u}_i)$ 与其统计平均值的偏差入手。

设已给定一个 $I(\boldsymbol{u}_i)$ 与 $H(U^L)$ 的偏差控制范围 $L\varepsilon>0$。根据契比雪夫不等式，某一样矢量 $\boldsymbol{u}_i$ 的自信息与 $LH(U)$ 的偏差超过 $L\varepsilon$（即事件 $|I(\boldsymbol{u}_i)-LH(U)|\geqslant L\varepsilon$ 发生）的概率满足：

$$\begin{aligned}P\{|I(\boldsymbol{u}_i)-LH(U)|\geqslant L\varepsilon\}&=P\left\{\left|\frac{I(\boldsymbol{u}_i)}{L}-H(U)\right|\geqslant\varepsilon\right\}\\&\leqslant\frac{L\sigma^2(U)}{(L\varepsilon)^2}=\frac{\sigma^2(U)}{L\varepsilon^2}\end{aligned} \tag{7.19}$$

由式(7.10)我们知道，L 次扩展信源 U^L 中有 n_L 个不同的样矢量 $\boldsymbol{u}_i$，这些样矢量所包含的自信息在 U^L 输出的平均信息量中有不同的贡献。根据 $\boldsymbol{u}_i$ 载荷的自信息与其统计平均值的偏差，我们可以将这 n^L 个样矢量分类并构成两个互补的集合 A_ε 和 $\overline{A_\varepsilon}$。组成这两个集合的样矢量分别满足：

$$A_\varepsilon=\left\{\boldsymbol{u}_i:\left|\frac{I(\boldsymbol{u}_i)}{L}-H(U)\right|<\varepsilon\right\} \tag{7.20}$$

$$\overline{A_\varepsilon}=\left\{\boldsymbol{u}_i:\left|\frac{I(\boldsymbol{u}_i)}{L}-H(U)\right|\geqslant\varepsilon\right\} \tag{7.21}$$

由不等式(7.19)可以知道，集合 $\overline{A_\varepsilon}$ 所表示的事件发生的概率为

$$P(\overline{A_\varepsilon})\leqslant\frac{\sigma^2(U)}{L\varepsilon^2}$$

根据概率分布律，集合 $\overline{A}_\varepsilon$ 所表示的事件发生的概率还应当满足：

$$P(\overline{A_\varepsilon})\geqslant 0$$

于是，有

$$0\leqslant P(\overline{A_\varepsilon})\leqslant\frac{\sigma^2(U)}{L\varepsilon^2} \tag{7.22}$$

另外，由于集合 A_ε 和 $\overline{A_\varepsilon}$ 是互补的，因此集合 A_ε 所表示的事件发生的概率为

$$1\geqslant P(A_\varepsilon)\geqslant 1-\frac{\sigma^2(U)}{L\varepsilon^2} \tag{7.23}$$

可以看出，当 $L\to\infty$ 时，有

$$P(A_\varepsilon)\to 1,\quad P(\overline{A_\varepsilon})\to 0 \tag{7.24}$$

即 L 次扩展信源 U^L 中的 n^L 个样矢量全部属于集合 A_ε，而集合 $\overline{A_\varepsilon}$ 中样矢量的个数为零。

由样矢量划分集合的定义可知，如果样矢量 $\boldsymbol{u}_i\in\overline{A_\varepsilon}$，则其携带的自信息 $I(\boldsymbol{u}_i)$ 与 U^L 的熵 $H(U^L)$ 具有很大的偏差，表明该样矢量对于 U^L 输出信息的贡献很小。反之，如果样矢

量 $\boldsymbol{u}_i \in A_\varepsilon$，则它所携带自信息 $I(\boldsymbol{u}_i)$ 与 U^L 的熵 $H(U^L)$ 的偏差在规定的范围 ε 之内，该样矢量的自信息在 U^L 输出平均信息量的计算中发挥了较大的作用。因此，在无失真信源编码中，集合 A_ε 中的样矢量 $\boldsymbol{u}_i$ 应当重点考虑。

那么，L 次扩展信源 U^L 的 n^L 个样矢量中，属于集合 A_ε 的样矢量个数满足怎样的关系呢？能否保证在 m^K 个可用码字中找到一一对应的编码关系？我们将从 $\boldsymbol{u}_i \in A_\varepsilon$ 的自信息的特点入手，找出集合 A_ε 中样矢量的个数所满足的不等式关系。

设集合 A_ε 中样矢量 $\boldsymbol{u}_i$ 的个数为 M_ε，显然，$M_\varepsilon \leqslant n^L$。

由集合 A_ε 的定义可知，样矢量 $\boldsymbol{u}_i \in A_\varepsilon$ 所具有的共同特点是：它们的自信息与信源熵的偏差在给定的范围 ε 之内，即对于 $\boldsymbol{u}_i \in A_\varepsilon$，有：

$$-\varepsilon < \frac{I(\boldsymbol{u}_i)}{L} - H(U) < \varepsilon$$

$$L(H(U)-\varepsilon) < I(\boldsymbol{u}_i) < L(H(U)+\varepsilon)$$

$$L(H(U)-\varepsilon) < -\log P(\boldsymbol{u}_i) < L(H(U)+\varepsilon)$$

$$-L(H(U)-\varepsilon) > \log P(\boldsymbol{u}_i) > -L(H(U)+\varepsilon)$$

因此

$$2^{-L(H(U)-\varepsilon)} > P(\boldsymbol{u}_i) > 2^{-L(H(U)+\varepsilon)} \tag{7.25}$$

由于事件 A_ε 发生的概率为 $\boldsymbol{u}_i \in A_\varepsilon$ 的概率之和，因此依式(7.25)左端可以得到，事件 A_ε 发生的概率满足下面的关系：

$$P(A_\varepsilon) = \sum_{\boldsymbol{u}_i \in A_\varepsilon} P(\boldsymbol{u}_i) \leqslant M_\varepsilon \times \max_{\boldsymbol{u}_i \in A_\varepsilon} P(\boldsymbol{u}_i) < M_\varepsilon \times 2^{-L(H(U)-\varepsilon)} \tag{7.26}$$

同理，由式(7.25)右端的不等式关系可以得到：

$$P(A_\varepsilon) = \sum_{\boldsymbol{u}_i \in A_\varepsilon} P(\boldsymbol{u}_i) \geqslant M_\varepsilon \times \min_{\boldsymbol{u}_i \in A_\varepsilon} P(\boldsymbol{u}_i) > M_\varepsilon \times 2^{-L(H(U)+\varepsilon)} \tag{7.27}$$

在前面的讨论中，我们根据契比雪夫不等式以及集合 A_ε 和 $\overline{A_\varepsilon}$ 的互补性，从严格的概率关系得出了事件 A_ε 发生概率的取值范围(见式(7.23))。同时，我们也通过对样矢量 $\boldsymbol{u}_i \in A_\varepsilon$ 的概率求和，得到了式(7.26)和式(7.27)指出的 $P(A_\varepsilon)$ 的上、下界。显然，式(7.26)和式(7.27)所指出的上、下界应当在式(7.23)所规定的范围之内。

于是，事件 A_ε 发生的概率的上界(见式(7.26)的右端)一定大于式(7.23)所规定的最小值，即

$$M_\varepsilon \times 2^{-L(H(U)-\varepsilon)} > 1 - \frac{\sigma^2(U)}{L\varepsilon^2} \tag{7.28}$$

同理，事件 A_ε 发生的概率的下界(见式(7.27)的右端)也不会超过式(7.23)所规定的最大值，即

$$M_\varepsilon \times 2^{-L(H(U)+\varepsilon)} < 1 \tag{7.29}$$

因此，集合 A_ε 中样矢量 $\boldsymbol{u}_i$ 的个数 M_ε 满足不等式：

$$\left(1 - \frac{\sigma^2(U)}{L\varepsilon^2}\right) 2^{L(H(U)-\varepsilon)} < M_\varepsilon < 2^{L(H(U)+\varepsilon)} \tag{7.30}$$

应用上述分析所得到的基本关系和不等式，我们便可以证明无失真等长信源编码定理。

7.2.3　等长信源编码定理

定理 7.1（等长信源编码定理）　由 L 个符号组成、每个符号的熵为 $H(U)$ 的离散无记忆平稳信源符号序列 $U_1U_2\cdots U_L$ 可以用 K 个符号 $X_1X_2\cdots X_K$（每个符号有 m 种取值可能）进行等长编码。对于任意的 $\varepsilon>0$，$\delta>0$，只要：

$$\frac{K}{L}\log m \geqslant H(U)+\varepsilon \tag{7.31}$$

当 L 足够大时，就可使译码差错小于 δ。

反之，当

$$\frac{K}{L}\log m \leqslant H(U)-2\varepsilon \tag{7.32}$$

时，译码差错一定是有限值。当 L 足够大时，译码几乎必定出错。

证明：这个定理由正定理和逆定理两部分组成。下面先证明正定理。

设离散无记忆信源有 n 种输出符号，则其 L 次扩展信源 U^L 有 n^L 种不同的样矢量 $\boldsymbol{u}_i$。根据前面给出的无失真信源编码的基本思想，在正定理的证明中我们并不把这 n^L 种样矢量 $\boldsymbol{u}_i$ 全部进行一一对应的编码，而是只保证对扩展信源 U^L 输出信息贡献较大的属于集合 A_ε 中的 M_ε 个样矢量 $\boldsymbol{u}_i$ 实现一一对应的编码。由定理给出的条件可知，由于码符号集中有 m 个码符号，因此可以组成 m^K 个长度为 K 的码字，于是只要 m^K 大于 M_ε，就能够实现上述编码方案。

定理中的式(7.31)给出的条件等价于：

$$\log m^K \geqslant L(H(U)+\varepsilon)$$

即

$$m^K \geqslant 2^{L(H(U)+\varepsilon)}$$

而由式(7.30)可知：

$$M_\varepsilon < 2^{L(H(U)+\varepsilon)}$$

所以，在满足式(7.31)给定的条件时，一定能够满足 $m^K>M_\varepsilon$。

因此，我们总可以找到一种编码方法，使得集合 A_ε 中的 M_ε 个样矢量 $\boldsymbol{u}_i$ 与 m^K 个码字有一一对应的关系，译码将不会产生错误。但是，此时并没有保证集合 $\overline{A_\varepsilon}$ 中的样矢量 $\boldsymbol{u}_i$ 也可以找到一一对应的输出码字。因此，译码中样矢量 $\boldsymbol{u}_i\in\overline{A_\varepsilon}$ 可能会产生错误。

由此可知，经过这样的编码处理，系统产生的失真仅由样矢量 $\boldsymbol{u}_i\in\overline{A_\varepsilon}$ 的译码差错造成，系统的译码差错概率 p_e 不会超过 $\overline{A_\varepsilon}$ 所发生的概率。由前面关于集合 A_ε 与 $\overline{A_\varepsilon}$ 的划分的讨论和式(7.22)所指出的概率关系可知，在此条件下译码差错概率满足：

$$p_e \leqslant P(\overline{A_\varepsilon}) \leqslant \frac{\sigma^2(U)}{L\varepsilon^2} \tag{7.33}$$

当 ε^2、σ^2 均为确定值时，对于任意小的 $\delta>0$，当 L 足够大，即

$$L > \frac{\sigma^2(U)}{\varepsilon^2\delta} \tag{7.34}$$

时，译码差错的概率 p_e 一定满足 $p_e<\delta$，并且有

$$\lim_{L\to\infty} p_e = 0$$

正定理得证。

在逆定理中，式(7.32)所给出的条件等价于：

$$\log m^K \leqslant L(H(U)-2\varepsilon)$$

即

$$m^K \leqslant 2^{L(H(U)-2\varepsilon)} = 2^{L(H(U)-\varepsilon)} \times 2^{-L\varepsilon} \tag{7.35}$$

表明在逆定理所给出的条件下，编码器可以使用的码字个数受到了限制，需要判断在逆定理条件下属于集合 A_ε 的样矢量 $\boldsymbol{u}_i$ 能否找到一一对应的编码。

假定集合 A_ε 中样矢量 $\boldsymbol{u}_i$ 的个数仅为其最小值。依据式(7.30)指出的集合 A_ε 中样矢量 $\boldsymbol{u}_i$ 的个数 M_ε 的取值范围，M_ε 的取值也必须满足：

$$M_\varepsilon > \left(1-\frac{\sigma^2(U)}{L\varepsilon^2}\right)2^{L(H(U)-\varepsilon)} \tag{7.36}$$

即集合 A_ε 中的样矢量 $\boldsymbol{u}_i$ 最少时，M_ε 也不可能小于其下界。

为了找出此 M_ε 的下界与满足逆定理条件时码字个数 m^K 的关系，将式(7.35)除以式(7.36)，便有

$$\frac{m^K}{M_\varepsilon} < \frac{2^{-L\varepsilon}}{1-\dfrac{\sigma^2(U)}{L\varepsilon^2}} \tag{7.37}$$

当信源符号序列长度 L 大于某定值 L_0 时，由于 $2^{-L\varepsilon}$ 减小得更快，必可使 $\dfrac{m^K}{M_\varepsilon}<1$，即

$$m^K < M_\varepsilon$$

因此，集合 A_ε 中，一定会有一部分样矢量 $\boldsymbol{u}_i$ 在 m^K 个码字中找不到一一对应的变换关系，使得译码时无法正确地恢复而出错。所以，在逆定理的条件下，译码差错不可能是无限小。

进一步，若 $L\to\infty$，则由式(7.37)可知

$$\frac{m^K}{M_\varepsilon} \to 0 \tag{7.38}$$

由式(7.24)得：$P(A_\varepsilon)\to 1$ 而 $P(\overline{A_\varepsilon})\to 0$，即当 $L\to\infty$ 时，L 次扩展信源中的 n^L 个样矢量 $\boldsymbol{u}_i$ 全部在集合 A_ε 中，然而可以使用的码字的总数 m^K 却远小于 n^L，以至于集合 A_ε 中的绝大多数样矢量 $\boldsymbol{u}_i$ 无法找到一一对应的码字，译码时几乎必定出错，此即逆定理所给出的结论。

证毕。

定理 7.1 表明，在进行等长无失真信源编码时，信源的平均符号熵 $H(U)$ 是编码器输出信息率的一个临界值。由式(7.9)可知，当编码器允许的输出信息率，即信源每输出一个信源符号编码器必须输出的信息率为

$$R' = \frac{K}{L}\log m$$

时，只要 $R'>H(U)$，即编码器允许的输出信息率大于信源输出的信息率(即信源的熵)，编码器就一定可以做到几乎无失真，译码差错接近于零(见式(7.33))，条件是信源符号序列的长度 L 足够大(见式(7.34))。反之，当 $R'<H(U)$ 时，不可能构成无失真的编码，即不存在一种编码器，能够使译码差错的概率趋近于零。$R'=H(U)$ 为编码器的临界状态，可能会产生译码差错，也可能不产生译码差错。

例如对于信源 U，当 n 个信源符号为等概率分布且 n 恰为 2 的整数幂，即 $n=2^{K_0}$ 时，信源的熵为

$$H(U)=\log n=\log 2^{K_0}=K_0 \quad \text{比特 / 符号}$$

采用二进制 K_0 位的等长编码可以实现 n 个信源符号与 2^{K_0} 个码字之间一一对应的变换关系，译码无失真且编码器的输出信息率恰为 $R'=H(U)$。如果信源非等概分布，则在 $R'=H(U)$时一般会产生译码差错。

定理 7.1 从理论上阐明了无失真且编码效率接近于 1 的理想等长信源编码器的存在性。按照定理 7.1 及其证明过程中的处理方法，离散无记忆信源的 L 次扩展信源 U^L 输出的样矢量 $\boldsymbol{u}_i$ 可以用 K 个码符号(码符号集有 m 个元素)组成的等长码进行编码。在给定 ε 和 δ 之后，按照式(7.34)规定的信源符号序列长度 L 的值计算每一个样矢量 $\boldsymbol{u}_i$ 的概率，将概率较大的样矢量选为集合 A_ε 中的元素，并使 $P(A_\varepsilon)=1-\delta$，然后将 A_ε 中的样矢量使用一一对应的码字编码表示，便完成了对 L 次扩展信源 U^L 的编码处理。只要所取的 δ 足够小，译码差错的概率 $p_e<\delta$ 将足够小。这时编码器的编码效率为

$$\eta=\frac{\dfrac{H(U^L)}{K}}{\log m}=\frac{\dfrac{LH(U)}{K}}{\log m}=\frac{H(U)}{\dfrac{K}{L}\log m}\leqslant\frac{H(U)}{H(U)+\varepsilon} \tag{7.39}$$

只要所取的 ε 足够小，η 便接近于 1。

但是，要使 ε、δ 足够小，便要求 L 足够大。对于 $p_e\to 0$，$\eta\to 1$ 的理想编码器，必须使 $L\to\infty$。显然，这是无法实现的。

【例 7.1】 已知某信源的概率空间为

$$[U,\ P(u_i)]=\begin{bmatrix} u_1 & u_2 & u_3 & u_4 & u_5 & u_6 & u_7 & u_8 \\ 0.40 & 0.18 & 0.10 & 0.10 & 0.07 & 0.06 & 0.05 & 0.04 \end{bmatrix}$$

可以求出 $H(U)=2.5524$(比特/符号)和 $\sigma^2(U)=7.82$。若要求编码效率为 90%，即

$$\eta=\frac{H(U)}{R}=\frac{H(U)}{H(U)+\varepsilon}=0.90$$

可解得 $\varepsilon\approx 0.28$。

设译码差错率为 10^{-6}，即

$$p_e=\delta=10^{-6}$$

由式(7.34)可知，信源符号序列的长度必须满足：

$$L\geqslant\frac{\sigma^2(U)}{\varepsilon^2\delta}=\frac{7.82}{0.28^2\times 10^{-6}}\approx 10^8$$

可见，在差错率和编码效率的要求并不是很高的情况下，必须把近 10^8 个信源符号一起编码，才能够满足要求。

因此，我们需要研究更加有效并且便于工程实现的编码方法，如变长编码的方法。

7.3　变长码的基本概念

从码符号集 $X=\{x_1, x_2, \cdots, x_m\}$中可以选出若干位码元(如 K 位)组成码字 $X_1X_2\cdots X_K$。在由这样的码字构成的码本中，如果码字的长度 K 并不是一个固定不变的量，则这样的一组

码字便被称为变长码。

下面首先讨论有关变长码的一些基本概念和关系。

7.3.1 平均码长与编码效率

由前面几节的讨论我们知道，所谓信源编码，是将信源符号 u_i 或扩展信源的样矢量 $\boldsymbol{u}_i$ 与码本 C 中的码字 $X_1X_2\cdots X_K$ 构成映射关系。因为变长码码本 C 中码字的长度不同，所以经过信源编码处理，与不同的 u_i 或 $\boldsymbol{u}_i$ 对应的码符号序列(码字)所含有的码符号个数即编码长度(码长)各不相同。为了衡量变长编码的效果，需要引入平均码长的概念。

对于离散无记忆信源：

$$[U,\ P]=\begin{bmatrix} u_1 & u_2 & \cdots & u_n \\ P(u_1) & P(u_2) & \cdots & P(u_n) \end{bmatrix}$$

在实现一一对应的无失真变长信源编码后，相应于各信源符号的码字：

$$W_1,\ W_2,\ \cdots,\ W_n$$

分别具有码长：

$$K_1,\ K_2,\ \cdots,\ K_n$$

已知信源符号 u_i 发生的概率为 $P(u_i)$，码长 K_i 发生的概率也为 $P(u_i)$，于是我们可以知道这组码字的平均码长为

$$\overline{K}=\sum_{i=1}^{n}P(u_i)K_i \tag{7.40}$$

平均码长表示了每一个信源符号平均使用的码元个数，其单位为码符号/信源符号。

若已知信源的熵为 $H(U)$，即信源输出一个信源符号时平均输出的信息量为 $H(U)$比特/信源符号，那么，变长信源编码器输出一位码元时所输出的平均信息率为

$$R=\frac{H(U)}{\overline{K}}\quad \text{比特 / 码符号} \tag{7.41}$$

已知码符号集由 m 个符号组成，编码器输出的最大信息率便为 $\log m$ 比特/码符号。若编码器的输出信息率达到了此最大信息率，即 $R=\log m$，则表明通过编码消除了信源符号中的冗余，编码效率为 100%。

在等长信源编码中，我们定义了编码效率来衡量编码器去除信源信息冗余的能力。因为等长编码中 $K=\overline{K}$，由式(7.4)和式(7.5)我们可以推出变长信源编码时的编码效率和编码冗余，即

$$\eta=\frac{R}{\log m}=\frac{H(U)}{\overline{K}\ \log m} \tag{7.42}$$

和

$$\gamma_{\mathrm{d}}=1-\eta=\frac{\overline{K}\ \log m-H(U)}{\overline{K}\ \log m} \tag{7.43}$$

同样，对于 L 次扩展信源 U^L，熵为 $H(U^L)=LH(U)$，其变长无失真信源编码是将输出的样矢量 $U_1U_2\cdots U_L$ 变换成长度不同的码符号序列 $X_1X_2\cdots X_K$。设码符号集 $X=\{x_1,\ x_2,\ \cdots,\ x_m\}$，样矢量的平均码长为 $\overline{K}_L$(单位为比特/样矢量)。于是，由式(7.42)和式(7.43)可知，扩展信源 U^L 的变长无失真信源编码的编码效率和编码冗余为

$$\eta = \frac{R}{\log m} = \frac{H(U^L)}{\overline{K}_L \log m} = \frac{LH(U)}{\overline{K}_L \log m} = \frac{H(U)}{\frac{\overline{K}_L}{L}\log m} \tag{7.44}$$

$$\gamma_d = 1 - \eta = \frac{\overline{K}_L \log m - H(U^L)}{\overline{K}_L \log m} = \frac{\frac{\overline{K}_L}{L}\log m - H(U)}{\frac{\overline{K}_L}{L}\log m} \tag{7.45}$$

其中，$\overline{K}_L/L$ 表示与一个信源符号相对应的平均码长。

7.3.2　变长码的结构与分类

如果对信源 U 进行变长编码，我们对于经常出现(概率较大)的信源符号可以用一个短码(K_i 较小)表示，而对于出现概率很小的信源符号使用长码(K_i 较大)表示，则平均码长 $\overline{K} = \sum_{i=1}^{n} P(u_i)K_i$ 相对于等长编码时将会减小。因此，选择恰当的变长编码可以提高编码效率。

然而，在各类通信系统和电子信息系统的设计中，并不是任意一种变长码都可以使用，即工程中允许使用的变长码必须满足一定的条件。

在下面的例子中，我们给出几种不同的变长码。虽然它们的平均码长 $\overline{K}$ 都较小，但是有的码却不能够在工程中使用。

【例 7.2】　对于信源：

$$[U, P] = \begin{bmatrix} u_1 & u_2 & u_3 & u_4 \\ 0.5 & 0.25 & 0.125 & 0.125 \end{bmatrix}$$

使用码符号集 $X=\{0, 1\}$ 作二进制变长编码。此处我们给出四种编码方案：

信源符号	编码 A	编码 B	编码 C	编码 D
u_1	0	0	0	0
u_2	0	1	01	10
u_3	1	00	011	110
u_4	10	10	0111	111

对于编码 A，信源符号 u_1 和 u_2 的码字都是 0，当接收端收到码符号 0 时便无法确定信源输出的是哪一个符号。

对于编码 B，如果接收到符号序列 00，则译码时无法确定信源的输出是 u_3 还是 u_1u_1。

对于编码 C 和编码 D，设接收到一个码符号序列 0010110…。编码 C 解码得到信源符号序列 $u_1u_2u_3\cdots$，而编码 D 的译码结果为信源符号序列 $u_1u_1u_2u_3\cdots$。

由上述分析可知，编码 C 和编码 D 能够用于无失真变长编码，而编码 A 和编码 B 则不能唯一、正确地恢复信源符号和信源符号序列，不能用于信源编码。

因此，在各类通信系统和电子信息系统中使用的信源编码方案必须具有一定的性质，满足特定的码字结构要求。

1. 唯一可译码

对于每一个有限长的信源符号序列，其编码形成的码符号序列不与任何其他的码符号

序列相重合，则称这样的码是唯一可译码，否则为非唯一可译码。

工程中使用的编码方案必须是这种满足唯一可译性或单义可译性的唯一可译码。由唯一可译码所构成的码字序列，可以不需要任何间隔符号就能够在接收端正确地分离码字并译出相应的信源符号或信源符号序列。

在例 7.2 中，编码 A 和编码 B 的译码结果都可能会产生歧义，编码 A 和编码 B 是非唯一可译码。而对于某一个码符号序列，编码 C 和编码 D 分别得到了一个唯一的信源符号序列，因此编码 C 和编码 D 是满足唯一可译性的唯一可译码。

虽然编码 C 和编码 D 都是唯一可译码，但是它们两者之间也有各自不同的特点。

编码 C 中，u_1 的码字 0 事实上构成了其他信源符号的编码码字的字头，即信源符号 u_2、u_3 和 u_4 的码字中起始的部分均为 u_1 的码字 0。同理，u_2 的码字 01 构成了 u_3 和 u_4 的码字 011 和 0111 的字头；u_3 的码字 011 是 u_4 码字 0111 的字头。由于码字结构的这一特点，对一个码符号序列进行码字分离处理时，若某一码字的最后一位码符号出现，则我们并不能即时地完成该码字的分离。因为这组码符号可能是某一信源符号的码字，也可能是另外一些信源符号的码字的字头，只有当下一个信源符号的码字中的第一位码符号被接收到时，我们才能完成前一个码字与下一个码字的分离。

与编码 C 相比，编码 D 的结构有不同的特点。编码 D 中的任何信源符号的码字都不构成其他码字的字头。于是对一个码符号序列的码字分离，不吸引下一个信源符号的码字出现，只要当前信源符号的码字接收完毕，即该码字的最后一位码符号出现，便可即时地完成该码字的分离处理。

2. 字首码

若任一信源符号码字的前面任一起始部分都不构成其他信源符号的码字，则称具有这种结构的编码为字首码。

字首码可以实现码字的即时分离，又称为即时码(可以即时地译码)。

显然，字首码满足唯一可译性，即字首码是唯一可译码的一类子码，字首码一定是唯一可译码。反之，唯一可译码不一定是字首码，有些唯一可译码可能不满足字首性。

根据字首码的结构特点我们知道，例 7.2 中，编码 D 是字首码，而编码 C 的结构不满足字首性，故编码 C 是非字首码。

按照字首性结构的要求仍然可以得到不同的编码方案，这些不同的编码方案的平均码长可能不同。

例如，对于例 7.2 给出的信源 U 的四个信源符号，我们可再构造一种编码 E：

$$u_1 \to 0,\ u_2 \to 101,\ u_3 \to 110,\ u_4 \to 111$$

编码 E 不仅具有唯一可译性，而且满足字首性，也是字首码。因为编码 D 和编码 E 中 u_2 码字的长度不同，所以 $\overline{K}_E > \overline{K}_D$。

3. 紧致码

平均码长最短的字首码称为紧致码。

通过上面对变长码的各种结构特点和性质的讨论，我们可以将变长码划分为如图 7-2 所示的几种类型。

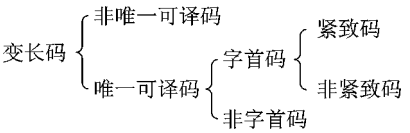

图 7－2　变长码的分类

信源编码的主要目的是去除信源的信息冗余，提高系统的有效性，因此，在考虑变长无失真信源编码中，希望平均码长尽量地短。然而，通过对各类变长码的结构特点的分析，我们应当明确，在通信系统和电子信息系统的设计中，所追求的平均码长 $\overline{K}$ 最短(有效性最高)是有条件的。最优的变长编码并不是单纯地追求平均码长 $\overline{K}$ 最短，而是在满足唯一可译性和字首性的条件下追求平均码长 $\overline{K}$ 最短，即对于某一离散无记忆信源 U 和 L 次扩展信源 U^L，寻找其变长无失真编码的紧致码。

紧致码是满足唯一可译性和字首性时平均码长最短的码。下面我们给出一种表示和构造字首码的方法，然后讨论一组字首码的码长所满足的关系，即字首码存在的条件。

7.3.3　码的树形图表示

设码符号集为 $X=\{x_1, x_2, \cdots, x_m\}$，由 X 中选出 K 个码符号构成的码字 $X_1X_2\cdots X_K$ 可以使用 m 进制的树形图来表示。这样的树形图我们也称为码树图。

所谓树形图，是一种由根、枝和节点表示的树状图形。

图 7－3 所示即为一树形图。其中，根是树形图的出发点(如图中的点 A)。由根可以伸出枝，枝的另外一端称为节点，每一个节点又可以再伸出新的枝，不再伸出枝的节点即为叶。在下面的讨论中，我们将继续伸枝的节点称为中间节点，见图 7－3 中的点 B 和 C，而将不再伸枝的节点即树叶称为终端节点，见图 7－3 中的点 D、E、F 和 G。

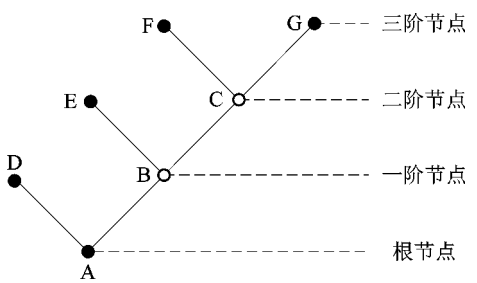

图 7－3　码树图

如果每一个根或中间节点最多可以伸出 m 个枝，则树形图是 m 进制的。由根到叶可以串接不同节数的枝，距根为 k 节的节点称为 k 阶节点。如果一个树形图中串接的最大枝数为 K，即阶数最高的终端节点为 K 阶，则称树形图是 K 阶的树形图。由此可知，图 7－3 所示为一个二进制三阶的树形图。

在一个 m 进制 K 阶的树形图中，如果所有的 $K-1$ 阶节点都伸出了 m 个枝，全部 K 阶节点都成为终端节点，则这样的树形图是一个 K 阶的满树，否则该树形图是非满树。显然，m 进制 K 阶的满树有 m^K 个终端节点。

图 7－4 所示为一个三进制三阶的满树，它有 $m^K=3^3=27$ 个终端节点。在图 7－5 中，一阶节点 B 和二阶节点 C、D、E、F 都没有伸出枝，树形图的终端节点中有一个一阶节点、四个二阶节点和六个三阶节点。因此，图 7－5 是一个非满树。

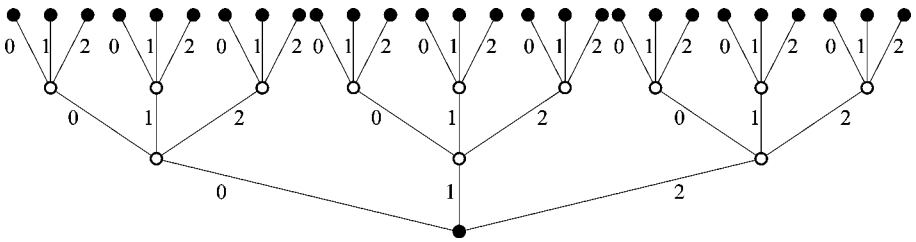

图 7－4 满树码树图

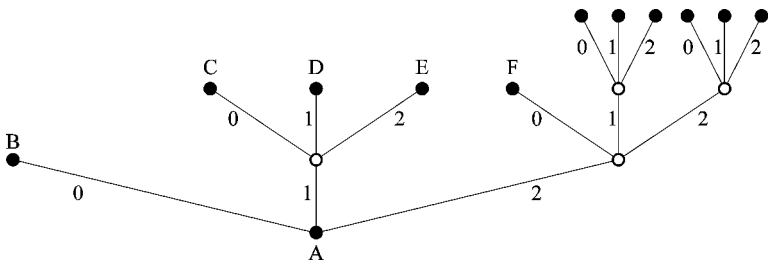

图 7－5 非满树码树图

以树形图的根作为起始点，沿着根和中间节点伸出的枝可以走到终端节点。由树形图的结构可以清楚地看出，从根到每一个终端节点所经过的路径是完全不同的。如果我们将根或中间节点所伸出的一组枝分别标注 0、1、…、m，则由根到每一个终端节点所经过的不同路径便可以用互不相同的一组有序数据(序列)来表示。

对于一组信源符号，我们可以设计一个树形图并将信源符号放置在树形图的终端节点上，从树形图的根到信源符号的路径构成的数据序列便可用作该信源符号的编码。由于从根到不同信源符号的路径不同，因此这组码字互不相同且一定满足唯一可译性，构成唯一可译码。在上述考虑中，我们没有在树形图的中间节点上安排信源符号，在由根到每一个信源符号的路径中不会遇到其他信源符号。因此任何信源符号的码字都不会成为其他信源符号码字的前缀。显然，这样构成的码字也同时满足了字首性，即为一组字首码。

因此，我们可以得到一个重要结论：在一个树形图中，只要不在任何中间节点安排信源符号，而是将所有的信源符号放在终端节点，所构成的码字就一定满足字首性。反过来讲，任何一种字首码都可以用一个信源符号全部处于终端节点的树形图来表示。

由此可知，树形图是表示和构造字首码的一种有效的方法。

如果一种编码的 m 进制 K 阶树形图是一个满树，则共有 m^K 个信源符号可以被分别安排在其 m^K 个终端节点上。此时由树形图的根到每一个信源符号的路径长度相同，均为树形图的阶数 K，所表示的是一组长为 K 的等长码。由此也可以看出，等长码满足字首性，也是一种字首码。

若一种字首码的 m 进制 K 阶树形图是一个非满树，则由根到每一个信源符号的路径

的长度可能不同。因此，若信源符号全部处于非满树的终端节点，则该树形图表示的是一组变字长的字首码。

7.3.4　字首码存在性定理

由上面的分析我们知道，工程中使用的变长码应当满足字首性。字首码存在性定理指出了一组字首码存在的条件。

定理 7.2（字首码存在性定理）　设信源符号集 $U=\{u_1, u_2, \cdots, u_n\}$，其码字长度分别为 $K_1, K_2, \cdots, K_n$ 的 m 进制的字首码存在的充分必要条件是：

$$\sum_{i=1}^{n} m^{-K_i} \leqslant 1 \tag{7.46}$$

此不等式由克拉夫特(Kraft)于 1949 年提出并加以证明，因此称为克拉夫特不等式。

码字的树形图可以直观地表示出一组字首码所满足的基本关系，下面我们结合码字树形图来证明克拉夫特不等式。

证明：取码符号集 $X=\{x_1, x_2, \cdots, x_m\}$，并对信源符号 $u_1, u_2, \cdots, u_n$ 进行 m 进制的变字长编码，设各信源符号的码字：

$$W_1, W_2, \cdots, W_n$$

分别具有码长：

$$K_1, K_2, \cdots, K_n \tag{7.47}$$

下面首先证明字首码存在的必要性。

我们构造一个 m 进制 K 阶的满树，显然该满树具有 m^K 个 K 阶节点(如图 7－4 所示的三进制三阶满树)。

假定我们所得到的一组码字 $W_1, W_2, \cdots, W_n$ 是字首码，且这组字首码可以用一个 m 进制 $K-1$ 阶的树形图加以表示，则这组码字的码长满足：

$$\max K_i \leqslant K-1$$

由字首码的结构和树形图表示时的基本特点我们知道，此时信源符号所在的节点必为该树形图的某一终端节点。

对于信源符号 u_i，已知其码字 W_i 的长度是 K_i，则该信源符号 u_i 一定位于树形图的某一个 K_i 阶的节点上。因为 W_i 是字首码，所以该 K_i 阶的节点成为一个终端节点，即此节点不能够再伸出新枝。于是，与此 K_i 阶节点相关联的阶次高于 K_i 的中间节点和终端节点便不能被使用，在 K 阶满树中的 K 阶节点个数将减少 m^{K-K_i}。

对于信源符号 $u_1, u_2, \cdots, u_n$，我们已编码得到了 n 个字首码，即由 K 阶满树中分别选出了阶数为 $K_1, K_2, \cdots, K_n$ 的 n 个节点作为终端节点。于是，在原 K 阶满树中，去掉的 K 阶节点的总数为

$$\sum_{i=1}^{n} m^{K-K_i}$$

显然，去掉的 K 阶节点的总数不会超过原 K 阶满树中 K 阶节点的总数 m_K，并满足：

$$\sum_{i=1}^{n} m^{K-K_i} \leqslant m^K$$

即

$$\sum_{i=1}^{n} m^{-K_i} \leqslant 1$$

因此对于字首码 W_1，W_2，…，W_n，其码字长度 K_1，K_2，…，K_n 一定满足克拉夫特不等式。

必要性得证。

接下来证明充分性。

假设式(7.47)表示的这组信源符号码字的码长满足克拉夫特不等式。显然，式(7.46)等价于：

$$\sum_{i=1}^{n} m^{K-K_i} \leqslant m^K$$

构造一个 m 进制 K 阶满树，其中 $K>\max K_i$。对于此满树中的 m^K 个 K 阶节点，可以按照下述方法划分为 n 组。

第 i 组($i=1$，2，…，n)中的节点满足：包含 m^{K-K_i} 个 K 阶节点且以同一个 K_i 阶的节点 A_i 作为出发点(母节点)。由树形图的结构可以看出，若以某一个 K_i 阶的节点作为母节点，则树形图的 K 阶($K>K_i$)节点中有 m^{K-K_i} 个节点是与其对应的子节点。因此，这样的划分是合理的。于是，我们可以将信源符号 u_i 放置在 K_i 阶的节点 A_i 上。按照字首码的结构要求，节点 A_i 便成为该树形图中的一个终端节点。由树形图的根到节点 A_i 即信源符号 u_i 的路径所构成的码字 W_i 满足字首性且码字长度为 K_i。

如果一组码字的长度满足克拉夫特不等式，则符合这组码长的字首码一定存在。

证毕。

在上面的定理中，我们给出了字首码存在的充分必要条件，而由证明的过程也可看出，唯一可译码存在的充分必要条件也满足克拉夫特不等式。事实上，在不改变码字长度的条件下，任何唯一可译码均可用一组字首码来代替。因此，下面我们将主要讨论字首码的构造方法。

7.4 变长信源编码定理

变长无失真信源编码是将信源符号 U 或符号序列 $U_1U_2\cdots U_L$ 变换成长度不同的码符号序列 $X_1X_2\cdots X_K$，以减小或消除信源中的冗余，提高系统的有效性。由于不同码字的长度不同，因此计算其平均码长是衡量编码效果的常用方法。然而，由上面的分析我们已经明确，在通信系统和电子信息系统的设计中，我们所追求的有效性是有条件的，即必须是在满足唯一可译性和字首性的条件下，寻找平均码长 $\overline{K}$ 最短的编码，也就是找出其紧致码。无失真变长信源编码定理给出了离散无记忆信源 U 及其扩展信源 U^L 的紧致码的平均码长 $\overline{K}$ 所达到的下界。

定理 7.3 已知离散无记忆信源 U 的熵为 $H(U)$，若使用码符号集 $X=\{x_1, x_2, \cdots, x_m\}$ 进行变字长无失真编码，则总可以找到一组字首码，其码字的平均长度满足：

$$\frac{H(U)}{\log m} \leqslant \overline{K} \leqslant 1+\frac{H(U)}{\log m} \tag{7.48}$$

定理 7.3 给出了紧致码的平均码长 $\overline{K}$ 与信源熵 $H(U)$、码符号个数 m 之间的关系。由

定理 7.3 可以看出，平均码长 $\overline{K}$ 以 $H(U)/\log m$ 为极限。当 $\overline{K}$ 小于此极限时，m 元的变字长字首码和唯一可译码不存在。同时，定理也给出了 $\overline{K}$ 的一个上界。但是由前面关于变长码的讨论我们应当明确，这并不表明在 $\overline{K}>1+\dfrac{H(U)}{\log m}$ 时不能够得到字首码，而是由于我们希望 $\overline{K}$ 尽可能地短，因此所追求的紧致码的平均码长不应超过 $1+\dfrac{H(U)}{\log m}$。

证明：设有离散无记忆信源：

$$[U,\ P]=\begin{bmatrix} u_1 & u_2 & \cdots & u_n \\ P(u_1) & P(u_2) & \cdots & P(u_n) \end{bmatrix}$$

若信源符号 u_i 使用长为 K_i 的码字表示，并使

$$-\frac{\log P(u_i)}{\log m}\leqslant K_i\leqslant 1-\frac{\log P(u_i)}{\log m}\quad i=1,2,\cdots,n \tag{7.49}$$

显然，满足式(7.49)的整数 K_i 一定存在。因为 $P(u_i)\geqslant 0$，所以若将式(7.49)的每一端都乘以 $P(u_i)$，则此不等式中的关系不变，成为

$$-P(u_i)\frac{\log P(u_i)}{\log m}\leqslant P(u_i)K_i\leqslant P(u_i)-P(u_i)\frac{\log P(u_i)}{\log m}\quad i=1,2,\cdots,n \tag{7.50}$$

将式(7.50)表示的 n 个不等式的每一端分别相加，即不等式中的每一端对 i 求和，得到：

$$\sum_{i=1}^{n}\left(-P(u_i)\frac{\log P(u_i)}{\log m}\right)\leqslant\sum_{i=1}^{n}P(u_i)K_i\leqslant\sum_{i=1}^{n}\left(P(u_i)-P(u_i)\frac{\log P(u_i)}{\log m}\right)$$

$$\frac{-1}{\log m}\sum_{i=1}^{n}P(u_i)\log P(u_i)\leqslant\sum_{i=1}^{n}P(u_i)K_i\leqslant\sum_{i=1}^{n}P(u_i)-\frac{1}{\log m}\sum_{i=1}^{n}P(u_i)\log P(u_i)$$

即

$$\frac{1}{\log m}H(U)\leqslant\overline{K}\leqslant 1+\frac{1}{\log m}H(U)$$

于是，定理 7.3 中的结论得证。

但是，由于我们要得到的码字必须是字首码，因此还需要考察式(7.49)所表示的这组码字的码长 $K_i(i=1,2,\cdots,n)$ 是否满足克拉夫特不等式(7.46)。

因为 K_i 是大于零的整数，所以只要其取值范围中的最小值能够满足克拉夫特不等式，按照式(7.49)确定码字长度就可构造出字首码。

于是，我们考察式(7.49)左端的不等关系，可知：

$$K_i\log m\geqslant-\log P(u_i)$$

此即

$$\log m^{K_i}\geqslant\log\frac{1}{P(u_i)}$$

同取以 2 为底的指数(反对数)得

$$m^{K_i}\geqslant\frac{1}{P(u_i)}$$

便可得到：

$$P(u_i)\geqslant m^{-K_i}$$

若将此不等式的两端分别对 $i=1,2,\cdots,n$ 求和，则不等式关系不变，即

$$\sum_{i=1}^{n} P(u_i) \geqslant \sum_{i=1}^{n} m^{-K_i}$$

于是

$$1 \geqslant \sum_{i=1}^{n} m^{-K_i}$$

由式(7.49)确定的一组码字的长度 K_i 满足克拉夫特不等式。所以，对于 $i=1,2,\cdots,n$，若信源符号 u_i 使用式(7.49)给出的长度为 K_i 的一组码字表示，则必可找到一种编码方式，使码字满足字首性。

证毕。

对于离散无记忆信源的 L 次扩展信源 U^L，其字首码的存在性由下述定理描述。

定理 7.4 对于由 L 个符号组成的、每个符号的熵为 $H(U)$ 的离散无记忆信源，若取码符号集为 $X=\{x_1, x_2, \cdots, x_m\}$，则一定可以找到一种无失真编码方法构成字首码，使信源 U 中每个符号所需要的码字平均长度满足：

$$\frac{H(U)}{\log m} \leqslant \frac{\overline{K}_L}{L} \leqslant \frac{1}{L} + \frac{H(U)}{\log m} \tag{7.51}$$

当 $L \to \infty$ 时，有：

$$\lim_{L\to\infty} \frac{\overline{K}_L}{L} = \frac{H(U)}{\log m} \tag{7.52}$$

此处，我们考虑的是信源符号序列的变长编码，即对其 L 次扩展信源 U^L 进行编码。定理 7.4 中，$\overline{K}_L$ 表示 L 次扩展信源 U^L 中每一个样矢量 $\boldsymbol{u}_i$ 的码字的平均码长。由于样矢量 $\boldsymbol{u}_i$ 由 L 个信源符号组成，因此对于信源 U^L 中的每一个信源符号 $\boldsymbol{u}_i$ 而言，码字的平均长度便为 $\overline{K}_L/L$。显然，最有效的编码可以达到最短的平均码长 $\overline{K}_L/L$。

证明：设有熵为 $H(U)$ 的离散无记忆信源：

$$[U, P] = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \\ P(u_1) & P(u_2) & \cdots & P(u_n) \end{bmatrix}$$

其 L 次扩展信源为

$$[U^L, P] = \begin{bmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_{n^L} \\ P(\boldsymbol{u}_1) & P(\boldsymbol{u}_2) & \cdots & P(\boldsymbol{u}_{n^L}) \end{bmatrix}$$

其中：

$$\boldsymbol{u}_i = (u_{i_1}, u_{i_2}, \cdots, u_{i_L})$$

$$P(\boldsymbol{u}_i) = \prod_{l=1}^{L} P(u_{i_l}) \quad i = 1, 2, \cdots, n^L; \quad i_1, i_2, \cdots, i_L = 1, 2, \cdots, n$$

且

$$H(U^L) = LH(U)$$

依定理 7.4 的要求，取编码符号集 $X=\{x_1, x_2, \cdots, x_m\}$ 对信源 U 和 L 次扩展信源 U^L 进行码字变长无失真编码。设对样矢量 $\boldsymbol{u}_i (i=1, 2, \cdots, n^L)$ 进行编码后，码字的平均长度为 $\overline{K}_L/L$。由定理 7.3 的结论可知，总可以找到一种编码方法使此平均码长满足式(7.48)，即

$$\frac{H(U^L)}{\log m} \leqslant \overline{K}_L \leqslant 1 + \frac{H(U^L)}{\log m}$$

$$\frac{LH(U)}{\log m} \leqslant \overline{K}_L \leqslant 1+\frac{LH(U)}{\log m}$$

将不等式的三端同除以样矢量中信源符号序列的长度 L，便有：

$$\frac{H(U)}{\log m} \leqslant \frac{\overline{K}_L}{L} \leqslant \frac{1}{L}+\frac{H(U)}{\log m}$$

即得到式(7.51)。

显然，当 $L\to\infty$ 时，有 $\lim\limits_{L\to\infty}\dfrac{\overline{K}_L}{L}\dfrac{H(U)}{\log m}$。

于是，定理 7.4 中的结论得到了证明。由于证明中使用了定理 7.3 中的结论，因此 L 次扩展信源 U^L 的字首码也必定存在。

证毕。

若将定理 7.4 的结论推广到平稳遍历的有记忆信源，如马尔可夫信源，则有：

$$\lim_{L\to\infty}\frac{\overline{K}_L}{L}=\frac{H_\infty}{\log m} \tag{7.53}$$

定理 7.3 和定理 7.4 所表示的物理意义是十分明确的，即指出了无失真信源编码时每一个信源符号平均需要的 m 进制码符号的个数与信源熵的关系。定理 7.3 和定理 7.4 指出，无失真信源编码平均码长的理论极限就是信源的熵。如果平均码长 $\overline{K}=\dfrac{\overline{K}_L}{L}<\dfrac{H(U)}{\log m}$，则不可能找到满足唯一可译性的编码方案，必定产生译码差错和失真。同时由定理 7.3 和定理 7.4 还可看出，如果是对 L 次扩展信源 U^L 进行编码，则信源符号序列愈长，平均码长愈短，当 $L\to\infty$ 时，平均码长 $\overline{K}$ 可以达到极限值。可见，减小平均码长 $\overline{K}$ 的代价是增加信源符号序列的长度，进而增加了编码处理的复杂性。

对于扩展信源的无失真变长编码，我们也可以通过计算编码效率来衡量一种字首码的有效性。根据式(7.44)，编码效率为

$$\eta=\frac{H(U^L)}{\overline{K}_L\log m}=\frac{H(U)}{\dfrac{\overline{K}_L}{L}\log m}$$

而由定理 7.4 指出的结论即式(7.51)可知：

$$\frac{\overline{K}_L}{L}\log m \leqslant \frac{\log m}{L}+H(U)$$

于是，结合式(7.44)和定理 7.4，取 $\dfrac{\overline{K}_L}{L}\log m$ 为其最大值，可以求出变长编码能够达到的编码效率的下限，即

$$\eta=\frac{H(U)}{\dfrac{\overline{K}_L}{L}\log m}>\frac{H(U)}{\dfrac{\log m}{L}+H(U)}$$

可见，L 愈大，编码效率愈高，只有信源符号序列的长度 $L\to\infty$ 时，才能使 $\eta\to 1$。因此，变长信源编码定理也是一个极限定理。

【例 7.3】　已知离散无记忆信源 $[U,\ P]=\begin{bmatrix} u_0 & u_1 \\ \dfrac{5}{6} & \dfrac{1}{6} \end{bmatrix}$，其熵为

$$H(U)=\frac{5}{6}\log\frac{6}{5}+\frac{1}{6}\log 6=\log 6-\frac{5}{6}\log 5\approx 0.650 \quad \text{比特/符号}$$

此信源输出的符号中的信息冗余为

$$\gamma=1-\frac{H(U)}{\log m}=1-\frac{0.650}{\log 2}=1-0.65=35\%$$

(1) 取码符号集 $X=\{0,1\}$ 构成信源 U 的编码：

$$u_0\to 0,\quad u_1\to 1$$

这组码字的树形图如图 7-6 所示。因为图 7-6 是一个二元一阶的满树，所以这组码字为 1 比特的等长字首码。

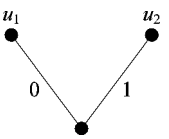

图 7-6 信源 U 的码数图

可以求出这组码字的平均码长：

$$\overline{K}_1=1 \quad \text{比特/符号}$$

编码效率：

$$\eta=\frac{H(U)}{\overline{K}_1\log m}=\frac{0.650}{1}=65\%$$

因此，编码冗余为

$$\gamma_d=1-\eta=1-0.65=35\%$$

可见，对于给定的二元信源，采用 1 比特的等长码虽然简单且易于实现，但是并没有减小信源的信息冗余，对提高系统的有效性没有帮助。

由编码定理可知，若想提高编码效率，则应当对扩展信源进行编码。

(2) 考虑 U 的二次扩展信源 U^2 并给出如下编码：

u_i	概率$\left(\frac{1}{36}\right)$	编码	码长(比特)
u_0u_0	25	0	1
u_0u_1	5	11	2
u_1u_0	5	100	3
u_1u_1	1	101	3

图 7-7 给出了这组码字的树形图。

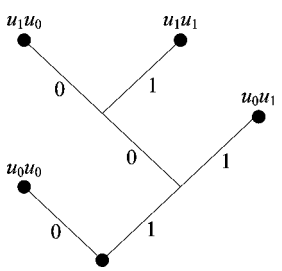

图 7-7 扩展信源 U 的码数图

图 7-7 是一个二进制三阶的非满树，由于信源符号都在树形图的终端节点上，因此这组码为字首码。

可以求出：

$$\overline{K}_2 = \frac{1}{36}[25\times 1+5\times 2+6\times 3] = \frac{53}{36} = 1.472 \quad \text{比特 / 两个符号}$$

$$\overline{K} = \frac{\overline{K}_2}{2} \approx 0.736 \quad \text{比特 / 符号}$$

编码效率为

$$\eta = \frac{H(U)}{\overline{K}\log m} = \frac{0.65}{0.736\times 1} \approx 88.3\%$$

编码冗余减小为

$$\gamma_d = 1-\eta = 11.7\%$$

(3) 对于 U 的三次扩展信源 U^3，我们给出如下编码：

u_i	概率$\left(\frac{1}{216}\right)$	编码	码长(比特)
$u_0u_0u_0$	125	0	1
$u_0u_0u_1$	25	100	3
$u_0u_1u_0$	25	101	3
$u_0u_1u_1$	5	11100	5
$u_1u_0u_0$	25	110	3
$u_1u_0u_1$	5	11101	5
$u_1u_1u_0$	5	11110	5
$u_1u_1u_1$	1	11111	5

图 7-8 为这组码字的树形图。

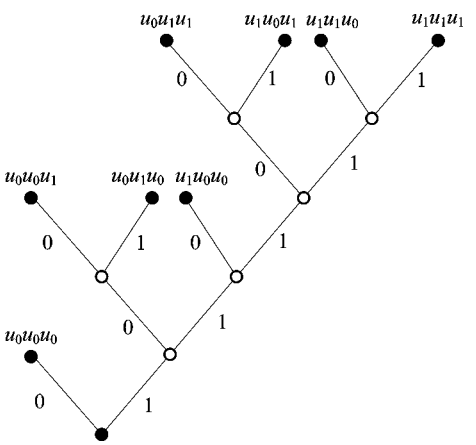

图 7-8　三次扩展信源 U 的码数图

由树形图可知，这组码是字首码。

$$\overline{K}_3 = \frac{1}{216}[125+3\times 25\times 3+5\times(5\times 3+1)] = \frac{430}{216} \approx 1.991 \quad \text{比特/三个符号}$$

$$\overline{K} = \frac{\overline{K}_3}{3} = 0.6636 \quad \text{比特/符号}$$

编码效率和编码冗余为

$$\eta = \frac{H(U)}{\overline{K}\log m} = \frac{0.65}{0.6636} \approx 97.95\%$$

$$\gamma_d = 1 - \eta = 2.05\%$$

可见，对信源 U 的扩展信源编码并增加输出信源符号序列的长度时，可以提高编码器的编码效率。事实上，如果 L 进一步增加，则编码效率 η 将随之进一步提高并趋近于 1。例 7.3 所表现出的趋势与编码定理的结论是一致的。

另一方面，与等长信源编码方法相比，在达到相同的编码效率时，变长编码所要求的信源符号序列长度要小得多。

在例 7.1 中，已知 $H(U)=2.55$ 比特/符号，仍取 $m=3$ 并且要求编码效率 $\eta>90\%$。若进行变长编码，则根据：

$$\eta > \frac{H(U)}{H(U)+\frac{1}{L}} = 0.9$$

可以求出：

$$L = \frac{1}{0.28} \approx 4$$

只要信源符号序列的长度大于 4，即可满足 $\eta>0.9$ 的要求。

此处的讨论中我们给出的是充分条件。实际上，只要选择恰当的变长编码方法，便可以在 $L=1$ 时，即仅对信源 U 进行变长编码使编码效率达到 97.7%。而由关于等长编码方法的讨论我们已经知道，为了满足 $\eta>90\%$ 的要求，需要将 $L\approx10^8$ 个信源符号进行等长编码。

所以，对于离散无记忆信源及其扩展信源的无失真编码处理，变长编码能够以较低的编码复杂度有效地提高编码效率，与等长信源编码方法相比，具有突出的优点。因此，变长编码是无失真信源编码中的常用方法。

7.5 最佳编码定理与统计编码方法

前几节我们深入讨论了无失真信源编码中的一些基本问题，特别是通过无失真信源编码定理分析了离散无记忆信源的编码所能达到的极限，为高效无失真编码方法的研究提供了重要的理论依据。

Shannon 的变长无失真信源编码定理不是一个构造性的定理，它所指出的只是紧致码的存在性，表明对于离散无记忆信源 U，一定存在一种平均码长接近 $\frac{H(U)}{\log m}$ 的编码（紧致码）。然而，这一存在性定理并没有告诉我们怎样得到这样一组紧致码，即由这一编码定理我们还不能够得到紧致码的构造方法。因此，针对不同信源和信息系统的特点与应用要求，研究紧致码或接近编码定理所指出的理论极限的高效无失真信源编码方法一直是信源编码理论和应用研究中的主要内容之一。

对于离散无记忆信源：

$$[U,\ P] = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \\ P(u_1) & P(u_2) & \cdots & P(u_n) \end{bmatrix} \tag{7.54}$$

假设我们已找到一组与其一一对应的无失真变字长信源编码，相应于各信源符号的码字：

$$W_1, W_2, \cdots, W_n \tag{7.55}$$

分别具有码长：

$$K_1, K_2, \cdots, K_n \tag{7.56}$$

如果式(7.55)所示的码字满足唯一可译性和字首性，并且式(7.56)表示一组码字的平均码长最短，则称式(7.55)给出的这组码字是信源 U 的最优码。

这一节我们将讨论构造最优码时码字长度分配和码符号分配与信源符号发生的概率之间的关系，以及最优编码处理中应当遵循的基本规则，并由此引出统计编码的一般思想和处理方法。

定理 7.5(最优码长分配规则)　在对离散无记忆信源 U 进行变字长的无失真编码时，如果使码字的长度严格地按照信源符号概率大小的相反顺序排列，即出现概率大的信源符号使用短码表示，出现概率小的信源符号使用长码表示，则其平均码长一定不大于按照任何其他符号排列方式所得编码的平均码长。

证明：对于式(7.54)所示的离散无记忆信源 U，假设式(7.55)和式(7.56)是一组按照定理 7.5 规定的码长分配规则所得到的一组编码，这组码字的平均码长为

$$\overline{K} = \sum_{i=1}^{n} K_i P(u_i) \tag{7.57}$$

其中，概率为 $P(u_i)$ 的信源符号 u_i 所对应的码字长度为 K_i。

根据定理中的要求，每一个信源符号码字的码长严格地与信源符号发生的概率成反比，即如果

$$P(u_l) \geqslant P(u_s)$$

则

$$K_l \leqslant K_s \tag{7.58}$$

其中：$l, s=1, 2, \cdots, n$ 且 $l \neq s$。

如果按照这种码长排列方法所得到的码字的平均码长最短，则由任何其他的码长排列方式所得到的平均码长都将大于式(7.57)所示的平均码长。为了证明这一论点，我们调整这组码字与信源符号的对应关系，观察其平均码长的变化情况。为使码长变化的关系简单且不失一般性，我们可以仅将信源符号 u_l 的码字与信源符号 u_s 的码字互换，而不改变其他信源符号的码字。于是信源符号 u_l 的码字长度成为 K_s，信源符号 u_s 的码字长度成为 K_l，其他信源符号的码字长度不变。

信源符号 u_l 和 u_s 的码字互换后，编码系统的平均码长改变为

$$\begin{aligned}\overline{K}' &= \overline{K} - K_l P(u_l) - K_s P(u_s) + K_s P(u_l) + K_l P(u_s) \\ &= \overline{K} + P(u_l) \times (K_s - K_l) - P(u_s) \times (K_s - K_l) \\ &= \overline{K} + (K_s - K_l) \times (P(u_l) - P(u_s))\end{aligned}$$

因为 $K_s - K_l \geqslant 0$ 且 $P(u_l) - P(u_s) \geqslant 0$，所以一定有 $\overline{K}' \geqslant \overline{K}$。

可见，对于这组给定的码字，只有使每一个信源符号的码字长度严格地与其发生的概率成反比，才能构成信源 U 的一组平均码长最短的码。

证毕。

虽然定理 7.5 仍然不是一个构造性的编码定理，但是它指出了构造最优码所遵从的一个基本关系，即在离散无记忆信源的无失真编码中，应将编码的码字长度与信源符号的概率统计关系联系起来。

按照定理 7.5 所指出的关系，对于某信源 U 或者其 L 次扩展信源 U^L，我们应当对其概率分布特性有深入的了解，然后依据各个信源符号和信源符号序列出现的概率大小，对它们分配不同长度的码字，将出现概率大的信源符号(序列)用短码表示，出现概率小的信源符号(序列)则用长码表示。由于这种编码处理的基本出发点是信源的统计特性，因此通过将码字的长度与信源符号(序列)发生的概率相匹配，可使码字长度的统计平均值达到其最小值，从而提高系统的有效性，这类编码方法称为统计编码或概率匹配编码，并成为变长无失真信源编码的基本方法。

根据定理 7.5 指出的码长分配规则，编码处理中应当首先将各信源符号按照其概率的大小重新排列。此处，我们不失一般地使信源 U 中的信源符号按其发生概率由大到小的顺序进行排列，即在式(7.54)所示的信源 U 的概率空间中，使

$$P(u_1) \geqslant P(u_2) \geqslant \cdots \geqslant P(u_n) \tag{7.59}$$

并且使式(7.55)所示的码字的码字长度(见式(7.56))满足：

$$K_1 \leqslant K_2 \leqslant \cdots \leqslant K_n \tag{7.60}$$

在这样的前提下，我们给出有关构造二元最优码的两个定理。

定理 7.6 对于给定的信源，存在一个最优的满足字首性的二元码，其最小概率的两个码字 W_{n-1} 和 W_n 的码长最长且相等，它们所构成的码符号序列之间只有最后一位码符号取值不同，即如果 W_{n-1} 的末位码符号取 0，则 W_n 的末位码符号为 1。

证明：取码符号集为 $X=\{0, 1\}$，则给定信源的二元字首码可以用一个二元的非满树来表示(如图 7-3 所示的码树图)。

按照定理 7.6 的码长分配规则，应当使具有最小概率的两个信源符号 u_{n-1} 和 u_n 所对应的码字 W_{n-1} 和 W_n 最长，同时根据字首码的结构我们知道，由树形图的根到信源符号 u_{n-1} 和 u_n 所处的终端节点的路径将只有最后一步不同。于是，信源符号 u_{n-1} 和 u_n 的码字 W_{n-1} 和 W_n 具有相同的码符号序列长度，且 W_{n-1} 和 W_n 只有最后一位码符号取值不同。如果不是这样，如假设 $K_n=K_{n-1}+1$，则 W_n 的码符号序列中的最后一位可以除去而不会影响码的唯一可译性和字首性(参见图 7-3)。

如果有三个以上概率最小的信源符号的码字长度相同，则总可以在不改变平均码长的前提下调换这些信源符号的码字，使得信源符号 u_{n-1} 和 u_n 所处的终端节点是由一个母节点延长一步而得到的两个子节点，它们的码符号序列之间只有最后一位取值不同。

于是，定理 7.6 的最优二元码中，使具有最小概率的两个信源符号的码字具有相同的编码长度，并且它们的码符号序列中只有最后一位码符号取值不同的编码一定存在。
证毕。

定理 7.6 给出了一种构造二进制最优码时应当遵循的码符号分配规则，即概率最小的两个信源符号的码符号序列的长度应当相同，且这两个码符号序列只有最后一位码符号不同，分别为 0 和 1。

为了构造平均码长最短的最优码，按照这样的码符号分配规则和概率匹配的编码原则，首先应当在给定的信源 U 中找出概率最小的两个信源符号 u_{n-1} 和 u_n，分别对 u_{n-1} 和 u_n 分配码符号 0 和 1，以确定它们的码符号序列中的最后一位码符号。

为了能够按照这样的码字分配规则继续处理，应当将信源 U 做适当的组合变形，然后继续寻找概率最小的两个符号并对其码符号序列的最后一位分配不同的码符号，直至完成

信源 U 的编码处理。

于是，我们将信源 U 的符号重新组合，构造出信源 U 的一个辅助集合：

$$[U', P] = \begin{bmatrix} u_1' & u_2' & \cdots & u_{n-1}' \\ P(u_1') & P(u_2') & \cdots & P(u_{n-1}') \end{bmatrix} \tag{7.61}$$

其中：

$$\begin{cases} u_i' = u_i \\ P(u_i') = P(u_i) \\ u_{n-1}' = u_{n-1} \bigcup u_n \\ P(u_{n-1}') = P(u_{n-1}) + P(u_n) \end{cases} \qquad i = 1, 2, \cdots, n-2 \tag{7.62}$$

对于这一辅助集合 U'，我们也将其包含的符号按照概率由大到小的顺序重新排列，并对概率最小的两个符号 u_{n-1}' 和 u_{n-2}' 的码符号序列中的最后一位分配码符号，即分别赋以 0 和 1。

如此重复，直至信源的辅助集合中只包含两个符号，分别对这两个符号的码符号序列中的最后一位赋以码符号 0 和 1 为止，就完成了信源 U 的编码处理。

但是，对于辅助集合 U' 的这种处理能否保证所得的码字一定是信源 U 的最优码呢？下面的定理给出了肯定的结论。

定理 7.7　对于辅助集合 U' 为最优的码字对原信源的符号集 U 也是最优的。

证明：对于给定的信源 U，其概率空间如式(7.54)所示。通过信源符号的组合，我们可以得到式(7.61)所示的辅助集合 U'。辅助集合 U' 中符号的概率满足式(7.62)。

对于辅助集合 U'，假设已找到一组最优码：

$$W_1', W_2', \cdots, W_{n-1}' \tag{7.63}$$

相应的码字长度为

$$K_1', K_2', \cdots, K_{n-1}'$$

按照前面讨论的编码策略，可以构造出信源 U 的一组码字：

$$W_1, W_2, \cdots, W_n$$

其码字长度为

$$\begin{aligned} K_i &= K_i' && i = 1, 2, \cdots, n-2 \\ K_i &= K_{n-1}' - 1 && i = n-1, n \end{aligned}$$

信源 U 的这组码字的平均码长为

$$\begin{aligned} \overline{K} &= \sum_{i=1}^{n} P(u_i) K_i = \sum_{i=1}^{n-2} P(u_i) K_i + P(u_{n-1}) K_{n-1} + P(u_n) K_n \\ &= \sum_{i=1}^{n-2} P(u_i') K_i' + (P(u_{n-1}) + P(u_n))(K_{n-1}' + 1) \end{aligned}$$

因为 $u_{n-1}' = u_{n-1} \bigcup u_n$，$P(u_{n-1}') = P(u_{n-1}) + P(u_n)$，所以有：

$$\begin{aligned} \overline{K} &= \sum_{i=1}^{n-1} P(u_i') K_i' + (P(u_{n-1}) + P(u_n)) \\ &= \overline{K}' + (P(u_{n-1}) + P(u_n)) \end{aligned} \tag{7.64}$$

式中，$(P(u_{n-1}) + P(u_n))$ 与辅助集合 U' 中的码字分配无关，对于给定的信源 U 它是一个常量；$\overline{K}' = \sum_{i=1}^{n-1} P(u_i') K_i'$ 是辅助集合 U' 的编码的平均码长。已知式(7.63)所示的码字是一

组最优码，其平均码长 $\overline{K}'$ 对于辅助集合 U' 而言是一个最小值。因此，式(7.64)所示的这组码字的平均码长对于信源 U 而言必定也是最短的。

证毕。

定理 7.7 给出了一个重要的结论：辅助集合 U' 的最优码也是信源 U 的最优码。于是寻求具有 n 个符号的信源 U 的最优码的过程可以转化为寻求具有 $n-1$ 个符号的辅助集合 U' 的最优码。

应用这一结论并加以推广，我们可以得出一个构造最优码的编码策略：对信源中的符号逐次地组合变形构成符号个数递减的辅助集合，对每一个辅助集合都应用概率匹配的编码思想和最优码的码符号分配规则进行处理得到辅助集合的最优码，最终构成的码字一定是信源 U 的最优码。有了这样的构造最优码的编码策略，我们就可以讨论变长无失真信源编码的基本方法了。

7.6 变长编码基本方法

7.6.1 霍夫曼编码

霍夫曼编码是一种高效的无失真信源编码方法，在各类无失真数据压缩和数据传输系统中得到了广泛的应用。

1952 年，霍夫曼(D. A. Huffman)提出了一种构造字首码的编码方法，一般称为霍夫曼码。由于这种编码方法直接应用了最优码的构造策略，所得到的码字具有最短的平均码长，因此霍夫曼码是一种最优码。

下面我们给出二进制霍夫曼码的编码处理步骤。

设有离散无记忆信源：

$$[U,\ P]=\begin{bmatrix} u_1 & u_2 & \cdots & u_n \\ P(u_1) & P(u_2) & \cdots & P(u_n) \end{bmatrix}$$

如果取码符号集 $X=\{0,\ 1\}$ 对信源符号 $u_i(i=1,\ 2,\ \cdots,\ n)$ 进行霍夫曼编码，则编码处理的基本步骤如下：

(1) 将 n 种信源符号按照概率由大到小的顺序重新排列，设

$$P(u_1)\geqslant P(u_2)\geqslant\cdots\geqslant P(u_n)$$

(2) 取出概率最小的两个符号，分别分配码符号 0 和 1，而后将这两个符号合并为一个新符号，新符号的概率是分配码符号的两个符号的概率之和。将此新符号与未分配码符号的其他符号按照概率由大到小的顺序重新排列。

(3) 重复第(2)步的处理直至合并后只有一个符号，其概率为 1。

(4) 从最后一次合并、仅包含一个符号(概率为 1)的辅助集合开始，逐级向前返回到每一个信源符号，由返回到信源符号的路径所构成的码符号序列即为信源符号所对应的变长码。

【例 7.4】 已知离散无记忆信源：

$$[U,\ P]=\begin{bmatrix} u_1 & u_2 & u_3 & u_4 & u_5 & u_6 & u_7 \\ 0.4 & 0.2 & 0.1 & 0.1 & 0.1 & 0.05 & 0.05 \end{bmatrix}$$

若取码符号集 $X=\{0,\ 1\}$，试求出其霍夫曼码，检验这组码的字首性并计算编码效率。

解：先计算此信源的熵和信源的信息冗余。

$$H(U)=-\sum_{i=1}^{7}P(u_i)\log P(u_i)$$

$$=-0.4\log0.4-0.2\log0.2-0.3\log0.1-0.1\log0.05$$

$$\approx 2.422 \quad \text{比特/符号}$$

$$\gamma=1-\frac{H(U)}{\log 7}=1-\frac{2.422}{2.807}\approx 1-0.863=0.137$$

可见，此信源输出的符号中，13.7%是不包含信息的多余部分。通过霍夫曼编码，我们可以减小这种多余的成分，提高系统的有效性。

图 7-9 给出了按照霍夫曼码的处理步骤构造信源 U 的霍夫曼码的过程。图 7-10 为这组霍夫曼码的树形图。

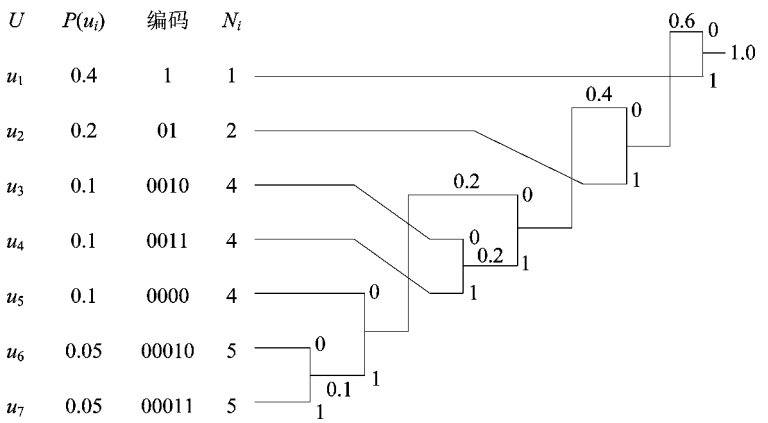

图 7-9　霍夫曼码的构造过程

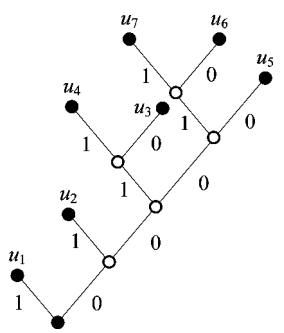

图 7-10　霍夫曼码的树形图

由这组霍夫曼码的树形图可以看出，各个信源符号都被安置在树形图的终端节点上，因此这组码字是满足字首性的字首码。

计算平均码长和编码效率：

$$\overline{K}=\sum_{i=1}^{7}P(u_i)K_i=0.4\times1+0.2\times2+0.3\times4+0.1\times5=2.5 \quad \text{比特/符号}$$

$$\eta=\frac{H(U)}{\overline{K}\log m}=\frac{2.422}{2.5}\approx 0.969=96.9\%$$

编码后仍然保留的冗余减少为

$$\gamma_d = 1 - \eta = 0.031 = 3.1\%$$

由霍夫曼码的编码步骤可以看出，霍夫曼码是直接应用最优码的编码策略而构成的。这不仅保证了码字长度的分配与信源符号的概率成反比，而且在每次合并后概率最小的两个符号只有最后一位码符号不同，使得短码得到了充分利用。因此，霍夫曼码一定是最优的紧致码。

结合无失真编码的应用，我们还应当了解霍夫曼码的一些工程特点。

(1) 按照霍夫曼编码方法所得到的信源 U 的霍夫曼码并非是唯一的。

造成霍夫曼码不唯一的原因主要有以下两个方面：

① 码符号 0、1 的分配方式是随意的。

在每一次的码符号分配处理时，对于辅助集合中概率最小的两个符号而言，无论是获得 0 还是获得 1，都符合字首码的构造规则。然而，不同的 0、1 分配结果将使它们的码符号序列中相对应的某一位码符号取值不同，构成了不同的码字。但是，这时由根到每一个信源符号所经过的路径长度是相同的，得到的不同的霍夫曼码具有相同的码字长度 K_i。

显然，由于码符号 0、1 随意分配所构成的不同霍夫曼码的平均码长 $\overline{K}$ 一定相同，因此它们是完全等价的。

② 若辅助集合中有两个以上的符号具有相同的概率，则重新排列的结果可能不同。

在由信源缩减合并所构成的辅助集合中，合并而成的新符号可能会与其他符号有相同的概率。在前面讨论的霍夫曼编码步骤中，对于这些具有相同概率的符号在重新排列时的前后位置并无严格规定。然而，不同的排列顺序将产生不同的路径，构成不同的霍夫曼码。因为排列顺序的改变会使某些路径的长短发生变化，所以不同的排列顺序所构成的霍夫曼码将具有不同的码字长度 K_i 分布关系。但是，由于这些不同的霍夫曼码都是按照最优码的构造规则得到的，因此它们都是最优码，即它们的平均码长 $\overline{K}$ 一定相同且达到其最小值。

【例 7.5】 对于例 7.4 给出的信源 U，构造出另一组霍夫曼码。此时，如果辅助集合中有两个以上符号的概率相同，则我们采用如下的排序规则：将在前一步处理中获得了码符号并被合并的新符号尽量向前提。图 7－11 给出了构造信源 U 的霍夫曼码的过程。图 7－12 为这组霍夫曼码的树形图。

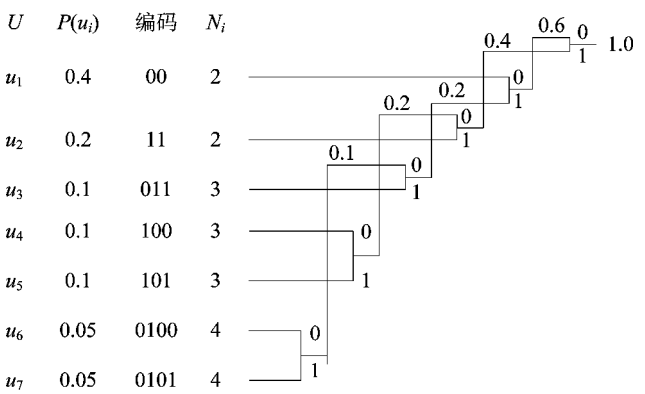

图 7－11 霍夫曼编码图

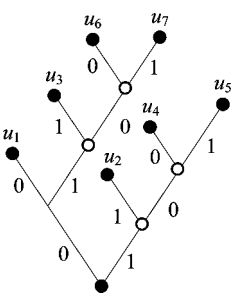

图 7－12 霍夫曼码树形图

由编码结果可以明显地看出，这组霍夫曼码与例 7.4 的霍夫曼码的码字不同，码长分布也不同。此时的平均码长为

$$\overline{K} = 0.6 \times 2 + 0.3 \times 3 + 0.1 \times 4 = 2.5 \quad \text{比特 / 符号}$$

可见，两种霍夫曼码的平均码长相同，它们都是最优码。

(2) 虽然不同的霍夫曼码有相同的平均码长 $\overline{K}$，但是它们的码长的方差可能不同。

对于给定的一组信源概率分布，虽然不同霍夫曼码的码长分布具有相同的统计平均值 $\overline{K}$，但是每一码长偏离其均值的程度可能会有很大的差别，即码字长度 K_i 的方差可能不同：

$$\sigma^2 = E[(K_i - \overline{K})^2] = \sum P(u_i)(K_i - \overline{K})^2 \tag{7.65}$$

在例 7.4 和例 7.5 中，我们得到了信源 U 的两组霍夫曼码。已求出它们的平均码长均为 2.5 比特，但是它们的方差分别为

$$\sigma_1^2 = 0.4 \times (1 - 2.5)^2 + 0.2 \times (2 - 2.5)^2 + 0.3 \times (4 - 2.5)^2 + 0.1 \times (5 - 2.5)^2 = 2.25$$

$$\sigma_2^2 = 0.6 \times (2 - 2.5)^2 + 0.3 \times (3 - 2.5)^2 + 0.1 \times (4 - 2.5)^2 = 0.45$$

在电子信息系统和通信系统的设计中，如果码字长度的变化较大，则将会增加编、译码处理的实现难度。因此，我们希望构造出的信源 U 最优码的码字长度变化要小。所以，我们通常选择码字长度的方差较小的霍夫曼码。

由前面的分析可以看出，霍夫曼编码过程中每一次符号的重新排列都会使得编码路径改变，这将使得码符号序列的长度随之改变。于是，为了得到码长方差较小的霍夫曼码，我们给出一种符号排序的操作规则，即辅助集合中的符号按照概率由大到小重新排列时，应当将由前一步处理中合并得到的新符号尽量向前排列。

这样的排序规则尽可能地减少了小概率信源符号被多次分配码符号的机会，从而使各个信源符号获取码符号的次数尽可能均匀，霍夫曼码码长的方差减小。

由前面的讨论我们已经明确，霍夫曼编码构造出了一种平均码长最短的最优码，因此霍夫曼码性能最优，是无失真信源编码中的常用方法。但是，在编码过程中，每一次码符号分配之前都需要将辅助集合中的符号按照概率进行比较和重新排序。如果信源符号集中的符号数目较大，则编码处理的复杂度将急剧增加。

为了减小编码处理的复杂度，于是出现了一些其他的变长编码方法。

7.6.2　香农-范诺编码

香农-范诺(Shannon - Fano)编码也是一种根据信源的统计关系构造字首码的一种方法。

设有离散无记忆信源：

$$[U, P] = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \\ P(u_1) & P(u_2) & \cdots & P(u_n) \end{bmatrix}$$

如果取码符号集 $X=\{0, 1\}$ 对信源符号 $u_i (i=1, 2, \cdots, n)$ 进行香农-范诺编码，则编码处理的基本步骤如下：

(1) 将 n 种信源符号按照概率由大到小的顺序重新排列。不妨设

$$P(u_1) \geqslant P(u_2) \geqslant \cdots \geqslant P(u_n)$$

（2）沿此排列顺序在信源符号集中找出一个分割点 N，将 U 划分为两个不相交的子集合 U_1 和 U_2：

$$U_1 = \begin{bmatrix} u_1 & u_2 & \cdots & u_N \\ P(u_1) & P(u_2) & \cdots & P(u_N) \end{bmatrix}$$

$$U_2 = \begin{bmatrix} u_{N+1} & u_{N+2} & \cdots & u_n \\ P(u_{N+1}) & P(u_{N+2}) & \cdots & P(u_n) \end{bmatrix}$$

要求两个子集合中各符号的概率之和相等或尽量接近，即对于分割点 N，应当使 $\sum_{i=1}^{N} P(u_i)$ 与 $\sum_{i=N+1}^{n} P(u_i)$ 相等或尽量相近。

（3）对子集合 U_1 中的信源符号和子集合 U_2 中的信源符号分别分配码符号 0 和 1。

（4）重复第(2)步和第(3)步的处理，即将 U_1 和 U_2 分别划分为更小的子集合，对子集合中的信源符号分配不同的码符号，直至每一个子集合中只有一个信源符号。

（5）每次分配的码符号所组成的码符号序列即为信源符号的变长香农-范诺编码。

【例 7.6】 已知离散无记忆信源为

$$[U, P] = \begin{bmatrix} u_1 & u_2 & u_3 & u_4 & u_5 & u_6 & u_7 & u_8 \\ 0.40 & 0.18 & 0.10 & 0.10 & 0.07 & 0.06 & 0.05 & 0.04 \end{bmatrix}$$

取码符号集 $X=\{0, 1\}$，构造其香农-范诺编码，检验码的字首性并计算编码效率。

解：按照香农-范诺编码的处理步骤，对信源 U 及其子集合进行划分并分配码符号的处理过程，所得到的各个信源符号的码字、码长关系分别如图 7-13 和图 7-14 所示。

由图 7-13 和图 7-14 可以看出，香农-范诺编码实际上是一种构造码的树形图的方法。如果将信源 U 看做树形图的根，则它的两个子集合是树形图中的两个发生概率平衡的一阶节点，通过分配码符号可以构成由根到两个一阶节点的不同路径。如果对每一个节点继续划分、分配路径，就构成了由根到每一个终端节点(信源符号)的非满树的码树图。

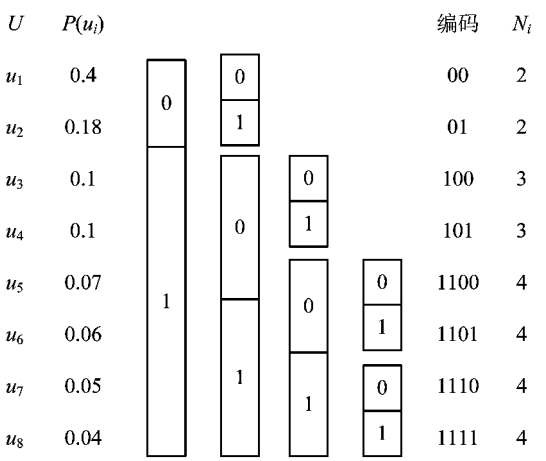

图 7-13 香农-范诺编码构造图

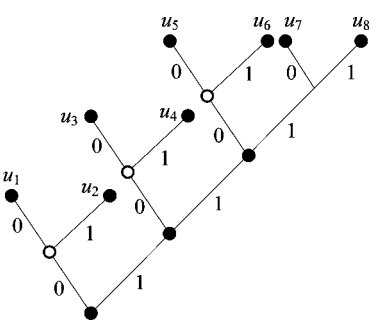

图 7-14 香农-范诺编码树形图

显然，这样构成的香农-范诺编码满足字首性，是一组字首码。

在例 7.1 中我们已求出此信源的熵为

$$H(U) \approx 2.5524 \quad \text{比特 / 符号}$$

可以求出这组香农-范诺编码的平均码长和编码效率为

$$\overline{K} = \sum_{i=1}^{8} P(u_i)K_i = 0.58 \times 2 + 0.2 \times 3 + 0.22 \times 4 = 2.64 \quad \text{比特 / 符号}$$

$$\eta = \frac{H(U)}{\overline{K}\log m} = \frac{2.5524}{2.64} \approx 0.967 = 96.7\%$$

在香农-范诺编码的过程中，第一步便将信源符号按照概率由大到小的顺序进行重新排列，此后在每一次信源的划分中要求子集合的概率尽量平衡，这些措施保证了香农-范诺码的长度与信源符号发生的概率成反比，符合统计编码的基本思想。因此香农-范诺编码也是一种高效的变长编码方法。

由于香农-范诺编码仅仅对信源符号做了一次按照概率由大到小顺序的重新排列，因此与霍夫曼编码相比其优点是降低了处理复杂度，易于软、硬件实现。然而也正是由于只有一次重新排列的处理，香农-范诺编码没有严格地遵循最优码的构造策略，因此不能够保证短码被充分利用，香农-范诺码一般不是最优码。特别是在每次划分时，如果不同子集合中信源符号的概率之和相差较大，则平均码长便会增加，使得编码效率降低。

对于例 7.6 中的信源 U，若进行霍夫曼编码，则可以得到：

$$\overline{K} = 2.61 \quad \text{比特 / 符号}$$

$$\eta = 97.8\%$$

的最优码。

所以，香农-范诺码不是最优码。只有对于某些特殊的信源概率分布(如呈 2 的负幂)，香农-范诺编码才有可能构成编码效率等于 1 的最优码。

7.6.3　算术编码

算术编码是一种非分组编码算法，由 J. Rissanen 等人提出。它从全序列出发，采用递推形式的连续编码。它不是将单个的信源符号映射成一个码字，而是将整个输入序列的符号依据它们的概率映射为实数轴上[0，1)区间内的一个小区间，再在该小区间内选择一个代表性的二进制小数，作为实际的编码输出。在假设信源为二元平稳的马尔可夫源以后，需要预先存储的不再是码字，而是由信源序列状态确定的一些参数。算术编码比霍夫曼编码、游程编码等熵编码算法都复杂，但它不需要传送像霍夫曼码表一类的编码表。算术编码具有自适应的能力，它的编码效率优于霍夫曼编码，所以算术编码是实现高效数据压缩的有力工具。

按照无失真信源编码定理，对于离散无记忆信源，霍夫曼编码的平均码长可以接近其下限，即编码符号的熵 $H(U)$。考虑达到该极限的情况，此时其平均码长为

$$\overline{K} = \sum_i P(u_i)K_i$$

而

$$H(U) = -\sum_i P(u_i)\log P(u_i)$$

显然，只有当对符号 u_i 分配的码长满足：

$$K_i = -\log P(u_i)$$

时，才能达到理想情况。但是分配给编码符号的码长 K_i 只能是整数而不能是小数，因此要求信源符号概率 $P(u_i)$ 等于 2 的负整数次幂时，霍夫曼编码才能达到最佳结果，香农-范诺编码同样如此。

从原理上讲，算术编码同霍夫曼编码类似，用较短的码表示出现概率小的信源符号，用较长的码表示出现概率大的信源符号。但是算术编码方法绕过了霍夫曼编码用整数码长的代码表示一个信源符号的想法，而是用一个浮点数表示一个信源符号流。因此，算术编码可以更加接近于无失真压缩的理论极限。

算术编码将被编码的信源符号流表示成实数半开区间[0，1)中的一个数值间隔。这个间隔随着信息流中每一个信源符号的加入逐步减小，每次减小的程度取决于当前加入的信源符号的先验概率。先验概率高，则减小的程度低，表示它只需要在原有基础上增加较少的位数；先验概率低，则减小的程度高，需要在原基础上增加较多的位数来表示它。因此，符号流越长，代表它的间隔就越小，编码表示这一间隔所需要的比特位数就越多。

下面用一个简单的例子说明算术编码的基本工作原理。

【例 7.7】 设有信源：

$$[U, P] = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ P(a_1) & P(a_2) & P(a_3) & P(a_4) \end{bmatrix}$$

已知此信源发出符号 $a_1a_2a_3a_4a_4\cdots$，试对此符号序列的前五位进行算数编码。

解：按照每个符号出现的概率对区间 [0，1)进行划分，显然每个符号都有一个对应的子区间，如表 7－1 所示。

表 7－1 算术编码的子区间划分

信源符号	概　率	初始子区间
a_1	0.2	[0，0.2)
a_2	0.2	[0.2，0.4)
a_3	0.4	[0.4，0.8)
a_4	0.2	[0.8，1.0)

对符号序列“$a_1a_2a_3a_4a_4$”进行编码的过程如图 7－15 所示。

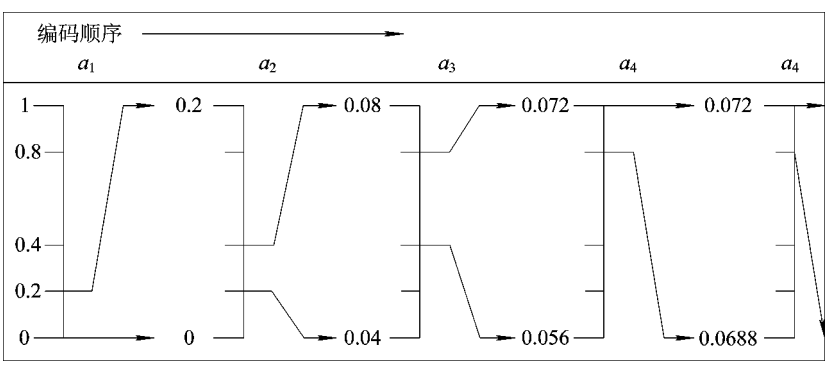

图 7－15 算术编码流程图

表 7－2 给出了每一步编码过程所对应的输出数值范围，开始假定符号占据整个半开区间[0，1)。对 a_1 进行编码，根据表 7－1 把符号串区间设置为[0，0.2)，然后将该区间扩展成和原区间[0，1)一样，端点用新的字符串区间标准。对下一个字符 a_2 进行编码，新的区间上边界为 0.2×0.2＝0.04，下边界为 0.2×0.4＝0.08，区间长度为 0.04；处理 a_3 时，新上边界为 0.04＋0.04×0.4＝0.056，下边界为 0.04＋0.04×0.8＝0.072，区间长度为 0.016；以此类推，最后处理 a_4，新上边界为 0.0688＋0.0032×0.8＝0.071 36，下边界为 0.0688＋0.0032×1.0＝0.072。

表 7－2　算术编码过程

编 码 过 程	输出数值范围
初始	[0，1)
输入 a_1 编码	[0，0.2)
输入 a_2 编码	[0.04，0.08)
输入 a_3 编码	[0.056，0.072)
输入 a_4 编码	[0.0688，0.072)
输入 a_4 编码	[0.07136，0.072)

对于解码器而言，并不需要知道编码输出最终数值范围的两个端点值，只需要知道这一范围内的任何一个数值(例如最终数值的下边界值)就可以解码，例如本例中，0.07136 就可以表示对符号序列“$a_1a_2a_3a_4a_4$”的编码。

从例 7.7 中可以看出，算术编码从全序列出发，采用递推形式的连续编码。它不是将单个的信源符号映射成一个码字，而是将整个输入序列的符号依据它们的概率映射为实数轴上[0，1)区间内的一个小区间，再在该小区间内选择一个代表性的小数，作为实际的编码输出。尤其当假设信源为二元平稳的马尔可夫信源以后，需要预先存储的不再是码字，而是由信源序列状态确定的一些参数。实际应用中，算术编码比霍夫曼编码等熵编码算法都复杂，但它不需要传送像霍夫曼码表一类的编码表。另外，算术编码具有自适应能力，即使在没有信源统计数据的情况下，只要监视一小段时间内码符号出现的频度，不管统计是平稳的还是非平稳的，编码的信息传输率总能趋近于信源的熵，其编码效率优于霍夫曼编码的效率。

7.6.4　变长码的抗信道误码能力分析

由前面的讨论可以知道，变字长统计编码使信源符号的码字长度与其发生的概率相匹配，可以有效地消除或减小信源的冗余度，是提高电子信息系统和通信系统有效性的重要手段。但是，由于有效性和可靠性之间所存在的矛盾关系，这种高效的变字长编码十分“脆弱”，因此如果信道中存在噪声，则译码结果将会产生错误。

信道中噪声的影响所产生的译码差错主要有两种形式，即译码错误和码字分离错误。

例 7.4 所得到的信源符号的霍夫曼码中，信源输出符号 u_7 的霍夫曼码为 00011。如果在数据的传输过程中由于信道中噪声的影响第五位码发生错误，使这一码符号序列成为 00010，则接收端的译码结果将成为 u_6，造成译码差错。如果信道中噪声的影响使这一码符号序列成为 00001，则在接收端进行译码处理时，由前四位码 0000 可以译码得到 u_5，而第五位的码符号 1 译码为 u_1。于是，此码符号序列的译码结果成为 u_5u_1，产生了码字分离错

误。显然，对于由多个信源符号的编码所构成的码符号序列，信道噪声的影响将会使码字的分离和译码产生严重错误。

由此可以看出，变字长的统计编码是一种高效的无失真压缩编码方法，但是变长码的抗信道误码能力较低。在变长统计编码应用系统的设计中，需要采取抗信道误码的保护措施，以实现信息的无失真传输。

习 题 7

7.1 设以 8000 样本/秒的速率抽样一语音信号，并以 $M=2^8=256$ 级对抽样均匀量化，设抽样值取各量化值的概率相等，且抽样间相互统计独立。

(1) 求每抽样的信息熵；

(2) 如以十进制数表示的 M 进制抽样序列为“126 24 53”，求以二进制数表示时该 M 进制序列的形式；

(3) M 进制抽样序列“126 24 53”所对应的二进制序列的长度是多少？

7.2 设有离散无记忆信源 $[U,P]=\begin{bmatrix} u_1 & u_2 \\ 0.1 & 0.9 \end{bmatrix}$，计算下面三种编码的平均码长 $\overline{N}$ 和编码效率。(注：若 L 次扩展信源中每矢量的平均码长为 $\overline{N}_L$，则每信源符号的平均码长为 $\overline{N}=\overline{N}_L/L$。)

(1) $u_1\to 0$，$u_2\to 1$；

(2) 作二次扩展 $\boldsymbol{u}_2$：

u_1u_1	u_1u_2	u_2u_1	u_2u_2
010	011	00	1

(3) 作三次扩展 $\boldsymbol{u}_3$：

$u_1u_1u_1$	$u_1u_1u_2$	$u_1u_2u_1$	$u_1u_2u_2$	$u_2u_1u_1$	$u_2u_1u_2$	$u_2u_2u_1$	$u_2u_2u_2$
010000	010001	01001	011	0101	000	001	1

7.3 设一信源有六个可能的输出，概率分布如表 7-3 所示，表中给出了对应的编码 A、B、C、D、E、F。

表 7-3 题 7.3 表

$P(a_i)$		A	B	C	D	E	F
a_1	$\frac{1}{2}$	000	0	0	0	0	0
a_2	$\frac{1}{4}$	001	01	10	10	10	100
a_3	$\frac{1}{16}$	010	011	110	110	1100	101
a_4	$\frac{1}{16}$	011	0111	1110	1110	1101	110
a_5	$\frac{1}{16}$	100	01111	11110	1011	1110	111
a_6	$\frac{1}{16}$	101	011111	111110	1101	1111	011

(1) 这些码中哪些是唯一可译码？

(2) 这些码中哪些是即时码？

(3) 对所有唯一可译码求其平均码长 $\overline{N}$。

(4) 利用树形图验证 A、C、E 满足字首性。

(5) 对于码 B：

① 用树形图验证它不是字首码；

② 这组码长是否满足克拉夫特不等式；

③ 若不改变每一码字的码长，那么是否存在满足字首性的编码？若存在，请找出一种？

(6) 对于码 D：

① 用树形图验证它不是字首码；

② 若不改变每一码字的码长，则是否存在字首码？为什么？

7.4　有两个信源 X 和 Y：

$$[X, P(x)] = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.20 & 0.19 & 0.18 & 0.17 & 0.15 & 0.10 & 0.01 \end{bmatrix}$$

$$[Y, P(x)] = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 \\ 0.49 & 0.14 & 0.14 & 0.07 & 0.07 & 0.04 & 0.02 & 0.02 & 0.01 \end{bmatrix}$$

(1) 分别用霍夫曼码编出二进制变长唯一可译码，并计算编码效率，画出树形图。

(2) 分别用香农-范诺码编出二进制变长唯一可译码，并计算编码效率，画出树形图。

(3) 从 X、Y 两种不同信源比较这两种编码方法的优缺点。

7.5　设一信源有 $K=3$，$P(s_k)$分别为 0.5、0.4 和 0.1。

(1) 求二进制霍夫曼编码及效率。

(2) 求二次扩展霍夫曼编码及效率。

7.6　某信源按 $P(0)=3/4$，$P(1)=1/4$ 的概率产生统计独立的二元序列。

(1) 试求 N_0，使当 $N>N_0$ 时，$P\{I(a_i)|N-H(S)\geqslant 0.05\}\leqslant 0.01$，其中 $H(S)$是信源的熵。

(2) 试求当 $N=N_0$ 时典型序列集 A_ε 中含有的信源序列个数 M_ε。

7.7　设有无记忆二元信源，其概率为 $p_1=0.005$，$p_0=0.995$。信源输出 $N=100$ 的二元序列。在长为 $N=100$ 的信源序列中只对含有 3 个或小于 3 个“1”的各信源序列构成一一对应的一组等长码。

(1) 求码字所需的最小长度；

(2) 计算等长码典型序列中的 ε 值；

(3) 考虑没有给予编码的信源序列出现的概率，该等长码引起的错误概率 p_e 是多少？若从契比雪夫不等式考虑，则 p_e 应是多少？试加以比较。

7.8　求概率分布为$\left(\frac{1}{3}, \frac{1}{5}, \frac{1}{5}, \frac{2}{15}, \frac{2}{15}\right)$的信源的二元霍夫曼码。讨论此码对于概率分布为$\left(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right)$的信源也是最佳二元码。

7.9　设信源符号集：

$$\begin{pmatrix} S \\ P \end{pmatrix} = \begin{pmatrix} s_1 & s_2 \\ 0.1 & 0.9 \end{pmatrix}$$

(1) 求 $H(S)$和信源冗余度。

(2) 设码符号为 $X=\{0, 1\}$，编出 S 的紧致码，并求 S 的紧致码的平均码长 $\bar{L}$。

(3) 把信源的 N 次无记忆扩展信源 S^N 编成紧致码，试求 $N=2, 3, 4$ 和∞时的平均码长$\left(\frac{\bar{L}_N}{N}\right)$。

7.10 设码 C 是均匀概率分布 $P=\left(\frac{1}{n}, \cdots, \frac{1}{n}\right)$的信源的二元霍夫曼码。设码 C 中各码字的码长为 l_i，又设 $n=a2^k$，而 $1\leqslant a\leqslant 2$。

(1) 证明在此等概率信源的所有即时码中，码 C 的总码长 $T=\sum_i l_i$ 最短。

(2) 证明码 C 满足 $\sum_i r^{-l_i}=1$。

(3) 证明码 C 中，对所有 i 满足 $l_i=L$ 或 $l_i=L-1$，其中 $L=\max_i l_i$。

(4) 设 u 是码长为 $L-1$ 的码字个数，v 是码长为 L 的码字个数，根据 a、k 确定 u、v 和 L。

7.11 若有一信源：

$$\begin{bmatrix} S \\ P \end{bmatrix} = \begin{bmatrix} s_1 & s_2 \\ 0.8 & 0.2 \end{bmatrix}$$

每秒钟发出 2.66 个信源符号。将此信源的输出符号送入某个二元信道中进行传输(假设信道是无噪无损的)，而信道每秒钟只传递 2 个二元符号。试问信源不通过编码能否直接与信道连接？通过适当编码能否在此信道中进行无失真传输？若能连接，试说明如何编码并说明其原因。

7.12 香农码的信源符号集 $A=\{1, 2, \cdots, q\}$，其概率分布为 $p_1, p_2, \cdots, p_q$。以下面方法对信源符号进行编码。首先将概率分布按大小顺序排列 $p_1\geqslant p_2\geqslant\cdots\geqslant p_q$，定义累积分布函数 $F_i=\sum_{k=1}^{i-1} p_k$，F_i 是所有小于 i 符号的概率和。于是，符号 i 的码字是取 F_i 的二进数的小数 l_i 位，若有尾数则进位到第 l_i 位，其中 $l_i=[\log(1/p_i)]$。

(1) 证明这样编出的码是即时码；

(2) 证明 $H(S)\leqslant L< H(S)+1$，其中 L 为平均码长；

(3) 对于概率分布为(0.5, 0.25, 0.125, 0.125)的信源进行编码，求其各码字；

(4) 由(3)中得的码是否与此信源的霍夫曼码正好完全一致？若一致，试说明这种完全一致的一般原理。

第 8 章　限失真信源编码

由第 3 章的讨论我们知道，在图 8－1 给出的简单信息传输系统中，系统的信息传输率为

$$R = I(X; Y)$$

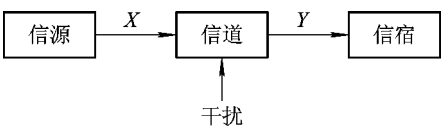

图 8－1　简单信息传输系统

由于平均互信息 $I(X; Y)$是信源概率分布 $P(x)$和信道转移概率分布 $P(y|x)$的函数，并且

$$I(X;Y) = I[P(x); P(y \mid x)]$$

是信源概率分布 $P(x)$的上凸函数，是信道转移概率分布 $P(y|x)$的下凸函数，因此如果信息传输系统中的信源不同($P(x)$不同)或者信道不同($P(y|x)$不同)，则该信息传输系统的信息传输率 R 不同。

第 5 章中我们讨论了使用固定信道传输不同信源信息的情况。由于此时信道转移概率分布 $P(y|x)$固定不变，平均互信息 $I(X; Y)$仅随着信源概率分布 $P(x)$的改变而改变，且是信源概率分布 $P(x)$的上凸函数，因此，对于给定的信道($P(y|x)$固定)，仅改变信源($P(x)$可变)时可以求出 $I(X; Y)$的极大值$\max\limits_{P(x)} I(X; Y)$，此极大值反映出了给定信道信息传输的能力，并被定义为给定信道的信道容量：

$$C = \max_{P(x)} I(X; Y)$$

与此类似，如果使用不同的信道传输同一信源($P(x)$固定)输出的信息，则系统的信息传输率将随着信道转移概率 $P(y|x)$的变化而改变。由于平均互信息 $I(X; Y)$是信道转移概率分布 $P(y|x)$的下凸函数，因此一定可以找到一个信道，当使用该信道传输给定信源输出的信息时，系统的信息传输率最小。此最小的平均互信息$\min\limits_{P(y|x)} I(X; Y)$反映出了传输该给定信源信息时必须有的最小信息传输率。显然，$\min\limits_{P(y|x)} I(X; Y)$与$\max\limits_{P(x)} I(X; Y)$是一个成对偶关系的理论问题。

对于一个给定的信源，可以使用不同的信源编码方法表示其输出的信息。从信息传输系统的有效性考虑，我们希望找出一种信源编码方法(可以看做某种信道)，采用该方法表示给定信源输出的符号时，信息传输率 R 最小，以构成一种最有效的信息传输系统。

由平均互信息 $I(X; Y)$的非负性可知：

$$\min I(X;Y)=0$$

但是当

$$I(X;Y)=0$$

即

$$H(Y)-H(Y\mid X)=0$$

$$H(Y\mid X)=H(Y)$$

时，信道的转移概率：

$$P(y\mid x)=P(y)$$

表明该信道的输出 Y 与其输入 X 相互独立，由 Y 根本得不到关于信源 X 的任何信息。由此可见，若使用

$$R=I(X;Y)=0$$

的信道传输给定信源输出信息，则信道的输出 Y 与信源的输出 X 相比面目全非，信息传输系统中产生了严重的失真，以至于完全丧失了传输信息的能力。

因此，关于信息传输系统有效性的分析，应当将信息传输率 R 与系统产生的失真联系起来，对于给定的信源($P(x)$固定)，度量 Y 与 X 之间的失真，并在一定的失真限度内寻求信息传输率 R，即平均互信息的最小值 $\min\limits_{P(y|x)} I(X;Y)$。

第 7 章我们讨论的无失真信源编码是一类在完全不允许失真(即失真度为零)的条件下的信源编码问题。根据无失真信源编码定理，信源的信息熵是进行无失真编码时编码器信息传输率的理论下界。本章讨论的信息率失真函数表示在某一失真限度下，传输给定信源信息时需要的最小信息传输率。

本课程主要针对离散无记忆信源，介绍信息率失真理论的基本概念，给出失真的度量、信息率失真函数 $R(D)$的定义和性质。下面我们将在允许一定限度内的信息损失的条件下，分析平均互信息 $I(X;Y)$的最小值问题，在最简单的条件下介绍信息率失真函数的计算，给出保真度准则下的限失真信源编码定理，初步讨论限失真信源编码的基本原理和主要方法。

8.1 失真函数与平均失真

由于人类生理、心理特性的影响，在实际的信息传递过程中，人们往往并不要求完全无失真地表示信源输出的信息，而是要求在一定保真度(失真限度)条件下，近似地还原信源输出的消息。为了讨论一定失真限度条件下的最小信息传输率，首先需要对信源、信道的输出给出一个定量的失真测度。

考虑图 8-2 所示的信息传输系统。

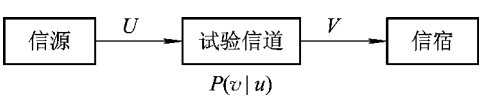

图 8-2 简单信息传输系统

设离散无记忆信源：

$$[U,\ P]=\begin{bmatrix} u_1 & u_2 & \cdots & u_r \\ P(u_1) & P(u_2) & \cdots & P(u_r) \end{bmatrix}$$

经信道 $P(v_j|u_i)$ 传输后，接收符号集为 $V=[v_1,\ v_2,\ \cdots,\ v_s]$。

1. 单个符号的失真函数

对于信道的每一对输入 u_i 和输出取值 v_j，指定一个非负数：

$$d(u_i,\ v_j)\geqslant 0 \quad i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots,\ s \tag{8.1}$$

表示 u_i 与 v_j 之间的误差或失真，被称为单个符号的失真函数。

显然，若 $v_j=u_i$，则用 v_j 表示 u_i 时便没有产生失真，于是，$d(u_i,\ v_j)=0$；若 $v_j\neq u_i$，则 $d(u_i,\ v_j)>0$。失真函数定量地描述了用 v_j 表示 u_i 时引起的失真、损失、风险和主观感觉上的差异大小，其具体的度量方式应视应用场合而定。

常用的失真函数有以下几种。

(1) 均方失真：

$$d(u_i,\ v_j)=(u_i-v_j)^2$$

(2) 绝对失真：

$$d(u_i,\ v_j)=|\ u_i-v_j\ |$$

(3) 相对失真：

$$d(u_i,\ v_j)=\frac{|\ u_i-v_j\ |}{|\ u_i\ |}$$

(4) 汉明失真：

$$d(u_i,\ v_j)=\delta(u_i-v_j)=\begin{cases}0 & u_i=v_j \\ a & u_i\neq v_j\end{cases}\qquad (a>0)$$

2. 失真函数的表示

由于信源 U 有 r 种可能取值，而接收符号集 V 有 s 种可能取值，因此，对于不同的 i、j 取值，U、V 之间共有 $r\times s$ 个失真函数 $d(u_i,\ v_j)$。它们可以排列成一个 $r\times s$ 的矩阵，即

$$\boldsymbol{D}=\begin{bmatrix} d(u_1,\ v_1) & d(u_1,\ v_2) & \cdots & d(u_1,\ v_s) \\ d(u_2,\ v_1) & d(u_2,\ v_2) & \cdots & d(u_2,\ v_s) \\ \vdots & \vdots & & \vdots \\ d(u_r,\ v_1) & d(u_r,\ v_2) & \cdots & d(u_r,\ v_s) \end{bmatrix}$$

此矩阵称为失真函数矩阵。

3. K 维随机矢量的失真函数

若信道的输入与输出是 K 维随机矢量，则可以将单个符号的失真函数的定义加以推广。

设信源输出的符号序列 $\boldsymbol{U}=(U_1,\ U_2,\ \cdots,\ U_K)$，其中每一随机变量 $U_i(i=1,2,\cdots,K)$ 取自同一符号集 $[u_1,\ u_2,\ \cdots,\ u_r]$，所以信源输出的符号序列 $\boldsymbol{U}$ 共有 r^K 种可能的取值。接收端的符号序列为 $\boldsymbol{V}=(V_1,\ V_2,\ \cdots,\ V_K)$，其中每一个随机变量 $V_j(j=1,\ 2,\ \cdots,\ K)$ 取自同一符号集 $[v_1,\ v_2,\ \cdots,\ v_s]$，因此，$\boldsymbol{V}$ 共有 s^K 种可能的取值。

又设 $\boldsymbol{u}_l=(u_{l_1}, u_{l_2}, \cdots, u_{l_K})$，$\boldsymbol{v}_m=(v_{m_1}, v_{m_2}, \cdots, v_{m_K})$分别是 $\boldsymbol{U}$、$\boldsymbol{V}$ 的样矢量，则 $\boldsymbol{u}_l$ 与 $\boldsymbol{v}_m$ 之间的失真函数为

$$d_K(\boldsymbol{u}_l, \boldsymbol{v}_m)=\frac{1}{K}\sum_{i=1}^{K} d(u_{l_i}, v_{m_i}) \quad l=1,2,\cdots,r^K;\ m=1,2,\cdots,s^K \tag{8.2}$$

对于不同的输入、输出符号序列，共有 $r^K\times s^K$ 个失真函数，可写成矩阵形式 $\boldsymbol{D}_K$，称之为信道输入、输出为 K 维随机矢量时的失真函数矩阵，它是一个 $r^K\times s^K$ 阶矩阵。

【例 8.1】 已知 U、$V\in\{0, 1\}$，失真函数定义为 $\boldsymbol{D}=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}$。

若作二次扩展，即 $\boldsymbol{U}$，$\boldsymbol{V}\in\{00, 01, 10, 11\}$，则二维随机矢量 $\boldsymbol{u}$、$\boldsymbol{v}$ 之间的失真函数为

$$d_2(00, 00)=\frac{1}{2[d(0, 0)+d(0, 0)]}=0$$

$$d_2(00, 01)=\frac{1}{2[d(0, 0)+d(0, 1)]}=\frac{1}{2}$$

$$\vdots$$

$$d_2(11, 11)=\frac{1}{2[d(1, 1)+d(1, 1)]}=0$$

依次可写出 $\boldsymbol{u}$、$\boldsymbol{v}$ 之间的失真函数矩阵：

$$\boldsymbol{D}_2=\begin{bmatrix}0 & \frac{1}{2} & \frac{1}{2} & 1\\ \frac{1}{2} & 0 & 1 & \frac{1}{2}\\ \frac{1}{2} & 1 & 0 & \frac{1}{2}\\ 1 & \frac{1}{2} & \frac{1}{2} & 0\end{bmatrix}=\frac{1}{2}\begin{bmatrix}0 & 1 & 1 & 2\\ 1 & 0 & 2 & 1\\ 1 & 2 & 0 & 1\\ 2 & 1 & 1 & 0\end{bmatrix}$$

4. 平均失真

由失真函数 $d(u_i, v_j)$的定义可以看出，对于不同的 u_i、v_j，$d(u_i, v_j)$的值可能不同。因此，失真函数 $d(u_i, v_j)$是一个伴随着随机变量 u_i 和 v_j 的发生而发生的随机变量，发生的概率为 $P(u_i, v_j)$。于是，在信息传输系统中产生的失真可以使用失真函数的统计平均值来表示。

平均失真定义为

$$\bar{d}=E[d(u_i, v_j)]=\sum_{i=1}^{r}\sum_{j=1}^{s}P(u_i, v_j)d(u_i, v_j) \tag{8.3}$$

若已知信道转移概率 $P(v_j|u_i)$，则平均失真 $\bar{d}$ 可进一步表示为

$$\bar{d}=\sum_{i=1}^{r}\sum_{j=1}^{s}P(u_i)P(v_j\mid u_i)d(u_i, v_j) \tag{8.4}$$

可见，单个符号的失真函数 $d(u_i, v_j)$描述了某个信源符号经过信道传输后的失真大小，对于不同的信源符号和不同的接收符号，其值是不同的，而平均失真则是从总体上描述整个系统失真情况的一个量。

若信道的输入、输出为 K 维随机矢量 $\boldsymbol{U}$、$\boldsymbol{V}$，则系统的平均失真定义为

$$\overline{d}_K = E[d_K(\boldsymbol{u}, \boldsymbol{v})] = \sum_{l=1}^{r^K}\sum_{m=1}^{m^K} P(\boldsymbol{u}_l, \boldsymbol{v}_m) d_K(\boldsymbol{u}_l, \boldsymbol{v}_m)$$

$$= \sum\sum P(\boldsymbol{u}_l, \boldsymbol{v}_m) P(\boldsymbol{v}_{\mathrm{m}} \mid \boldsymbol{u}_l) d_K(\boldsymbol{u}_l, \boldsymbol{v}_m) \tag{8.5}$$

也可写成：

$$\overline{d}_K = \sum_{i=1}^{r^K}\sum_{j=1}^{s^K} P(\boldsymbol{u}_l) P(\boldsymbol{v}_m \mid \boldsymbol{u}_l) d_K(\boldsymbol{u}_l, \boldsymbol{v}_m)$$

$$= \frac{1}{K}\sum_{i=1}^{r^K}\sum_{j=1}^{s^K} P(\boldsymbol{u}_l) P(\boldsymbol{v}_m \mid \boldsymbol{u}_l) \sum_{l=1}^{K} d_K(u_{l_i}, v_{m_i}) \tag{8.6}$$

当信源与信道都是无记忆时，可进一步简化为

$$\overline{d}_K = \frac{1}{K}\sum_{i=1}^{K} E[d(u_{l_i}, v_{m_i})] = \frac{1}{K}\sum_{i=1}^{K} \overline{d}_i \tag{8.7}$$

其中，$\overline{d}_i$ 是 K 维随机矢量的第 i 个分量的平均失真。

如果离散信源是平稳的，则有 $\overline{d}_i = \overline{d}$ 和 $\overline{d}_K = \overline{d}$，即通过离散无记忆信道传输离散无记忆平稳信源输出的符号序列时，系统的平均失真等于信源输出单个符号的平均失真 $\overline{d}$。

8.2　信息率失真函数及其性质

8.2.1　D 失真许可准则与 D 失真许可信道

在实际的通信过程中，人们对于通信质量的要求一般是从统计平均的意义上进行考查和度量的。为了满足一定的通信质量要求，需要对系统中产生的失真加以限制，要求信息传输系统所产生的平均失真 $\overline{d}$ 不能够大于某一给定的常数 D，即满足：

$$\overline{d} \leqslant D \tag{8.8}$$

由于常数 D 为信息传输系统中所允许的失真限度，因此式(8.8)称为信息传输系统的 D 失真许可准则。

考虑一个使用不同信道 $P(v_j \mid u_i)$ 传输给定的信源 U 输出符号的信息传输过程，信道的输出符号为随机变量 V。显然，使用不同信道传输该给定信源信息时，系统的信息传输率 $R = I(U; V)$ 不同，产生的失真也不同。

在定义了失真函数 $d(u_i, v_j)$ 之后，我们总是希望在允许一定失真限度 D 的条件下，使得传输信源信息的信息传输率 $R = I(U; V)$ 尽可能地小。也就是说，在 D 失真许可准则 $(\overline{d} \leqslant D)$ 下，寻找传输该给定信源信息的信息传输率 $R = I(U; V)$ 的最小值。

由平均失真的定义式(8.4)可以看出，系统的平均失真 $\overline{d}$ 不仅与失真函数 $d(u_i, v_j)$ 有关，而且与信源的统计特性 $P(u_i)$ 和信道的转移概率 $P(v_j \mid u_i)$ 有关。应当注意，在给定信源 U，并且定义了 U、V 之间的失真函数 $d(u_i, v_j)$ 之后，系统的平均失真 $\overline{d}$ 仅由信道转移概率 $P(v_j \mid u_i)$ 决定，即 $\overline{d} = \overline{d}[P(v_j \mid u_i)]$。

如果要求信息传输系统所产生的平均失真 $\overline{d}$ 不能够大于某一给定的常数 D，则可以使用的信道受到了限制。于是，传输该给定信源信息的信道可以划分为两类：一类信道能够满足 D 失真许可准则的要求，即满足 $\overline{d} \leqslant D$；另一类信道中所产生的平均失真 $\overline{d} > D$。因

此，我们将信道 $P(v_j|u_i)$ 划分为两个集合：

$$P_D = \{P(v_j \mid u_i);\ \overline{d} \leqslant D\} \tag{8.9}$$

$$P_{\overline{D}} = \{P(v_j \mid u_i);\ \overline{d} > D\} \tag{8.10}$$

显然，如果信道 $P(v_j|u_i) \in P_D$，则使用该信道传输给定信源输出信息时，传输系统的平均失真将满足 D 失真许可准则。因此，集合 P_D 中的信道称为 D 失真许可信道。

说明：为了求出信道可变条件下信息传输率 $R=I(U;V)$ 的最小值，此处引用的信道 $P(v_j|u_i)$ 为一种假想的可变信道，称为试验信道。在实际问题的分析中，这些不同的试验信道可以看做不同的信源编码方法。

8.2.2 信息率失真函数

平均互信息的凸性表明，对于给定的信源，信息传输系统的信息传输率：

$$R = I(U;V) = I[P(v_j \mid u_i)]$$

是信道转移概率 $P(v_j|u_i)$ 的下凸函数。因此，若改变信道参数，则一定存在一个平均互信息的最小值，使得传输给定信源信息时系统的信息传输率最小。

但是，此时讨论的问题是在满足 D 失真许可要求下寻找重建给定信源信息时所必须有的最小平均信息量。虽然由平均互信息的下凸性知道此时 $R=I(U;V)$ 的最小值一定存在，但是由于规定了失真限度 D，因此传输给定信源信息所能够使用的信道受到了限制，即 $P(v_j|u_i) \in P_D$。

于是，我们可以在 D 失真许可信道集合 P_D 中选出一组信道转移概率 $P(v_j|u_i)$，使得信息传输率 $R=I(U;V)$ 达到最小值。此平均互信息的最小值表示在满足 D 失真许可要求 $(\overline{d} \leqslant D)$ 时传输给定信源信息的信息传输率 R 的最小值。

定义满足 D 失真许可要求的平均互信息的最小值为信息率失真函数，即

$$R(D) = \min_{P(v|u) \in P_D} \{I(U;V)\} = \min\{I(U;V);\ \overline{d} \leqslant D\} \tag{8.11}$$

简称率失真函数，单位是比特/符号或奈特/符号。

信息率失真函数 $R(D)$ 的定义表明，$R(D)$ 为给定信源和定义失真函数的前提下，在失真限度为 D（试验信道满足一定约束条件）时平均互信息 $I(U;V)$ 的最小值。因此信息率失真函数 $R(D)$ 具有明确的物理含义和实际指导意义。

信息率失真函数 $R(D)$ 与信道容量 C 具有对偶性。$R(D)$ 是在给定信源 $P(u)$ 和失真限度 D 时平均互信息 $I(U;V)$ 的最小值。这个最小值 $R(D)$ 是在给定信源的情况下，接收端以满足失真限度 D 的要求而重建信源信息时所必须获得的最小平均信息量。因此，$R(D)$ 反映了在满足一定失真限度 D 的要求下，给定信源输出的信息率可以压缩的程度，即最有效地表达信源输出信息时可以压缩到的最低信息率。显然，$R(D)$ 所表示的最低信息率是反映信源特性的参量，与试验信道无关。对于不同的信源，其 $R(D)$ 不同。信道容量 C 则是在给定信道 $P(v|u)$ 时平均互信息 $I(U;V)$ 的最大值。信道容量 C 反映了信道传输信息的能力，是在信道中实现可靠传输时该给定信道的最大信息传输率。信道容量 C 与信源无关，是反映信道特性的参量。对于不同的信道，其信道容量不同。

信息率失真函数 $R(D)$ 与信道容量 C 这两个概念在实际应用中有明显区别。

信道容量 C 的研究目的是掌握给定信道传输信息的能力，即该信道能够可靠传送的最

大信息量。在充分利用给定信道传输能力的条件下，使信息传输系统的信息传输率最大而错误概率任意小，这就是一般的信道编码问题。

信息率失真函数 $R(D)$的研究目的是掌握给定信源输出信息的特性。在满足 D 失真许可要求的条件下，人们希望使用尽可能少的码符号表达(传送)给定信源输出的信息。由于 $R(D)$恰为失真限度 D 之内传输给定信源输出信息时所必需的最小信息传输率，因此信息率失真函数 $R(D)$反映了信息传输系统的有效性研究，即信源编码和数据压缩应用中所面临的一个基本问题。

8.2.3　$R(D)$的基本函数关系

由上面的讨论可以看出，传输给定信源输出的符号时，如果信息传输系统中允许产生的平均失真限度为 D，则由 D 失真许可信道集合 P_D 中找出的平均互信息的最小值便为信息率失真函数 $R(D)$。显然，如果要求的平均失真限度 D 不同，则 D 失真许可信道集合 P_D 不同，所得到的最小信息传输率也不同。因此，信息率失真函数 $R(D)$是失真限度 D 的函数。

设已给定离散无记忆信源：

$$[U,\ P(u)]=\begin{bmatrix} u_1 & u_2 & \cdots & u_r \\ P(u_1) & P(u_2) & \cdots & P(u_r) \end{bmatrix} \tag{8.12}$$

对于试验信道：

$$P(v_j \mid u_i) \quad i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots,\ s \tag{8.13}$$

已定义 U、V 之间的 $r\times s$ 个失真函数 $d(u_i,\ v_j)\geqslant 0$。失真函数矩阵为

$$\boldsymbol{D}=\begin{bmatrix} d(u_1,\ v_1) & d(u_1,\ v_2) & \cdots & d(u_1,\ v_s) \\ d(u_2,\ v_1) & d(u_2,\ v_2) & \cdots & d(u_2,\ v_s) \\ \vdots & \vdots & & \vdots \\ d(u_r,\ v_1) & d(u_r,\ v_2) & \cdots & d(u_r,\ v_s) \end{bmatrix} \tag{8.14}$$

下面首先讨论信息率失真函数：

$$R(D)=\min_{P(v\mid u)\in P_D}\{I(U;\ V)\}=\min\{I(U;\ V);\ \overline{d}\leqslant D\}$$

所具有的基本函数关系，即考察 $R(D)$的值域和使得 $R(D)$有定义的失真限度 D 的取值范围。

1. $R(D)$的值域

信息率失真函数 $R(D)$的定义式(8.11)表明，$R(D)$本质上反映的是信息传输系统中信源 U 与信道输出 V 之间的平均互信息 $I(U;\ V)$。依据平均互信息 $I(U;\ V)$的基本性质，可以确定信息率失真函数 $R(D)$取值大小满足的关系。

由于平均互信息 $I(U;\ V)$具有非负性，即

$$I(U;\ V)\geqslant 0$$

因此信息率失真函数的取值不会小于 0，即

$$R(D)\geqslant 0$$

由平均互信息 $I(U;\ V)$的极值性，即

$$I(U;\ V)\leqslant H(U)$$

可以看出，信息率失真函数 $R(D)$ 的取值有界：

$$R(D) \leqslant H(U)$$

即信息率失真函数 $R(D)$ 的数值不会超过信源 U 的输出的平均信息量 $H(U)$。于是，信息率失真函数 $R(D)$ 的取值范围为

$$0 \leqslant R(D) \leqslant H(U)$$

2. $R(D)$ **的定义域的下界** $D_{\min}$

对于给定的信源（$P(u)$ 固定）和定义的失真函数 $d(u_i, v_j)$，使得信息率失真函数 $R(D)$ 有定义的失真限度 D 中的最小值 $D_{\min}$ 便为 $R(D)$ 的定义域的下界。

由前面关于失真函数的讨论可知，失真函数大于或等于零，即

$$d(u_i, v_j) \geqslant 0 \quad i = 1, 2, \cdots, r; \ j = 1, 2, \cdots, s$$

作为失真函数 $d(u_i, v_j)$ 的统计平均值，平均失真 $\overline{d}$ 也是一个非负的实数，即 $\overline{d} \geqslant 0$。因此，使得信息率失真函数 $R(D)$ 有意义的失真限度 D 的取值不可能小于零。由此可知，信息率失真函数 $R(D)$ 的定义域的下界满足 $D_{\min} \geqslant 0$。

在实际应用中，由于不同信源具有不同的统计关系，在不同的应用场合用 v_j 表示 u_i 时所产生的差异、损失与风险的大小也各不相同，因此，信息率失真函数 $R(D)$ 的定义域需要根据具体问题和应用要求加以确定。

对于以给定的信源 U（见式(8.12)）和失真函数矩阵 $\boldsymbol{D}$（见式(8.14)），可以求出 U 与 V 之间的平均失真：

$$\overline{d} = \sum_{i=1}^{r} \sum_{j=1}^{s} P(u_i) P(v_j \mid u_i) d(u_i, v_j) \tag{8.15}$$

由于信源概率分布 $P(u_i)$ 和失真函数 $d(u_i, v_j)$ 已确定，因此此时的平均失真 $\overline{d}$ 是信道参数 $P(v_j \mid u_i)$ 的函数。于是，改变试验信道参数时所取得的平均失真的最小值即为 $R(D)$ 定义域的下界，即

$$D_{\min} = \min_{P(v|u) \in P_D} \overline{d} = \min_{P(v|u) \in P_D} \sum_{i=1}^{r} \sum_{j=1}^{s} P(u_i) P(v_j \mid u_i) d(u_i, v_j) \tag{8.16}$$

对于信源输出符号集合中的某一符号 $u_i (i=1, 2, \cdots, r$，且为固定的参变量），不同试验信道 $P(v_j \mid u_i)$ 的输出 $v_j (j=1, 2, \cdots, s)$ 不同，则该试验信道传输某一信源符号 u_i 所产生的平均失真为

$$\sum_{j=1}^{s} P(v_j \mid u_i) d(u_i, v_j) \tag{8.17}$$

显然，使用的试验信道 $P(v_j \mid u_i)$ 不同，则平均失真的大小不同，并且一定存在一种试验信道 $P(v_j \mid u_i)$，使得式(8.17)所表示的平均失真达到其可能的最小值。

进一步分析可知，对于给定的信源和已定义的失真函数，如果可以找到一种试验信道，对于 $u_i (i=1, 2, \cdots, r)$ 都能够使式(8.16)达到其可能的最小值，则 U、V 之间的平均失真一定达到其最小值，且为信息率失真函数 $R(D)$ 的定义域的下界 $D_{\min}$。于是，式(8.16)可以表示为

$$D_{\min} = \sum_{i=1}^{r} P(u_i) \min_{P(v|u) \in P_D} \left[\sum_{j=1}^{s} P(v_j \mid u_i) d(u_i, v_j) \right] \tag{8.18}$$

由此可以看出，我们可以通过选择一种试验信道，使每一种信源符号所产生的平均失

真最小(即式(8.18)中的内和式达到其最小值)，从而得到 $R(D)$ 的定义域的下界 $D_{\min}$。

使用不同的 $v_j(j=1, 2, \cdots, s)$ 表示固定的 u_i 时，由于失真函数矩阵 $\boldsymbol{D}$ 中第 i 行的元素 $d(u_i, v_j)$ 互不相同，因此在失真函数矩阵 $\boldsymbol{D}$ 的第 i 行元素中一定可以找出一个最小值(也可能有若干个相同的最小值)。不妨设失真函数矩阵 $\boldsymbol{D}$ 的第 i 行中第 J 列失真函数的取值(此处记做 J_i)最小，便有：

$$\min_j d(u_i, v_j) = d(u_i, v_{J_i}) \quad i = 1, 2, \cdots, r$$

可以选择一种试验信道，使其满足：

$$\begin{cases} P(v_j \mid u_i) = 1 & j = J_i \\ P(v_j \mid u_i) = 0 & j \neq J_i \end{cases} \quad (j = 1, 2, \cdots, s) \tag{8.19}$$

那么，第 i 行的平均失真达到其最小值：

$$\sum_{j=1}^{s} P(v_j \mid u_i) d(u_i, v_j) = d(u_i, v_{J_i}) = \min_j d(u_i, v_j) \quad i = 1, 2, \cdots, r$$

由于失真函数矩阵 $\boldsymbol{D}$ 中每一行的平均失真都是最小的，因此系统总体的平均失真一定为其可能取值的最小值。于是，对于给定的信源和定义的失真函数，信息率失真函数 $R(D)$ 的定义域的下界为

$$D_{\min} = \sum_{i=1}^{r} P(u_i) \min[d(u_i, v_j)] \tag{8.20}$$

若失真函数矩阵 $\boldsymbol{D}$ 的每一行中至少有一个失真函数为零，则有 $D_{\min}=0$，否则，$D_{\min}\neq 0$。若给定失真限度为

$$D = D_{\min} = 0$$

则表示该信息传输系统必须无失真传输给定信源输出的符号。于是，信息传输率至少应等于信源的信息熵，即

$$R(0) = H(U)$$

如果失真函数矩阵 $\boldsymbol{D}$ 中某些列中包含了至少两个取值为零的元素，则表明给定信源符号集中的某些符号可以压缩、合并。此时，一般满足：

$$R(0) < H(U)$$

若失真函数矩阵 $\boldsymbol{D}$ 中至少有一行全部为非零元素，则有 $D_{\min}\neq 0$。此时

$$R(D_{\min}) \leqslant H(U)$$

【例 8.2】 设信源 $[U, P]=\begin{bmatrix} u_1 & u_2 \\ p & 1-p \end{bmatrix}(0\leqslant p\leqslant 1)$，$V\in[v_1, v_2, v_3]$，定义失真函数矩阵为

$$\boldsymbol{D} = \begin{bmatrix} 0 & 1 & \dfrac{1}{2} \\ 1 & 0 & \dfrac{1}{3} \end{bmatrix}$$

由式(8.20)可以求出信息率失真函数 $R(D)$ 的定义域的下界为

$$\begin{aligned} D_{\min} &= \sum_{i=1}^{2} P(u_i) \min_j [d(u_i, v_j)] \\ &= p \cdot \min\left[0, 1, \frac{1}{2}\right] + (1-p) \cdot \min\left[1, 0, \frac{1}{3}\right] = 0 \end{aligned}$$

由前面的分析和式(8.19)所示的关系可以看出，达到此最小允许失真限度的试验信道为

$$[P(v_j \mid u_i)] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

由于此试验信道是一个无噪无损信道，在信道的输出端观测到输出 V 之后，便可以消除关于信源 U 的不确定性，因此信道疑义度 $H(U|V)=0$，则有：

$$I(U;V) = H(U) - H(U \mid V) = H(U)$$

于是

$$R(D_{\min}) = R(0) = \min_{p_{ji} \in P_D} \{I(U;V)\} = H(U)$$

【例 8.3】 设信源 $[U, P]=\begin{bmatrix} u_1 & u_2 & u_3 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$，$V \in [v_1, v_2]$，定义失真函数矩阵为

$$\boldsymbol{D} = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \\ 1 & 0 \end{bmatrix}$$

由式(8.20)计算得：

$$D_{\min} = \sum_{i=1}^{3} P(u_i) \min_j [d(u_i, v_j)] = \frac{1}{3} \cdot 0 + \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot 0 = \frac{1}{6}$$

所以，对于此信源 U 和定义的失真函数，$R(D)$的定义域的下界为 $D_{\min}=\frac{1}{6}$。

根据式(8.19)，达到此最小允许失真限度的试验信道必须满足：

$$\begin{cases} P(v_1 \mid u_1) = 1 \\ P(v_1 \mid u_2) + P(v_2 \mid u_2) = 1 \\ P(v_2 \mid u_3) = 1 \end{cases}$$

由于满足约束条件 $P(v_1|u_2)+P(v_2|u_2)=1$ 的 $P(v_1|u_2)$和 $P(v_2|u_2)$可以有无穷多种组合，因此可以使得系统平均失真 $\overline{d}=D_{\min}=\frac{1}{6}$的试验信道有无数多个。这些试验信道的信道矩阵中每一列都有不止一个非零元素，所以 $H(U|V)\neq 0$，于是

$$R(D_{\min}) = R\left(\frac{1}{6}\right) = \min_{p_{ji} \in P_D} (I(U;V)) < H(U)$$

若将失真函数矩阵改成：

$$\boldsymbol{d}' = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}$$

则

$$D'_{\min} = \sum_{i=1}^{3} P(u_i) \min_j [d(u_i, v_j)] = 0$$

同理，达到此最小允许失真限度的试验信道必须满足：

$$\begin{cases} P(v_1 \mid u_1) = 1 \\ P(v_1 \mid u_2) + P(v_2 \mid u_2) = 1 \\ P(v_2 \mid u_3) = 1 \end{cases}$$

满足上述条件的试验信道的共同特点就是信道矩阵中每列有不止一个非零元素，所以，$H(U|V)\neq 0$，因此，$R(D'_{\min})=R(0)<H(U)$。

从失真函数矩阵 d' 也可以看出，信源符号 u_1 传递到符号 v_1 无失真，而信源符号 u_2 传递到 v_1 也无失真，因此，完全可以将信源的符号 u_1 和 u_2 合并成一个符号，使信源符号集由三个符号压缩成两个符号，并不引起任何失真。显然，压缩后信息传输率必然减小，所以得 $R(0)<H(U)$。

3. $R(D)$的定义域的上界 $D_{\max}$

由信息率失真函数的值域可知，$R(D)$在其定义域内的取值非负，只要失真限度 $D\geqslant 0$，则 $R(D)\geqslant 0$。为了能够更加明确地反映出信息率失真函数的函数变化关系和工程意义，此处需要进一步分析 $R(D)$随着失真限度 D 改变时其函数关系的特点。

当使用可变试验信道传输给定信源 U 时，如果信道不同，则不仅信息传输率 $R=I(U;V)$ 不同，V、U 之间的平均失真 $\overline{d}$ 也不同。由前面的讨论可以看出，随着系统平均失真 $\overline{d}$ 的增大，信息传输率 $R=I(U;V)$将单调减小，其极限情况为 $R=I(U;V)=0$。

由于 $I(U;V)=0$ 意味着信道的输出 V 与信源的输出 U 相互独立，因此接收者不能够由该信息传输系统得到信源输出的信息。然而对于一个实际的信息传输系统来说，在平均失真 $\overline{d}$ 增大而 $R=I(U;V)$单调递减的过程中，存在着某一个特定的平均失真取值 $\overline{d}_{\min}$。当 $\overline{d}<\overline{d}_{\min}$ 时，$R=I(U;V)>0$；当 $\overline{d}\geqslant\overline{d}_{\min}$ 时，$R=I(U;V)\equiv 0$。

于是，作为在 D 失真许可条件下传输给定信源信息时的平均互信息的最小值(即使得 $I(U;V)$由非零变为零所对应的平均失真取值)，对于考察信息率失真函数 $R(D)$的基本函数关系有重要意义。

因此，设给定信源的信息率失真函数为 $R(D)$，对于失真限度 $D\geqslant 0$，一定存在一个自变量的取值 $D=D_{\max}$，使得 $D<D_{\max}$ 时，有

$$0<R(D)\leqslant H(U)$$

当 $D\geqslant D_{\max}$ 时，有

$$R(D)\equiv 0$$

则 $D_{\max}$ 表示 $R(D)$定义域的上界值。

下面我们讨论和给出 $R(D)$定义域的上界 $D_{\max}$ 的确定方法。

由前面关于系统平均失真 $\overline{d}$ 与系统信息传输率 $R=I(U;V)$ 的讨论可以看出，当 $\overline{d}\geqslant\overline{d}_{\min}$ 时，$R=I(U;V)\equiv 0$。由于与此对应的试验信道的输入 U 与输出 V 相互独立，因此为了得到 $\overline{d}_{\min}$，构造试验信道转移概率：

$$P(v_j\mid u_i)=P(v_j)\quad i=1,2,\cdots,r;\ j=1,2,\cdots,s \tag{8.21}$$

于是可以求出信道输入 U 与输出 V 之间的平均失真：

$$\overline{d}=E[d(u_i,v_j)]=\sum_{i=1}^{r}\sum_{j=1}^{s}P(u_i)P(v_j\mid u_i)d(u_i,v_j)$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{s}P(u_i)P(v_j)d(u_i,v_j)$$

显然，在满足 U 与 V 相互独立(见式(8.21))的条件下，不同的试验信道中产生的平均失真不同，而 $\overline{d}_{\min}$ 恰为此类信道中平均失真 $\overline{d}$ 的下界(最小值)，于是

$$\overline{d}_{\min}=\min\Big\{\sum_{i=1}^{r}\sum_{j=1}^{s}P(u_i)P(v_j)d(u_i,\ v_j)\Big\}$$

$$=\min\Big\{\sum_{j=1}^{s}P(v_j)\sum_{i=1}^{r}P(u_i)d(u_i,\ v_j)\Big\}$$

其中，内和式表示了 U、V 统计独立时，失真函数矩阵 $\boldsymbol{D}$ 的每一列元素 $d(u_i,\ v_j)$ 对 $P(u_i)$ 的均值。

为使上式总和最小，我们可以按下述方法来选择试验信道。

首先计算失真函数矩阵 $\boldsymbol{D}$ 中每一列元素 $d(u_i,\ v_j)$ 的统计平均值：

$$\sum_{i=1}^{r}P(u_i)d(u_i,\ v_j)\quad j=1,\ 2,\ \cdots,\ s$$

而后在各列均值中找出最小者。此处不妨设第 k 列最小，即

$$\min_j\Big\{\sum_{i=1}^{r}P(u_i)d(u_i,\ v_j)\Big\}=\sum_{i=1}^{r}P(u_i)d(u_i,\ v_k)$$

相应地，取试验信道的转移概率为

$$P(v_j\mid u_i)=P(v_j)=\begin{cases}1 & j=k\\ 0 & j\neq k\end{cases}\quad(i=1,\ 2,\ \cdots,\ r;\ j=1,\ 2,\ \cdots,\ s)$$

于是，得到了 U、V 统计独立条件下的最小平均失真，即

$$\overline{d}_{\min}=\min\Big\{\sum_{j=1}^{s}P(v_j)\sum_{i=1}^{r}P(u_i)d(u_i,\ v_j)\Big\}$$

$$=\sum_{j=1}^{s}P(v_j)\cdot\min_j\Big\{\sum_{i=1}^{r}P(u_i)d(u_i,\ v_j)\Big\}$$

$$=\min_j\Big\{\sum_{i=1}^{r}P(u_i)d(u_i,\ v_j)\Big\}$$

显然，由这样的 U、V 相互独立的试验信道所得到的平均失真的最小值即为信息率失真函数 $R(D)$ 定义域的上界值 $D_{\max}$，即

$$D_{\max}=\min\Big\{\sum_{U,\ V}P(u)P(v)d(u,\ v)\Big\}$$

$$=\min\Big\{\sum_{j=1}^{s}P(v_j)\sum_{i=1}^{r}P(u_i)d(u_i,\ v_j)\Big\}\tag{8.22}$$

并且当 $D\geqslant D_{\max}$ 时，有

$$R(D)\equiv 0$$

【例 8.4】 设输入随机变量的概率空间为 $[U,\ P]=\begin{bmatrix}-1 & 0 & +1\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{bmatrix}$，输出符号集 $V\in\Big\{-\frac{1}{2},\ \frac{1}{2}\Big\}$，定义失真函数矩阵为

$$\boldsymbol{D}=\begin{bmatrix}1 & 3\\ 1 & 2\\ 3 & 1\end{bmatrix}$$

则由式(8.22)可知：

$$D_{\max}=\min_j\Big\{\sum_{i=1}^{3}P(u_i)d(u_i,v_j)\Big\}$$
$$=\min\Big\{\frac{1}{3}\times1+\frac{1}{3}\times1+\frac{1}{3}\times3,\ \frac{1}{3}\times3+\frac{1}{3}\times2+\frac{1}{3}\times1\Big\}$$
$$=\min\Big\{\frac{5}{3},\ \frac{6}{3}\Big\}=\frac{5}{3}$$

可以看出，达到最大失真限度 $D_{\max}$ 所对应的试验信道矩阵为

$$P(v_j\mid u_i)=P(v_j)=\begin{cases}1 & j=1\\ 0 & j=2\end{cases}$$

并且当 $D\geqslant\frac{5}{3}$ 时，有

$$R(D)\equiv0$$

8.2.4　信息率失真函数的性质

性质 1　在定义域 $[D_{\min},D_{\max}]$ 内，$R(D)$ 是 D 的下凸函数，即已知 D_1，$D_2\in[D_{\min},D_{\max}]$，对于任意给定的 $0\leqslant\theta\leqslant1$，有：

$$R(\theta D_1+(1-\theta)D_2)\leqslant\theta R(D_1)+(1-\theta)R(D_2)\tag{8.23}$$

证明：在 $R(D)$ 的定义域内任取两点，并有：

$$R(D_1)=\min_{p_{ji}\in P_{D_1}}\{I(U;V_1)\}=I(U;V_1,P_1(v_j\mid u_i))$$

其中，$P_1(v_j\mid u_i)$ 是恰使 $I(U;V_1)$ 达到最小值的试验信道。

同理有：

$$R(D_2)=\min_{p_{ji}\in P_{D_2}}\{I(U;V_2)\}=I(U;V_2,P_2(v_j\mid u_i))$$

其中，$P_2(v_j\mid u_i)$ 是恰使 $I(U;V_2)$ 达到最小值的试验信道。

令 $0\leqslant\theta\leqslant1$，并令

$$D'=\theta D_1+(1-\theta)D_2$$

且

$$P(v_j\mid u_i)=\theta P_1(v_j\mid u_i)+(1-\theta)P_2(v_j\mid u_i)$$

首先证明 $P(v_j\mid u_i)$ 是 D' 失真许可信道集合 $P_{D'}$ 内的信道，即使用信道 $P(v_j\mid u_i)\in P_{D'}$ 传输信源信息时，系统中产生的平均失真 $\overline{d}\leqslant D'$。

因为

$$\overline{d}(P(v_j\mid u_i))=\sum_i\sum_jP(u_i)P(v_j\mid u_i)d(u_i,v_j)$$
$$=\sum_i\sum_jP(u_i)[\theta P_1(v_j\mid u_i)+(1-\theta)P_2(v_j\mid u_i)]d(u_i,v_j)$$
$$=\theta\sum_i\sum_jP(u_i)P_1(v_j\mid u_i)d(u_i,v_j)$$
$$+(1-\theta)\sum_i\sum_jP(u_i)P_2(v_j\mid u_i)d(u_i,v_j)$$

可记成

$$\overline{d}=\theta\overline{d}_1+(1-\theta)\overline{d}_2$$

因为

$$P_1(v_j \mid u_i) \in P_{D_1}$$
$$P_2(v_j \mid u_i) \in P_{D_2}$$

所以

$$\overline{d}_1 \leqslant D_1$$
$$\overline{d}_2 \leqslant D_2$$

于是

$$\overline{d}(P(v_j \mid u_i)) = \theta\overline{d}_1 + (1-\theta)\overline{d}_2 \leqslant \theta D_1 + (1-\theta)D_2 = D'$$

因此

$$P(v_j \mid u_i) \in P_{D'}$$

即 $P(v_j|u_i)$是 D 失真许可信道集合 P_D 内的一个信道。因此有：

$$R(D') = R(\theta D_1 + (1-\theta)D_2) \leqslant I(U; V, P(v_j \mid u_i)) \tag{8.24}$$

另外，由于平均互信息 $I(U; V, P(v_j|u_i))$是信道转移概率的下凸函数，因此有：

$$I(U; V, P(v_j \mid u_i)) \leqslant \theta I(U; V_1, P_1(v_j \mid u_i)) + (1-\theta)I(U; V_2, P_2(v_j \mid u_i))$$
$$\leqslant \theta R(D_1) + (1-\theta)R(D_2) \tag{8.25}$$

由式(8.24)和式(8.25)可得：

$$R(\theta D_1 + (1-\theta)D_2) \leqslant R(D_1) + (1-\theta)R(D_2)$$

因此信息率失真函数 $R(D)$在定义域内是下凸的。

性质 2 $R(D)$在定义域$[D_{\min}, D_{\max}]$内是单调递减函数。

由上面的讨论可以看出，当传输给定信源输出的信息时，允许的失真限度 D 愈大，则传输该信源必须有的最小信息传输率将愈小。因此，依信息率失真函数的定义可知，在定义域$[D_{\min}, D_{\max}]$内，$R(D)$是单调递减函数。$R(D)$在定义域内的单调递减性可以由 $R(D)$的下凸性加以证明。

证明：设 D_1 为定义域内的任意一点。在$(D_1, D_{\max})$中任取一点 D_θ，使

$$D_\theta = (1-\theta)D_1 + \theta D_{\max}$$

其中，θ 是任意小的正数。

由于 $R(D)$在定义域内是下凸的，因此有：

$$R(D_\theta) = R((1-\theta)D_1 + \theta D_{\max}) \leqslant (1-\theta)R(D_1) + \theta R(D_{\max})$$

已知 $R(D_{\max})=0$，则

$$R(D_\theta) \leqslant (1-\theta)R(D_1)$$

由于 $0<1-\theta<1$，于是

$$R(D_\theta) < R(D_1)$$

即对于定义域$[D_{\min}, D_{\max}]$内的任意两点 D_1 和 D_θ，无论 D_θ 多么接近D_1，只要 $D_\theta>D_1$，则一定满足 $R(D_\theta)<R(D_1)$。

所以，信息率失真函数 $R(D)$在定义域$[D_{\min}, D_{\max}]$内是单调递减函数。

性质 3 $R(D)$在定义域$[D_{\min}, D_{\max}]$内是失真限度 D 的连续函数。

对于给定的信源U，信息传输率 $R=I(U; V)$是信道转移概率 $P(v_j|u_i)$的函数，即

$$I(U; V) = I[P(v_j \mid u_i)] = \sum_{i=1}^{r}\sum_{j=1}^{s} P(u_i)P(v_j \mid u_i)\log\frac{P(v_j \mid u_i)}{\sum_{i=1}^{r} P(u_i)P(v_j \mid u_i)}$$

在满足概率分布关系的条件下，如果使信道转移概率 $P(v_j|u_i)$ 有一个微小的变化，即使信道参数为

$$P(v_j \mid u_i) \pm \varepsilon \qquad \varepsilon > 0$$

则系统的信息传输率改变为

$$I(U;V) = I[P(v_j \mid u_i) \pm \varepsilon]$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{s} P(u_i)[P(v_j \mid u_i) \pm \varepsilon]\log \frac{[P(v_j \mid u_i) \pm \varepsilon]}{\sum_{i=1}^{r} P(u_i)[P(v_j \mid u_i) \pm \varepsilon]}$$

令 $\varepsilon \to 0$，则

$$P(v_j) \neq 0$$

因此，平均互信息 $I(U;V)$ 是信道转移概率 $P(v_j|u_i)$ 的连续函数。

对于给定信源 U，如果失真限度为 D，则 D 失真许可信道集合为

$$P_D = \{P(v_j \mid u_i);\ \overline{d} \leqslant D\}$$

由信息率失真函数 $R(D)$ 的定义可知，一定存在一组信道参数 $P_1(v_j|u_i) \in P_D$，使得：

$$I[P_1(v_j \mid u_i)] = \min_{P(v|u)\in P_D}\{I(U;V)\} = R(D)$$

若将失真限度调整为

$$D' = D + \Delta \qquad \Delta \geqslant 0$$

则满足 D' 失真许可的信道集合为

$$P_{D'} = \{P(v_j \mid u_i);\ \overline{d} \leqslant D + \Delta\}$$

同理，一定可以找到一组信道参数：

$$P_2(v_j \mid u_i) \in P_{D'}$$

使得：

$$I[P_2(v_j \mid u_i)] = \min_{P(v|u)\in P_{D'}}\{I(U;V)\} = R(D')$$

显然，当 $\Delta \to 0$ 时，$P_{D'} \to P_D$，且有：

$$P_2(v_j \mid u_i) = P_1(v_j \mid u_i) \pm \varepsilon \qquad \varepsilon \to 0$$

由于平均互信息 $I(U;V)$ 是信道转移概率 $P(v_j|u_i)$ 的连续函数，因此有：

$$\lim_{\varepsilon\to 0} I[P_1(v_j \mid u_i) \pm \varepsilon] = I[P_1(v_j \mid u_i)]$$

此式等价于：

$$\lim_{\Delta\to 0} R(D') = \lim_{\Delta\to 0}\{\min_{P(v|u)\in P_{D'}}\{I(U;V)\}\} = \min_{P(v|u)\in P_D}\{I(U;V)\}$$

于是

$$\lim_{\Delta\to 0} R(D+\Delta) = \lim_{\Delta\to 0}\{\min_{P(v|u)\in P_{D'}}\{I(U;V)\}\} = \min_{P(v|u)\in P_D}\{I(U;V)\} = R(D)$$

因此，信息率失真函数 $R(D)$ 在其定义域 $[D_{\min}, D_{\max}]$ 内是失真限度 D 的连续函数。

根据 $R(D)$ 的上述性质，可以做出 $R(D)$ 的大致函数曲线。

设给定信源 U，其信息熵为 $H(U)$，并且已定义信道输入 U 与输出 V 之间的失真函数矩阵 $\boldsymbol{D}$。如果失真函数矩阵 $\boldsymbol{D}$ 的每一行元素中的最小值为 0，且只有一个元素为 0，则由前面关于信息率失真函数 $R(D)$ 的基本函数关系的讨论可知，系统中允许的失真限度 D 的下限值为 $D_{\min}=0$，且有：

$$R(D_{\min}) = R(0) = H(U)$$

并且一定存在一个失真限度 $D_{\max}$，当 $D \geqslant D_{\max}$ 时，$R(D_{\max}) \equiv 0$。于是，在失真限度 D 和信息率失真函数 $R(D)$ 分别为横、纵坐标构成的坐标系中，可以标出信息失真函数的定义域 $[0, D_{\max}]$ 与函数中的两个点：$(0, H(U))$ 和 $(D_{\max}, 0)$。

由于信息失真函数 $R(D)$ 在其定义域 $[0, D_{\max}]$ 内是下凸、单调递减的连续函数，因此可以做出 $R(D)$ 在其定义域内的函数曲线，并且当 $D \geqslant D_{\max}$ 时，有 $R(D_{\max}) \equiv 0$，如图 8-3(a)所示。

如果对于给定的失真函数矩阵 $\boldsymbol{D}$，系统中允许的失真限度 $D \neq 0$，则可计算得到信息失真函数 $R(D)$ 的定义域的下界 $D_{\min}$ 和上界 $D_{\max}$，求出相应的信息率失真函数取值 $R(D_{\min})$ 和 $R(D_{\max}) \equiv 0$，依据信息失真函数 $R(D)$ 在其定义域 $[D_{\min}, D_{\max}]$ 内的下凸性、单调递减性和连续性，做出 $R(D)$ 的函数曲线，如图 8-3(b)所示。（注意：当 $D \geqslant D_{\max}$ 时，$R(D_{\max}) \equiv 0$。）

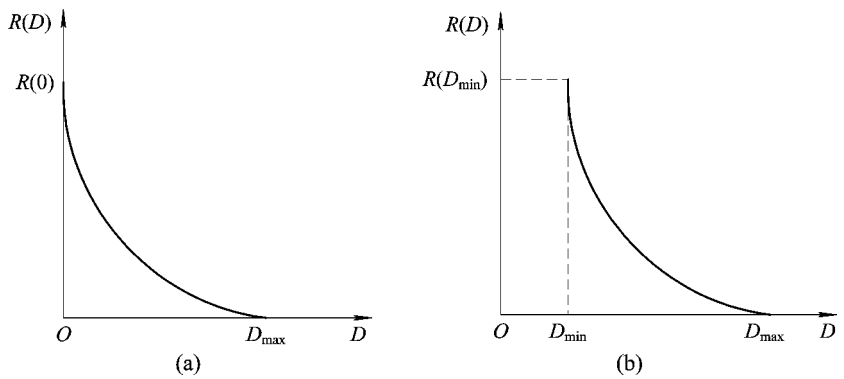

图 8-3 $R(D)$ 函数的曲线

8.3 信息率失真函数的计算

8.3.1 信息率失真函数的一般计算方法

由前面关于信息率失真函数的定义知道，$R(D)$ 是在满足失真限度 D 的条件下系统平均互信息 $I(X; Y)$ 的最小值。于是，对于给定的信源 U 和定义的失真函数矩阵 $\boldsymbol{D}$，首先需要找出满足失真限度 D 要求的信道，构成 D 失真许可信道集合 P_D，依据平均互信息 $I(X; Y)$ 的下凸性，从 D 失真许可信道集合 P_D 内选出一组信道参数函数 $P(v|u)$，使得平均互信息 $I(X; Y)$ 达到其最小值，得到信息率失真函数：

$$R(D) = \min_{P(v|u) \in P_D} \{I(U; V)\}$$

显然，这种在使用信道受限的条件下求出平均互信息最小值的问题（即信息率失真函数 $R(D)$ 的计算）是一个比较复杂和困难的问题。

设给定离散无记忆信源：

$$[U, P] = \begin{bmatrix} u_1 & u_2 & \cdots & u_r \\ P(u_1) & P(u_2) & \cdots & P(u_r) \end{bmatrix}$$

已知信道的输出符号集合为 $V=[v_1, v_2, \cdots, v_s]$，定义 U、V 之间的失真函数矩阵：

$$\boldsymbol{D}=\begin{bmatrix} d(u_1, v_1) & d(u_1, v_2) & \cdots & d(u_1, v_s) \\ d(u_2, v_1) & d(u_2, v_2) & \cdots & d(u_2, v_s) \\ \vdots & \vdots & & \vdots \\ d(u_r, v_1) & d(u_r, v_2) & \cdots & d(u_r, v_s) \end{bmatrix}$$

由上面关于信息率失真函数 $R(D)$的基本函数关系和性质可以知道，给定信源的信息率失真函数 $R(D)$的一般计算内容主要包括：

(1) 确定 $D_{\min}$及 $R(D_{\min})$；

(2) 确定 $D_{\max}$及 $R(D_{\max})$；

(3) 确定 $D_{\min}<D<D_{\max}$内 $R(D)$的解析式。

此处，我们以一个特定类型的信源 U 和失真函数矩阵 $\boldsymbol{D}$ 为例，介绍信息率失真函数 $R(D)$的一般计算思路和方法，以进一步理解和掌握 $R(D)$的基本函数关系和性质。

【例 8.5】　设二元对称信源 $[U, P]=\begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix}$，其中 $p\leqslant\frac{1}{2}$，接收符号集 $V=\{0, 1\}$，定义失真函数矩阵 $\boldsymbol{D}=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$，试求 $R(D)$在其定义域内的解析式。

解：首先求出信源 U 的信息熵：

$$H(U)=H(p, 1-p)=H_2(p)$$

(1) 确定 $D_{\min}$及 $R(D_{\min})$。由前面关于 $R(D)$的定义域下界的讨论可以得到：

$$\begin{aligned} D_{\min}&=\sum_{i=1}^{2}P(u_i)\min_j\{d(u_i, v_j)\} \\ &=P(u_1)\min\{d(u_1, v_1), d(u_1, v_2)\}+P(u_2)\min\{d(u_2, v_1), d(u_2, v_2)\}=0 \end{aligned}$$

可见，此系统中允许的最小失真限度 $D_{\min}=0$。可以找出使得 $D_{\min}=0$ 的试验信道为

$$[P(v\mid u)]=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

显然，此试验信道是一个无噪无损信道，可以求出：

$$R(D_{\min})=R(0)=I(U; V)=H(p, 1-p)=H_2(p)$$

(2) 确定 $D_{\max}$及 $R(D_{\max})$。

$$\begin{aligned} D_{\max}&=\min_j\Big\{\sum_{i=1}^{2}P(u_i)d(u_i, v_j)\Big\} \\ &=\min\Big\{\sum_{i=1}^{2}P(u_i)d(u_i, v_1), \sum_{i=1}^{2}P(u_i)d(u_i, v_2)\Big\} \\ &=\min\{1-p, p\} \\ &=p \quad \left(\text{因为 } p\leqslant\frac{1}{2}\right) \end{aligned}$$

于是，使得 $R(D)=0$ 时，失真限度 D 的最小值，即 $R(D)$定义域的上界为

$$D_{\max}=p$$

可以找出，使得系统平均失真恰为 $D_{\max}$ 的试验信道为

$$[P(v \mid u)] = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$

显然，这个试验信道的输出 V 与输入 U 相互独立，即无论信源 U 输出符号为 u_1 或 u_2，信道的输出符号都是 v_2。因此，当信源 U 输出符号 u_1 时一定发生传输错误，而当信源 U 输出符号 u_2 时，信道的输出与输入之间的失真为 0。由于 $P(u_1)=p$，因此此时系统的平均失真 $\overline{d}=D_{\max}=p$。

同时，由信息率失真函数的基本关系知：

$$R(D_{\max}) = 0$$

(3) 确定 $0<D<p$ 内 $R(D)$ 的解析式。

由 $R(D)$ 的定义式：

$$R(D) = \min_{p_{ij} \in P_D} \{I(U; V)\} = \min\{I(U; V); \overline{d} \leqslant D\}$$

可知，$R(D)$ 实际上是在满足保真度准则 $\overline{d} \leqslant D$ 条件下平均互信息 $I(U; V)$ 的最小值。由于 $R(D)$ 在定义域内是单调递减函数，因此 $R(D)$ 即为在 $\overline{d}=D$ 的条件下 $I(U; V)$ 的最小值，即

$$R(D) = \min\{I(U; V); \overline{d} = D\} = \min\{H(U) - H(U \mid V); \overline{d} = D\}$$

其中：

$$\begin{aligned} \overline{d} = E[d] &= \sum_{U, V} P(u)P(v \mid u)d(u, v) = \sum_{U, V} P(u, v)d(u, v) \\ &= P(u_1, v_2) + P(u_2, v_1) \end{aligned}$$

显然，此时的平均失真为错误概率：

$$\overline{d} = E[d] = P(U \neq V) = p_e$$

若使系统的平均失真等于错误概率：

$$\overline{d} = p_e = D$$

则有：

$$R(D) = \min\{H(U) - H(U \mid V); p_e = D\}$$

而根据范诺不等式($r=2$)有：

$$H(U \mid V) \leqslant H(p_e) + p_e \log(2-1) = H(p_e)$$

即

$$H(U \mid V) \leqslant H(D)$$

因此有：

$$\begin{aligned} I(U; V) &= H(U) - H(U \mid V) \geqslant H(U) - H(D) \\ &= H_2(p) - H(D) \end{aligned}$$

所以，一定存在一个试验信道$[P(v|u)]$，使得平均互信息 $I(U; V)$ 可以达到上述不等式的下界，于是，在 $0 \leqslant D \leqslant p$ 内满足 $\overline{d} \leqslant D$ 条件下信源的信息率失真函数 $R(D)$ 为

$$R(D) = H_2(p) - H(D)$$

上面我们只是假定一定存在一个试验信道$[P(v|u)]$，使得用该信道传输信源信息时系统的平均失真 $\overline{d} \leqslant D$，而接收端获得的平均信息量 $I(U; V)=H_2(p)-H(D)$。关于这一假设是否真的成立，还需要进一步证实。为证实这一点，就必须找到一个这样的试验信道，

使得其平均失真 $\overline{d}=D$，接收端获得的平均信息量度：

$$I(U;V)=H_2(p)-H(D)$$

为此，我们构造一个反向的试验信道，即将原信道的传递关系反向，使原信道的输入端为输出端，而输出端为输入端。按照失真限度 D 的要求，所构造的反向试验如图 8-4 所示。

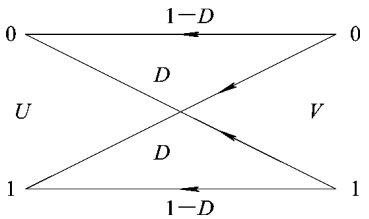

图 8-4　反向信道转移概率

其转移概率为

$$P(u=0\mid v=0)=1-D$$
$$P(u=1\mid v=0)=D$$
$$P(u=0\mid v=1)=D$$
$$P(u=1\mid v=1)=1-D$$

现在，我们来计算用该信道传输信源时系统的平均失真及接收端获得的平均互信息。

对于此反向试验信道，可以求出系统的平均失真和平均互信息。系统的平均失真：

$$\begin{aligned}\overline{d}=E[d]&=\sum_{U,V}P(u,v)d(u,v)\\&=\sum_{U,V}P(u)P(u\mid v)d(u,v)\\&=P(v=1)P(u=0\mid v=1)+P(v=0)P(u=1\mid v=0)\\&=\frac{1-p-D}{1-2D}\times D+\frac{p-D}{1-2D}\times D\\&=D\end{aligned}$$

系统的平均互信息 $I(U;V)$ 为

$$I(U;V)=H(U)-H(U\mid V)$$

而

$$\begin{aligned}H(U\mid V)&=\sum_{U,V}P(u,v)\log\frac{1}{P(u\mid v)}\\&=\sum_{i=1}^{2}\sum_{j=1}^{2}P(v_j)P(u_i\mid v_j)\log\frac{1}{P(u_i\mid v_j)}\\&=P(v_1)\sum_{i=1}^{2}P(u_i\mid v_1)\log\frac{1}{P(u_i\mid v_1)}+P(v_2)\sum_{i=1}^{2}P(u_i\mid v_2)\log\frac{1}{P(u_i\mid v_2)}\\&=P(v_1)H(D)+P(v_2)H(D)\\&=H(D)\end{aligned}$$

因此

$$I(U;V)=H(U)-H(U\mid V)=H_2(p)-H(D)$$

由此可见，所选择的反向试验信道满足失真许可准则 $\overline{d}\leqslant D$，且接收端获得的平均信息量达到平均互信息的下界，即

$$I(U;V)=H_2(p)-H(D)$$

但是上述结论是对一个构造的反向试验信道进行计算得到的，而此反向试验信道的构造是否合理，即此反向试验信道是否存在，还需要依据已知的信源概率分布 $P(u)$ 和反向试验信道的转移概率 $P(u|v)$，检验随机变量 V 是否满足概率分布关系的要求。由于在上面的分析中已经应用了 $P(v_1)+P(v_2)=1$ 的条件，因此只需验证随机变量 V 的概率分布是否满足：

$$0\leqslant P(v_1)\leqslant 1$$

设

$$P(v_1)=\alpha,\ P(v_2)=1-\alpha\quad 0<\alpha<1$$

则由已知条件可得：

$$\begin{aligned}P(u_1)&=P(v_1)P(u_1\mid v_1)+P(v_2)P(u_1\mid v_2)\\&=\alpha(1-D)+(1-\alpha)D=p\end{aligned}$$

解得：

$$\alpha=\frac{p-D}{1-2D}$$

由于已知

$$0\leqslant p\leqslant\frac{1}{2}$$

而

$$0\leqslant D\leqslant p$$

所以

$$0\leqslant P(v_1)=\alpha\leqslant\frac{1}{2}$$

因此，上面定义的反向试验信道参数合理，所构成的反向试验信道是存在的。

又因为

$$P(v_j\mid u_i)=\frac{P(u_i\mid v_j)P(v_j)}{P(u_i)}$$

所以我们得到了在满足 $\overline{d}\leqslant D$ 的条件下，使得平均互信息 $I(U;V)$ 达到最小值时所对应的试验信道。

于是，通过对信息率失真函数的定义域、值域和解析式的分析与计算，对于此例所给定的信源 U 和定义的失真函数，信息率失真函数 $R(D)$ 为

$$R(D)=\begin{cases}H(p)-H(D) & 0\leqslant D\leqslant p\\ 0 & D>p\end{cases}\quad \text{比特 / 符号}$$

图 8-5 给出了 $R(D)$ 的函数曲线。图 8-6 描述了在 p 取不同值时的 $R(D)$ 曲线。由此可见，对于同一个 D，信源分布越均匀，$R(D)$ 就越大，信源压缩的可能性就越小。反之，若信源分布越不均匀，即信源信息冗余度越大，则 $R(D)$ 就越小，压缩的可能性就越大。

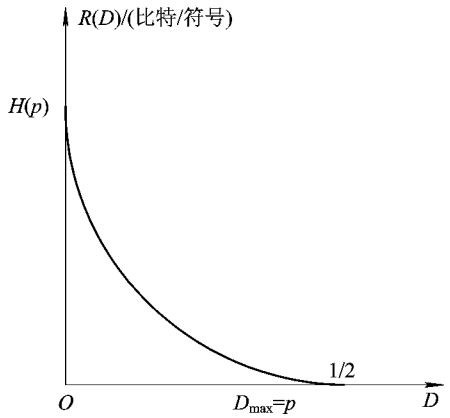

图 8-5　二元对称信源的 $R(D)$

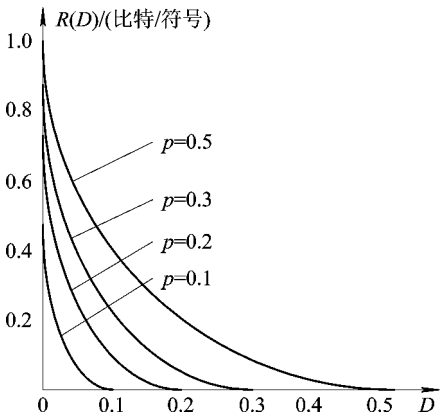

图 8-6　p 取不同值时的 $R(D)$ 函数

8.3.2　r 元等概信源和汉明失真函数的 $R(D)$

设信源、信宿的符号集相同，即 $U, V \in \{0, 1, 2, \cdots, r-1\}$，已知信源符号等概分布，定义失真函数为汉明失真，即

$$d(u_i, v_j) = \begin{cases} 0 & u_i = v_j \\ a(a>0) & u_i \neq v_j \end{cases}$$

则信源的信息率失真函数为

$$R(D) = \begin{cases} \log r - \dfrac{D}{a}\log(r-1) - H\left(\dfrac{D}{a}\right) & 0 \leqslant D \leqslant a\left(1-\dfrac{1}{r}\right) \\ 0 & D > a\left(1-\dfrac{1}{r}\right) \end{cases}$$

称为 r 元等概信源和汉明失真函数的 $R(D)$，也可称做 r 元信源在错误概率准则下的信息率失真函数。

下面我们来证明上述结论。

证明：对于给定的信源和汉明失真函数可以求出：

$$D_{\min} = \sum_{i=1}^{r} P(u_i) \min_j \{d(u_i, v_j)\} = 0$$

$$D_{\max} = \min_j \{\sum_{i=1}^{r} P(u_i) d(u_i, v_j)\}$$

$$= \min \underbrace{\left\{\frac{1}{r} \cdot a \cdot (r-1), \frac{1}{r} \cdot a \cdot (r-1), \cdots, \frac{1}{r} \cdot a \cdot (r-1)\right\}}_{\text{共 } r \text{ 列}}$$

$$= a\left(1-\frac{1}{r}\right)$$

以及

$$R(D_{\min}) = R(0) = H(U) = \log r \text{ 比特 / 符号}$$

$$R(D_{\max}) = 0$$

当 $D_{\min} < D < D_{\max}$，时，由于 $R(D)$ 为失真限度的单调递减函数，因此

$$R(D) = \min\{I(U;V);\ \bar{d} \leqslant D\}$$
$$= \min\{I(U;V);\ \bar{d} = D\} \quad (\text{因为 } R(D) \text{ 单调递减})$$

首先计算系统的平均失真 $\bar{d}$：

$$\bar{d} = E[d] = \sum_{U,V} P(u,v)d(u,v)$$
$$= \sum_{U,V\text{且}U\neq V} P(u,v)d(u,v) + \sum_{U,V\text{且}U=V} P(u,v)d(u,v)$$
$$= \sum_{U,V\text{且}U\neq V} P(u,v)d(u,v) \quad (\text{因为 } u=v \text{ 时},\ d(u,v)=0)$$
$$= a\sum_{U,V}\{P(u,v);\ u\neq v\}$$
$$= aP(U\neq V) = ap_e$$

其中，p_e 为信道的平均错误概率。

令 $\bar{d}=ap_e=D$，得 $p_e=D/a$，于是根据范诺不等式有：

$$H(U\mid V) \leqslant H(p_e) + p_e\log(r-1) = H\left(\frac{D}{a}\right) + \frac{D}{a}\log(r-1)$$

所以

$$I(U\mid V) = H(U) - H(U\mid V)$$
$$\geqslant \log r - H\left(\frac{D}{a}\right) - \frac{D}{a}\log(r-1)$$

$$R(D) = \min\{I(U;V);\ \bar{d}=D\} = \min I(U;V);\ p_e = \frac{D}{a}$$

因此，根据范诺不等式，一定存在一个试验信道$[P(v|u)]$，使得在该信道中平均失真 $\bar{d}=D$，且接收端获得的平均信息量 $I(U;V)$可以达到上述不等式的下界，即

$$H(U;V) = \log r - H\left(\frac{D}{a}\right) - \frac{D}{a}\log(r-1)$$

于是可得 r 元等概信源在汉明失真函数下的信息率失真函数 $R(D)$为

$$R(D) = \log r - H\left(\frac{D}{a}\right) - \frac{D}{a}\log(r-1) \quad 0 \leqslant D \leqslant a\left(1-\frac{1}{r}\right)$$

由于上述讨论中我们假定存在一个反向试验信道，因此对于反向试验信道的合理性还需要作进一步的论证，即设法找到一个试验信道，在该信道中平均失真 $\bar{d}=D$，且接收端获得的平均信息量：

$$H(U;V) = \log r - H\left(\frac{D}{a}\right) - \frac{D}{a}\log(r-1)$$

为此，我们构造一个反向信道，如图 8-7 所示，并用该反向信道作为试验信道。首先，我们需要验证一下该反向信道是否存在，或者说是否合理，即验证下列条件是否成立：

$$0 \leqslant P(v) \leqslant 1 \quad \text{且} \quad \sum_v P(v) = 1$$

根据已知条件 $P(u_i)=1/r$ 及反向信道的信道转移

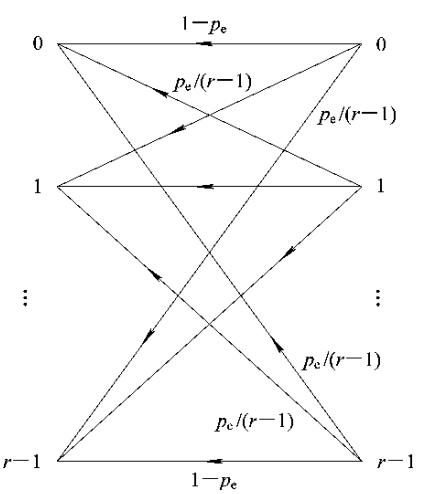

图 8-7 反向试验信道

概率：

$$P(u_i \mid v_j)=\begin{cases}1-p_e & u_i=v_j\\ \dfrac{p_e}{r-1} & u_i\neq v_j\end{cases}$$

可由公式：

$$P(u_i)=\sum_{j=1}^{r}P(u_i,v_j)=\sum_{j=1}^{r}P(v_j)P(u_i\mid v_j)\quad i=1,2,\cdots,r$$

求得 r 元方程组的解：

$$P(v_j)=\frac{1}{r}\quad j=1,2,\cdots,r$$

显然：

$$\begin{cases}0<P(v_j)<1 & j=1,2,\cdots,r\\ \sum\limits_{j=1}^{r}P(v_j)=1\end{cases}$$

所以该反向信道是一个合理的信道。

现在，我们用该反向信道作为试验信道，并计算在该试验信道中的平均失真 $\overline{d}$ 及接收端获得的平均信息量 $I(U;V)$：

$$\begin{aligned}\overline{d}&=E[d]=\sum_{U,V}P(u,v)d(u,v)\\&=\sum_{U}\sum_{V}P(v)P(u\mid v)d(u,v)\\&=\sum_{V}P(v)\sum_{U}P(u\mid v)d(u,v)\\&=\sum_{V}P(v)\left[(1-p_e)\cdot 0+\frac{p_e}{r-1}(r-1)\cdot a\right]\\&=ap_e\sum_{V}P(v)\\&=ap_e\\&=D\end{aligned}$$

$$\begin{aligned}H(U\mid V)&=\sum_{U}\sum_{V}P(v)P(u\mid v)\log\frac{1}{P(u\mid v)}\\&=\sum_{V}P(v)\sum_{U}P(u\mid v)\log\frac{1}{P(u\mid v)}\\&=\sum_{V}P(v)\left[(1-p_e)\log\frac{1}{1-p_e}+\left(\frac{p_e}{r-1}\log\frac{r-1}{p_e}\right)\cdot(r-1)\right]\\&=(1-p_e)\log\frac{1}{1-p_e}+p_e\log\frac{r-1}{p_e}\\&=(1-p_e)\log\frac{1}{1-p_e}+p_e\log\frac{1}{p_e}+p_e\log(r-1)\\&=H(p_e)+p_e\log(r-1)\end{aligned}$$

因此：

$$\begin{aligned} I(U;V) &= H(U) - H(U \mid V) \\ &= \log r - H(p_e) - p_e \log(r-1) \\ &\geqslant \log r - H\left(\frac{D}{a}\right) - \frac{D}{a}\log(r-1) \end{aligned}$$

可见，在所选择的反向试验信道中平均失真 $\bar{d}=D$，且接收端获得平均互信息达到最小值。因此，在汉明失真函数定义下，r 元等概信源的信息率失真函数 $R(D)$ 为

$$R(D) = \begin{cases} \log r - \dfrac{D}{a}\log(r-1) - H\left(\dfrac{D}{a}\right) & 0 \leqslant D \leqslant a\left(1-\dfrac{1}{r}\right) \\ 0 & D > a\left(1-\dfrac{1}{r}\right) \end{cases}$$

根据 r 的不同取值，可做出 $R(D)$ 的曲线，如图 8-8 所示。由曲线图可以看出，对于同一失真限度 D，r 越大，$R(D)$ 越大，信源压缩的可能性就越小；反之，r 越小，$R(D)$ 越小，信源压缩的可能性就越大。这一规律对于实际信源的量化分层及数据压缩有深刻的理论指导意义。

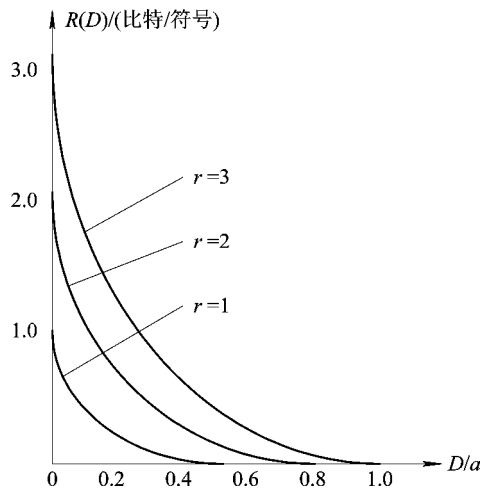

图 8-8 r 取不同值时的 $R(D)$ 函数

8.3.3 信息率失真函数的参量表示

根据 $R(D)$ 的定义式，对于给定的信源（$P(u)$ 固定），在定义了具体的失真函数后，就可求得信源的 $R(D)$ 函数，方法上与求信道的信道容量一样，是一个在有约束的条件下计算平均互信息 $I(U;V)$ 的极小值问题，即选择适当的试验信道 $P(v|u)$ 使平均互信息：

$$I(U;V) = \sum_{i=1}^{r}\sum_{j=1}^{s} P(u_i)P(v_j \mid u_i)\ln\frac{P(v_j \mid u_i)}{\sum_{k=1}^{r} P(u_k)P(v_j \mid u_k)} \tag{8.26}$$

最小化，并使得 $P(v_j|u_i)$ 满足以下约束条件：

$$P(v_j \mid u_i) \geqslant 0 \quad i = 1, 2, \cdots, r;\ j = 1, 2, \cdots, s \tag{8.27}$$

$$\sum_{j=1}^{s} P(v_j \mid u_i) = 1 \quad i = 1, 2, \cdots, r \tag{8.28}$$

和

$$\sum_{i=1}^{r}\sum_{j=1}^{s} P(u_i)P(v_j \mid u_i)d(u_i, v_j) = D \tag{8.29}$$

因此，可采用拉格朗日乘子法进行求解。但是如果要求得到明显的解析式，那么也是比较困难的，通常只能用参量形式来表达。即便如此，除简单情况外，实际计算仍然是相当困难的，尤其约束条件式(8.27)是求解 $R(D)$ 函数的主要障碍。因为应用拉格朗日乘子法解得的一个或几个 $P(v_j \mid u_i)$ 很可能是负的，所以在这种情况下，必须假设某些 $P(v_j \mid u_i)=0$，然后重新计算，这就使得计算复杂化了。目前，可采用收敛的迭代算法在计算机上求解 $R(D)$ 函数。

下面介绍拉格朗日乘子法求解 $R(D)$ 函数，并用 S 作为参量来表述信息率失真函数 $R(D)$ 和平均失真函数 $D(S)$。值得说明的是，为了讨论和书写方便，平均互信息表达式(式(8.26))中采用了自然对数的形式。

首先，构造一辅助函数 Φ，在该辅助函数中暂且不考虑约束条件式(8.27)，仅考虑约束条件式(8.28)和式(8.29)。其中，约束条件式(8.28)包含 r 个等式，取拉格朗日乘子 $\mu_i(i=1, 2, \cdots, r)$ 分别与之对应，并取拉格朗日乘子 S 与式(8.29)对应。

$$\Phi = I(U; V) - \mu_i \sum_{j=1}^{s} P(v_j \mid u_i) - SD \tag{8.30}$$

求平均互信息 $I(U; V)$ 的极小值，实质上就是求辅助函数 Φ 的一阶导数等于零的方程组的解。因为 $I(U; V)$ 是信道转移概率 $P(v|u)$ 的下凸函数，所以若极值存在，则它一定是极小值，即求：

$$\frac{\partial \Phi}{\partial P(v_j \mid u_i)} = 0 \quad i = 1, 2, \cdots, r; \ j = 1, 2, \cdots, s \tag{8.31}$$

因为

$$\begin{aligned}\frac{\partial I(U; V)}{\partial P(v_j \mid u_i)} &= P(u_i)\ln P(v_j \mid u_i) + P(u_i) - P(u_i)\ln P(v_j) - P(u_i) \\ &= P(u_i)\ln \frac{P(v_j \mid u_i)}{P(v_j)}\end{aligned}$$

又因为

$$\frac{\partial \mu_i \sum_{j=1}^{s} P(v_j \mid u_i)}{\partial P(v_j \mid u_i)} = \mu_i$$

$$\begin{aligned}\frac{\partial SD}{\partial P(v_j \mid u_i)} &= \frac{\partial S \sum_{i=1}^{r}\sum_{j=1}^{s} P(u_i)P(v_j \mid u_i)d(u_i, v_j)}{\partial P(v_j \mid u_i)} \\ &= SP(u_i)d(u_i, v_j)\end{aligned}$$

于是得：

$$\begin{aligned}\frac{\partial \Phi}{\partial P(v_j \mid u_i)} = P(u_i)\ln \frac{P(v_j \mid u_i)}{P(v_j)} - SP(u_i)d(u_i, v_j) - \mu_i = 0 \\ i = 1, 2, \cdots, r; \ j = 1, 2, \cdots, s\end{aligned} \tag{8.32}$$

当 $i=1, 2, \cdots, r$ 和 $j=1, 2, \cdots, s$ 分别取值时，式(8.32)共有 $r\times s$ 个方程，加上式(8.28)的 r 个方程，共有 $r+1+r\times s$ 个方程。而其中未知数 $u_i(i=1, 2, \cdots, r)$、S 和 $P(v_j|u_i)(i=1, 2, \cdots, r; j=1, 2, \cdots, s)$的个数也正好为 $r+1+r\times s$ 个，所以原则上求极小值的问题已解决。只需求解式(8.28)、式(8.29)和式(8.32)组成的方程组，即可求出 $I(U; V)$在约束条件下的极小值。

为了求解方便，令 $\mu_i=P(u_i)\ln\lambda_i$，并整理式(8.32)得：

$$\ln P(v_j \mid u_i)-\ln P(v_j)-Sd(u_i, v_j)-\ln\lambda_i=0 \quad i=1, 2, \cdots, r; j=1, 2, \cdots, s \tag{8.33}$$

求解式(8.33)可得 $r\times s$ 个信道转移概率：

$$P(v_j \mid u_i)=P(v_j)\lambda_i\exp[Sd(u_i, v_j)] \quad i=1, 2, \cdots, r; j=1, 2, \cdots, s \tag{8.34}$$

将式(8.34)对 j 求和得：

$$\sum_j P(v_j \mid u_i)=\sum_j P(v_j)\lambda_i\exp[Sd(u_i, v_j)]=1$$

即

$$\lambda_i=\frac{1}{\sum_j P(v_j)\exp[Sd(u_i, v_j)]} \quad i=1, 2, \cdots, r \tag{8.35}$$

再将式(8.34)两端同乘以 $P(u_i)$并对 i 求和得：

$$\sum_i P(u_i)P(v_j \mid u_i)=P(v_j)\sum_i \lambda_i P(u_i)\exp[Sd(u_i, v_j)]$$

若 $P(v_j)\neq 0$，则可得：

$$\sum_i \lambda_i P(u_i)\exp[Sd(u_i, v_j)]=1 \quad j=1, 2, \cdots, s \tag{8.36}$$

将式(8.35)代入式(8.36)得：

$$\sum_i \frac{P(u_i)\exp[Sd(u_i, v_j)]}{\sum_j P(v_j)\exp[Sd(u_i, v_j)]}=1 \quad j=1, 2, \cdots, s \tag{8.37}$$

由式(8.37)中的 S 个方程就可以解出 S 个输出符号的概率 $P(v_j)$，然后将所求的 $P(v_j)$代入式(8.35)即可求出 λ_i。

若将式(8.35)代入式(8.34)，则得：

$$P(v_j \mid u_i)=\frac{P(v_j)\exp[Sd(u_i, v_j)]}{\sum_j P(v_j)\exp[Sd(u_i, v_j)]} \quad i=1, 2, \cdots, r; j=1, 2, \cdots, s \tag{8.38}$$

将由式(8.37)求得的 $P(v_j)$代入式(8.38)，即可求得 $I(U; V)$取极小值时的试验信道的转移概率 $P(v_j|u_i)$。应该强调的是，这里所得的结果是以 S 为参量的表达式，而不是显式表达式。因而所得到的 $R(D)$的表达式也将是以 S 为参量的表达式。参量 S 对应的约束条件为式(8.29)，它与允许的失真限度 D 有关，所以以 S 为参量就相当于以 D 为参量。

将式(8.34)分别代入式(8.29)和式(8.26)，可得到以 S 为参量的信息率失真函数 $R(S)$和平均失真函数 $D(S)$。

$$\begin{aligned} D(S)&=\sum_{i=1}^{r}\sum_{j=1}^{s}P(u_i)P(v_j)\lambda_i\exp[Sd(u_i, v_j)]d(u_i, v_j)\\ &=\sum_{i=1}^{r}\sum_{j=1}^{s}P(u_i)P(v_j)\lambda_i d(u_i, v_j)\exp[Sd(u_i, v_j)] \end{aligned} \tag{8.39}$$

$$
\begin{aligned}
R(S) &= \sum_{i=1}^{r}\sum_{j=1}^{s} P(u_i)P(v_j)\lambda_i \exp[Sd(u_i, v_j)] \ln \frac{P(v_j)\lambda_i \exp[Sd(u_i, v_j)]}{P(v_j)} \\
&= \sum_{i=1}^{r}\sum_{j=1}^{s} P(u_i)P(v_j)\lambda_i \exp[Sd(u_i, v_j)] \ln\{\lambda_i \exp[Sd(u_i, v_j)]\} \\
&= \sum_{i=1}^{r}\sum_{j=1}^{s} P(u_i)P(v_j)\lambda_i \exp[Sd(u_i, v_j)] \ln\lambda_i \\
&\quad + \sum_{i=1}^{r}\sum_{j=1}^{s} P(u_i)P(v_j)\lambda_i \exp[Sd(u_i, v_j)] Sd(u_i, v_j) \\
&= \sum_{i=1}^{r}\sum_{j=1}^{s} P(u_i)P(v_j \mid u_i)\ln\lambda_i + SD(S) \\
&= SD(S) + \sum_{i=1}^{r} P(u_i)\ln\lambda_i \sum_{j=1}^{s} P(v_j \mid u_i) \\
&= SD(S) + \sum_{i=1}^{r} P(u_i)\ln\lambda_i
\end{aligned} \tag{8.40}
$$

可见，当参数 S 给定某一数值时，可由式(8.37)求出 $P(v_j)(j=1, 2, \cdots, s)$，再由式(8.35)计算出 $\lambda_i(i=1, 2, \cdots, r)$，将所求的结果代入式(8.39)和式(8.40)即可确定 $R(D)$ 的值。由于 $P(v_j)(j=1, 2, \cdots, s)$ 不能是负值，所以参量 S 的取值有一定的限制。现在我们来分析一下参量 S 的物理意义及其可能取值的范围。

由于 D 是参量 S 的函数，λ_i 也是 S 的函数，因此，也可以把 S 看成是 D 的函数，则 λ_i 也是 D 的函数。首先求 $R(S)$ 对 D 的导数，得：

$$
\begin{aligned}
\frac{\mathrm{d}R(S)}{\mathrm{d}D} &= \frac{\partial R(S)}{\partial S}\frac{\mathrm{d}S}{\mathrm{d}D} = \frac{\partial}{\partial S}\Big[SD(S) + \sum_{i=1}^{r} P(u_i)\ln\lambda_i\Big]\frac{\mathrm{d}S}{\mathrm{d}D} \\
&= \Big[D + S\frac{\mathrm{d}D}{\mathrm{d}S} + \sum_{i=1}^{r} P(u_i)\frac{1}{\lambda_i}\frac{\mathrm{d}\lambda_i}{\mathrm{d}S}\Big]\cdot\frac{\mathrm{d}S}{\mathrm{d}D} \\
&= S + \Big[D + \sum_{i=1}^{r} P(u_i)\frac{1}{\lambda_i}\frac{\mathrm{d}\lambda_i}{\mathrm{d}S}\Big]\cdot\frac{\mathrm{d}S}{\mathrm{d}D}
\end{aligned} \tag{8.41}
$$

将式(8.36)对 S 求导，则得：

$$
\sum_i P(u_i)\exp[Sd(u_i, v_j)]\frac{\mathrm{d}\lambda_i}{\mathrm{d}S} + \sum_i \lambda_i P(u_i)d(u_i, v_j)\exp[Sd(u_i, v_j)] = 0 \qquad j = 1, 2, \cdots, s
$$

将上式两端同乘以 $P(v_j)$，并对 j 求和，得：

$$
\sum_i\sum_j P(u_i)P(v_j)\exp[Sd(u_i, v_j)] + \sum_i\sum_j \lambda_i P(u_i)P(v_j)d(u_i, v_j)\exp[Sd(u_i, v_j)] = 0
$$

即

$$
\sum_i P(u_i)\frac{\mathrm{d}\lambda_i}{\mathrm{d}S}\sum_j P(v_j)\exp[Sd(u_i, v_j)] + D = 0
$$

考虑式(8.35)，上式可简化为

$$
\sum_i \frac{P(u_i)}{\lambda_i}\frac{\mathrm{d}\lambda_i}{\mathrm{d}S} + D = 0
$$

将上式代入式(8.41)可得：

$$
\frac{\mathrm{d}R(S)}{\mathrm{d}D} = S + 0 \times \frac{\mathrm{d}S}{\mathrm{d}D} = S \tag{8.42}
$$

式(8.42)表明，参量 S 是信息率失真函数 $R(D)$ 的斜率。由于信息率失真函数 $R(D)$ 是 D 的单调递减的下凸函数，所以斜率 S 必为非正且递增(即 $\mathrm{d}S/\mathrm{d}D>0$)。当 D 由 $D_{\min}$ 增大至 $D_{\max}$ 时，S 的数值也随之由 $S_{\min}$ 增至 $S_{\max}$。

当 $D_{\min}=0$ 时，由式(8.38)可知，等式可变诸因子 $P(u_i)$、$P(v_j)$、λ_i 都是非负值，而各 $d(u_i, v_j)$ 也不都等于零。因此，使满足 $D_{\min}=0$ 的条件必使指数的幂为负无穷大，即 $S_{\min}$ 应该趋于负无穷，也就是 $D_{\min}=0$ 处 $R(D)$ 的斜率应该趋于负无穷。当 $D>0$ 时，S 随 D 而增加。当 D 达到 $D_{\max}$ 时，S 也达到 $S_{\max}$，但它仍是负值，当然，最大值等于零。

一般情况下，在 $(D_{\min}, D_{\max})$ 区域内 S 是 D 的连续函数。当 $D>D_{\max}$ 时，$R(D)\equiv 0$，当然 $\dfrac{\mathrm{d}R(S)}{\mathrm{d}D}=S=0$。所以在 $D=D_{\max}$ 处，除某些特例外，S 在这一点上将不是连续的，它将从某一负值跳到零。

从上面的分析过程中可以看出，对于一般的离散信源来说，采用上述方法计算 $R(D)$ 是相当复杂的。

8.4 保真度准则下的信源编码定理

本节将阐述和证明信息率失真理论的基本定理。这些定理严格地证实了 $R(D)$ 函数确实是在允许失真为 D 的条件下，每个信源符号能够被压缩的最低值。

虽然本节的讨论局限于离散无记忆信源，但所述定理可以推广到连续信源、有记忆信源等更一般的情况。

8.4.1 信源编码定理及其逆定理

定理 8.1 (保真度准则下的信源编码定理) 设 $R(D)$ 为一离散无记忆信源的信息率失真函数，并且有有限的失真测度。对于任意 $D\geqslant 0$，$\varepsilon>0$，以及任意足够长的码长 k，一定存在一种信源编码 C，其码字个数为

$$M=\mathrm{e}^{\{k[R(D)+\varepsilon]\}} \tag{8.43}$$

而编码后的平均失真度：

$$d(C)\leqslant D+\varepsilon$$

如果用二元编码，$R(D)$ 取比特为单位，则上式 M 可写成：

$$M=2^{\{k[R(D)+\varepsilon]\}}$$

该定理又称为香农第三定理。该定理告诉我们：对于任意失真度 $D\geqslant 0$，只要码长 k 足够长，总可以找到一种编码 C，使编码后每个符号的信息传输率：

$$R'=\frac{\log M}{k}=R(D)+\varepsilon \tag{8.44}$$

即

$$R'\geqslant R(D)$$

而码的平均失真度：

$$d(C)\leqslant D$$

证明：令离散无记忆信源的符号集及其概率分布为

$$\begin{bmatrix} U \\ P \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \\ P(u_1) & P(u_2) & \cdots & P(u_n) \end{bmatrix}$$

接收再现的符号 $V=[v_1, v_2, \cdots, v_s]$。对于给定的有限失真函数 $d(u_i, v_j)$，存在 $R(D)$函数。总可以找到某一试验信道$[P(v|u)]$，使达到：

$$I(U; V) = R(D) \tag{8.45}$$

并且

$$E[d(u_i, v_j)] \leqslant D \tag{8.46}$$

现考虑信源 U 输出的是长度为 k 的信源符号序列 $\boldsymbol{U}=(U_1U_2\cdots U_k)$，其中每个变量 $U_l \in U$。因而

$$\boldsymbol{U} \in U^k = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{r^k}]$$

其中：

$$\boldsymbol{\alpha}_i = (u_{i_1} u_{i_2} \cdots u_{i_k}) \quad i_1, i_2, \cdots, i_k = 1, 2, \cdots, r$$

同理，接收再现的符号序列：

$$\boldsymbol{V} \in V^k = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_{s^k}]$$

其中：

$$\boldsymbol{\beta}_j = (v_{j_1} v_{j_2} \cdots v_{j_k}) \quad j_1, j_2, \cdots, j_k = 1, 2, \cdots, s$$

于是，根据失真函数的定义有：

$$d(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j) = \sum_{l}^{k} d(u_{i_l}, v_{j_l})$$

若假设满足式(8.45)、式(8.46)的试验信道是无记忆的，则根据信源和信道都是无记忆的可得：

$$P(\boldsymbol{\alpha}_i) = \prod_{l=1}^{k} P(u_{i_l}) \tag{8.47}$$

$$P(\boldsymbol{\beta}_j \mid \boldsymbol{\alpha}_i) = \prod_{l=1}^{k} P(v_{j_l} \mid u_{i_l}) \tag{8.48}$$

$$P(\boldsymbol{\alpha}_i\boldsymbol{\beta}_j) = \prod_{l=1}^{k} P(u_{i_l} v_{j_l}) \tag{8.49}$$

$$P(\boldsymbol{\beta}_j) = \prod_{l=1}^{k} P(v_{j_l}) \tag{8.50}$$

并有：

$$R_k(D) = kR(D) \tag{8.51}$$

现在定义一种信源编码 $C=\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_M\}$，它包含 M 个码字，并属于 V^k 空间的某一子集。每个码字是长度为 k 的接收序列。当用码 C 对信源$[U, P]$的输出进行编码时，就是将信源序列集 U^k 中的每个序列 $\boldsymbol{\alpha}_i$ ——映射到码 C 中。映射的原则是将每一个信源序列 $\boldsymbol{\alpha}_i$ 变换成失真最小的那个码字 $\boldsymbol{\beta}_j$，即

$$f(\boldsymbol{\alpha}_i) = \boldsymbol{\beta}_j \quad \boldsymbol{\alpha}_i \in U^k$$

使满足：

$$d(\boldsymbol{\alpha}_i, f(\boldsymbol{\alpha}_i)) = \min_{\boldsymbol{\beta}_j \in C} d(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j) \tag{8.52}$$

因为信源编码是一一对应的，所以可以把这种编码方法看成是一个特殊的假设信道 $P_0(\boldsymbol{\beta}_j \mid \boldsymbol{\alpha}_i)$，它满足：

$$P_0(\boldsymbol{\beta}_j \mid \boldsymbol{\alpha}_i) = \begin{cases} 1 & \text{当 } \boldsymbol{\beta}_j \in C \text{ 且 } d(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j) = d(\boldsymbol{\alpha}_i, f(\boldsymbol{\alpha}_i)) \text{ 时} \\ 0 & \text{其他} \end{cases}$$

因此，根据平均失真的定义式可知，该码中每个信源符号的平均失真为

$$\begin{aligned} d(C) &= \frac{1}{k} \sum_{U^k} \sum_{V^k} P(\boldsymbol{\alpha}_i) P_0(\boldsymbol{\beta}_j \mid \boldsymbol{\alpha}_i) d(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j) \\ &= \frac{1}{k} \sum_{U^k} P(\boldsymbol{\alpha}_i) d(\boldsymbol{\alpha}_i, f(\boldsymbol{\alpha}_i)) \end{aligned} \tag{8.53}$$

现在考虑码 C 是由随机选择的码字所组成的，即每个码字都是按照式(8.50)的概率分布 $P(\boldsymbol{\beta}_j)$从 V^k 集中任意选取的，并且每个码字的选取是彼此独立的，因此码 C 的发生概率为

$$P(C) = P(\boldsymbol{\beta}_1 \boldsymbol{\beta}_2 \cdots \boldsymbol{\beta}_M) = \prod_{j=1}^{M} P(\boldsymbol{\beta}_j) \tag{8.54}$$

在随机选择信源编码的情况下，可以得到 s^{kM} 种不同的码。在这 s^{kM} 种随机编码的信源码集中，总的平均失真度为

$$\overline{d}(C) = \sum_{C=1}^{s^{kM}} P(C) d(C) = \sum_{\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_M \in V^k} \cdots \sum P(\boldsymbol{\beta}_1) P(\boldsymbol{\beta}_2) \cdots P(\boldsymbol{\beta}_M) d(C)$$

式中，$d(C)$是随机选择的每种信源编码的平均失真度，与式(8.53)相同。所以将式(8.53)代入得：

$$\overline{d}(C) = \sum_{C=1}^{s^{kM}} P(C) d(C) = \sum_{\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_M \in V^k} \cdots \sum P(\boldsymbol{\beta}_1) P(\boldsymbol{\beta}_2) \cdots P(\boldsymbol{\beta}_M) \frac{1}{k} \sum_{U^k} P(\boldsymbol{\alpha}_i) d(\boldsymbol{\alpha}_i, f(\boldsymbol{\alpha}_i)) \tag{8.55}$$

将信源序列集 U^k 分成两个子集 S 和 T，即

$$S = \{\boldsymbol{\alpha}_i : d(\boldsymbol{\alpha}_i, f(\boldsymbol{\alpha}_i)) \leqslant k(D+\delta)\}$$

$$T = \{\boldsymbol{\alpha}_i : d(\boldsymbol{\alpha}_i, f(\boldsymbol{\alpha}_i)) > k(D+\delta)\}$$

其中，δ 是大于零的任意整数。所以得：

$$\overline{d}(C) = \sum_{\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_M \in V^k} \cdots \sum P(\boldsymbol{\beta}_1) P(\boldsymbol{\beta}_2) \cdots P(\boldsymbol{\beta}_M) \frac{1}{k} \left\{ \sum_{\boldsymbol{\alpha}_i \in S} P(\boldsymbol{\alpha}_i) d(\boldsymbol{\alpha}_i, f(\boldsymbol{\alpha}_i)) + \sum_{\boldsymbol{\alpha}_i \in T} P(\boldsymbol{\alpha}_i) d(\boldsymbol{\alpha}_i, f(\boldsymbol{\alpha}_i)) \right\} \tag{8.56}$$

根据子集 S 的定义，显然有：

$$\sum_{\boldsymbol{\alpha}_i \in S} P(\boldsymbol{\alpha}_i) d(\boldsymbol{\alpha}_i, f(\boldsymbol{\alpha}_i)) \leqslant k(D+\delta)$$

令 $D_{\max}$是 $d(u_i, v_j)$的最大失真度，所以

$$\max d(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j) = \max \sum_{l=1}^{k} d(u_{i_l}, v_{j_l}) = kD_{\max}$$

显然，子集 T 中的序列 $\boldsymbol{\alpha}_i$ 满足：

$$kD_{\max} > d(\boldsymbol{\alpha}_i, f(\boldsymbol{\alpha}_i)) > k(D+\delta)$$

代入式(8.56)得

$$\bar{d}(C) \leqslant D+\delta+D_{\max}\sum_{\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_M\in V^k}\cdots\sum P(\boldsymbol{\beta}_1)P(\boldsymbol{\beta}_2)\cdots P(\boldsymbol{\beta}_M)\sum_{\boldsymbol{\alpha}_i\in T}P(\boldsymbol{\alpha}_i) \tag{8.57}$$

根据式(8.52)可知，只有当码字 $\boldsymbol{\beta}_j\in C$，$d(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)>k(D+\delta)$时，$d(\boldsymbol{\alpha}_i,f(\boldsymbol{\alpha}_i))$才大于 $k(D+\delta)$。因此，设一示性函数：

$$\Delta(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)=\begin{cases}1 & \text{当 } d(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)\leqslant k(D+\delta)\text{ 时}\\ 0 & \text{当 } d(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)>k(D+\delta)\text{ 时}\end{cases}$$

这样使式(8.57)中有限制的求和号变为无限制的求和号，即

$$\begin{aligned}\sum_{\boldsymbol{\alpha}_i\in T}P(\boldsymbol{\alpha}_i)&=\sum_{\boldsymbol{\alpha}_i\in U^k}P(\boldsymbol{\alpha}_i)[1-\Delta(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_1)][1-\Delta(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_2)]\cdots[1-\Delta(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_M)]\\&=\sum_{\boldsymbol{\alpha}_i\in U^k}P(\boldsymbol{\alpha}_i)\prod_{j=1}^{M}[1-\Delta(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)]\end{aligned}$$

于是，式(8.57)变换成：

$$\begin{aligned}\bar{d}(C)&\leqslant D+\delta+D_{\max}\sum_{\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_M\in V^k}\cdots\sum P(\boldsymbol{\beta}_1)P(\boldsymbol{\beta}_2)\cdots P(\boldsymbol{\beta}_M)\sum_{\boldsymbol{\alpha}_i\in U^k}P(\boldsymbol{\alpha}_i)\prod_{j=1}^{M}[1-\Delta(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)]\\&=D+\delta+D_{\max}\sum_{\boldsymbol{\alpha}_i\in U^k}P(\boldsymbol{\alpha}_i)\sum_{\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_M\in V^k}\cdots\sum\prod_{j=1}^{M}P(\boldsymbol{\beta}_j)[1-\Delta(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)]\end{aligned}$$

若某函数 $f(x)$是定义在集合 A 上的，则有关系式：

$$\Big[\sum_{x\in A}f(x)\Big]^M=\sum_{x_1\in A}\cdots\sum_{x_M\in A}f(x_1)\cdots f(x_M)$$

根据上式可得：

$$\begin{aligned}\bar{d}(C)&=D+\delta+D_{\max}\sum_{\boldsymbol{\alpha}_i\in U^k}P(\boldsymbol{\alpha}_i)\Big\{\sum_{\boldsymbol{\beta}_j\in V^k}P(\boldsymbol{\beta}_j)[1-\Delta(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)]\Big\}^M\\&=D+\delta+D_{\max}\sum_{\boldsymbol{\alpha}_i\in U^k}P(\boldsymbol{\alpha}_i)\Big\{1-\sum_{\boldsymbol{\beta}_j\in V^k}P(\boldsymbol{\beta}_j)\Delta(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)\Big\}^M\end{aligned} \tag{8.58}$$

现在需要估计式(8.58)中括号内求和号的上限值。因此，进一步定义示性函数：

$$\Delta_0(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)=\begin{cases}1 & \text{当 } d(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)\leqslant k(D+\delta)\text{ 且 } I(\boldsymbol{\alpha}_i;\boldsymbol{\beta}_j)\leqslant k[R(D)+\delta]\text{ 时}\\ 0 & \text{其他}\end{cases}$$

其中：

$$I(\boldsymbol{\alpha}_i;\boldsymbol{\beta}_j)=\log\frac{P(\boldsymbol{\beta}_j\mid\boldsymbol{\alpha}_i)}{P(\boldsymbol{\beta}_j)}$$

而 $P(\boldsymbol{\beta}_j|\boldsymbol{\alpha}_i)$和 $P(\boldsymbol{\beta}_j)$分别满足式(8.48)和式(8.50)。

比较所定义的两种示性函数得：

$$\Delta_0(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)\leqslant\Delta(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)$$

因此

$$\sum_{\boldsymbol{\beta}_j\in V^k}P(\boldsymbol{\beta}_j)\Delta_0(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)\leqslant\sum_{\boldsymbol{\beta}_i\in V^k}P(\boldsymbol{\beta}_j)\Delta(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)$$

当 $\Delta_0(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)=1$ 时，$\boldsymbol{\alpha}_i$ 与 $\boldsymbol{\beta}_j$ 满足：

$$P(\boldsymbol{\beta}_j)\geqslant P(\boldsymbol{\beta}_j\mid\boldsymbol{\alpha}_i)\mathrm{e}^{k[R(D)+\delta]}$$

于是，式(8.58)中括号项为

$$\Big\{1-\sum_{\boldsymbol{\beta}_j\in V^k}P(\boldsymbol{\beta}_j)\Delta(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)\Big\}^M\leqslant\Big\{1-\sum_{\boldsymbol{\beta}_j\in V^k}P(\boldsymbol{\beta}_j)\Delta_0(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)\Big\}^M$$
$$\leqslant\Big[1-\mathrm{e}^{-k[R(d)+\delta]}\cdot\sum_{\boldsymbol{\beta}_j\in V^k}P(\boldsymbol{\beta}_j\mid\boldsymbol{\alpha}_i)\Delta_0(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)\Big]^M \tag{8.59}$$

利用以下不等式：

$$(1-xy)^M\leqslant 1-x+\mathrm{e}^{-yM}\quad 0\leqslant x,\ y\leqslant 1;\ M>0$$

现令

$$y=\mathrm{e}^{-k[R(D)+\delta]},\ x=\sum_{\boldsymbol{\beta}_j\in V^k}P(\boldsymbol{\beta}_j\mid\boldsymbol{\alpha}_i)\Delta_0(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)$$

代入式(8.59)**得：**

$$\Big\{1-\sum_{\boldsymbol{\beta}_j\in V^k}P(\boldsymbol{\beta}_j)\Delta(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)\Big\}^M\leqslant 1-\sum_{\boldsymbol{\beta}_j\in V^k}P(\boldsymbol{\beta}_j\mid\boldsymbol{\alpha}_i)\Delta_0(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)+\exp[-\mathrm{e}^{-k[R(D)+\delta]}M]$$

将上式代入式(8.58)**，得：**

$$\bar{d}(C)\leqslant D+\delta+D_{\max}\Big\{1-\sum_{\boldsymbol{\alpha}_i\in U^k}\sum_{\boldsymbol{\beta}_j\in V^k}P(\boldsymbol{\alpha}_i\boldsymbol{\beta}_j)\Delta_0(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)+\exp[-\mathrm{e}^{-k[R(D)+\delta]}M]\Big\}$$
$$=D+\delta+D_{\max}\Big\{\sum_{\boldsymbol{\alpha}_i\in U^k}\sum_{\boldsymbol{\beta}_j\in V^k}P(\boldsymbol{\alpha}_i\boldsymbol{\beta}_j)[1-\Delta_0(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)]+\exp[-\mathrm{e}^{-k[R(D)+\delta]}M]\Big\} \tag{8.60}$$

首先，若取 $M=\mathrm{e}^{k[R(D)+4\delta]}$**，则当** $k\to\infty$**时，式**(8.60)**右边大括号内最后一项趋于零。只要** k **选取足够大，就可使该项小于** $\delta/D_{\max}$**，因此，式**(8.60)**成为**

$$\bar{d}(C)\leqslant D+2\delta+D_{\max}\sum_{\boldsymbol{\alpha}_i\in U^k}\sum_{\boldsymbol{\beta}_j\in V^k}P(\boldsymbol{\alpha}_i\boldsymbol{\beta}_j)[1-\Delta_0(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)] \tag{8.61}$$

其次，考虑上式的最后一项。根据 $\Delta_0(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)$**的定义可知，式**(8.61)**最后一项求和号是有限制的。它只对或者满足条件** $d(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)>k(D+\delta)$**，或者满足条件** $I(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)>k[R(D)+\delta]$**的** $\boldsymbol{\alpha}_i$ **和** $\boldsymbol{\beta}_j$ **求和。根据概率关系得：**

$$\sum_{\boldsymbol{\alpha}_i}\sum_{\boldsymbol{\beta}_j}P(\boldsymbol{\alpha}_i\boldsymbol{\beta}_j)[1-\Delta_0(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)]\leqslant P\{d(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)>k(D+\delta)\}+P\{I(\boldsymbol{\alpha}_i;\boldsymbol{\beta}_j)>k[R(D)+\delta]\} \tag{8.62}$$

因为

$$d(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)=\sum_{l}^{k}d(u_{i_l},v_{j_l})$$

它是服从同一概率分布的 k **个彼此统计独立的随机变量之和，且每个随机变量** $d(u_i,v_j)$**的均值** $E[d(u_i,v_j)]\leqslant D$**，所以，根据弱大数定理有：**

$$\lim_{k\to\infty}P\{d(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)>k(D+\delta)\}=0$$

同理，因为

$$I(\boldsymbol{\alpha}_i,\boldsymbol{\beta}_j)=\sum_{l=1}^{l}I(u_{i_l};v_{j_l})$$
$$E[I(u_{i_l};v_{j_l})]=I(U;V)=R(D)$$

所以根据大数定理又可得：

$$\lim_{k\to\infty}P\{I(\boldsymbol{\alpha}_i;\boldsymbol{\beta}_j)>k[R(D)+\delta]\}=0$$

对于足够长的 k**，可使：**

$$\begin{cases} P\{d(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j) > k(D+\delta)\} < \dfrac{\delta}{D_{\max}} \\ P\{I(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j) > k(R(D)+\delta)\} < \dfrac{\delta}{D_{\max}} \end{cases} \tag{8.63}$$

联合式(8.61)～式(8.63)，当 k 足够大时，得：

$$\overline{d}(C) \leqslant D + 4\delta$$

令 $\varepsilon = 4\delta$ 得：

$$\overline{d}(C) \leqslant D + \varepsilon$$

而此时码字个数为

$$M = \mathrm{e}^{k[R(D)+\varepsilon]}$$

所以证得在随机选择的信源编码集中，每一码集含有 $M = \mathrm{e}^{k[R(D)+\varepsilon]}$ 个码字，并且总的平均失真度 $\overline{d}(C) \leqslant D+\varepsilon$。那么，在该随机信源编码集中至少必有一个码 C，它的平均失真度 $d(C) \leqslant d(C) \leqslant D+\varepsilon$。因此，只要证得码长 k 足够长，就一定存在一种信源编码，其含有 M 个码字，而码的平均失真度 $\overline{d}(C) \leqslant D+\varepsilon$。

定理 8.2(信源编码逆定理)　不存在平均失真度为 D，而平均信息传输率 $R' < R(D)$ 的任何信源编码，亦即对任意码长为 k 的信源编码 C，若码字个数 $M < \mathrm{e}^{k[R(D)]}$，则一定有 $d(D) > D$。

逆定理告诉我们：如果编码后平均每个信源符号的信息传输率 R' 小于信息率失真函数 $R(D)$，则不能在保真度准则下再现信源的消息。下面用反证法来证明此逆定理。

证明：设存在一种信源编码 $C = \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_M\}$，$M < \mathrm{e}^{k[R(D)]}$，它使得 $d(C) \leqslant D$。信源码与信源序列 $\boldsymbol{\alpha}_i$ 的变换关系仍采用式(8.52)的方法。这种编码方法可看成是一种特殊的试验信道 $P_0(\boldsymbol{\beta}_j \mid \boldsymbol{\alpha}_i)$，它满足：

$$P_0(\boldsymbol{\beta}_j \mid \boldsymbol{\alpha}_i) = \begin{cases} 1 & \text{当 } \boldsymbol{\beta}_j \in C \text{ 且 } d(\boldsymbol{\alpha}_i, \boldsymbol{\beta}_j) = d(\boldsymbol{\alpha}_i, f(\boldsymbol{\alpha}_i)) \text{ 时} \\ 0 & \text{其他} \end{cases}$$

根据假设在这个试验信道中，可得 $d(C) \leqslant D$。又因为在这个信道中，$H(V \mid U) = 0$，所以，平均互信息：

$$I(U; V) = H(V) \leqslant \log M$$

因为信源 U 是离散无记忆信源，所以有：

$$\sum_{l=1}^{k} I(U_l; V_l) \leqslant I(U; V) \leqslant \log M$$

将上式两端分别除以 k 得：

$$\frac{1}{k}\sum_{l=1}^{k} I(U_l; V_l) \leqslant \frac{1}{k}\log M$$

设 U_l 以平均失真 $D_l \leqslant D$ 再现，则必有：

$$R(D_l) \leqslant I(U_l; V_l)$$

又根据信息率失真函数的下凸性和单调递减性得：

$$\frac{1}{k}\log M \geqslant \frac{1}{k}\sum_{l=1}^{k} R(D_l) \geqslant R\left(\frac{1}{k}\sum_{l=1}^{k} D_l\right) \geqslant R(D)$$

因此得

$$: R' = \frac{1}{k}\log M \geqslant R(D)$$

即

$$M \geqslant \mathrm{e}^{k[R(D)]}$$

这个结果与前面的假设相矛盾，所以逆定理成立。

8.4.2 编码定理的意义

保真度准则下的信源编码定理及其逆定理在实际的通信理论中有着重要的意义。这两个定理证实了允许失真限度 D 确定后，总存在一种信源编码方法，使编码的信息传输率 R' 大于 $R(D)$ 且可任意接近于 $R(D)$，而平均失真度小于允许的失真限度 D。反之，若 $R'<R(D)$，则编码的平均失真度将大于 D。如果用二元码符号来进行编码，则在允许一定失真限度 D 的情况下，平均每个信源符号所需二元码符号个数的下限值就是 $R(D)$。由香农第三定理可知，$R(D)$ 确实是允许失真限度 D 的情况下信源压缩的下限值。比较香农第一定理和香农第三定理可知，当信源给定后，无失真信源压缩的极限值是信源的信息熵 $H(U)$；有失真压缩的极限值是信息率失真函数 $R(D)$。在给定某 D 后，一般 $R(D)<H(U)$。因此，香农第三定理是有失真(限失真)信源压缩的理论基础。

另外，把香农第二定理和香农第三定理结合起来，可得信息传输的另一个重要结论，即通过某信道来传输信源输出的消息，若信道的信道容量 $C>R(D)$，则在信源和信道处用足够复杂的处理后，总能以保真度 $D+\varepsilon$ 再现信源的消息；反之，若 $C<R(D)$，则不管如何处理，在信道的接收端总不能以保真度 D 的要求再现信源的消息。

在给定信源 U 和允许的失真限度 D 后，可以求得信源的信息率失真函数 $R(D)$。如果此信源通过某信道传输，而信道的信道容量满足 $C>R(D)$，那么，根据香农第三编码定理，总可以对给定的信源 U 先进行信源压缩编码，使编码的信息传输率 $R'\geqslant R(D)$，且编码的平均失真度 $d(C)<D$。此时 R' 必须满足：

$$C > R' \geqslant R(D)$$

然后，把压缩后的信源通过信道传输，由于上述不等式左边存在，因此根据香农第二编码定理，存在一种信道编码，使压缩后的信源通过信道传输后，错误概率趋于零。因此，在接收端再现信源的消息时，总的失真或错误不会超过允许的失真限度 D。这意味着引起的失真是由信源压缩造成的，而信道传输不会造成新的失真或错误；反之，若 $C<R(D)$，则不能保证信源压缩后的信息率 $R'<C$，所以，香农第二编码定理不成立。因此，信道中引起的失真或错误无法避免，必然使得在接收端再现信源的消息时，总的失真或错误大于 D。由此可见，可以把通信系统中的信源编码和信道编码的功能完全分开。信源编码所用的码字只对给定的信源和保真度是最佳的，并不涉及信道的具体性质，也允许信道编码的应用不考虑信源的具体性质。这样就可以构成一个理想的通信系统。首先，通过信源编码器，从长的信源符号序列中去除信息冗余或不必要的精度，只留下保真度准则确定的最少、最主要的信息；然后，由信道编码器重新加入特殊形式的冗余，以提高信道传输的抗干扰性。

下面以一个简单的例子来阐明如何进行限失真的信源编码和信息率失真函数 $R(D)$ 的实用意义。

【例 8.6】 设某二元无记忆信源：

$$\begin{bmatrix} U \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

显然，此信源的信息熵为 $H(U)=1$ 比特/信源符号。根据香农第一编码定理可知，若对此

信源进行无失真编码，则每个信源符号必须用一个二元符号来表示。此时，信源输出的信息率 $R=H(U)=1$ 比特/信源符号。假设此信源在再现时允许失真存在，并定义失真函数为汉明失真，即

$$d(u_i, v_j)=\begin{cases}0 & u_i=v_j\\ 1 & u_i\neq v_j\end{cases}$$

下面对信源 U 的三次扩展信源进行编码，即把信源输出的符号序列每 3 个符号分成一组，用矢量 $\boldsymbol{U}$ 表示。显然，三次扩展信源 U^3 中有 $2^3=8$ 种可能的取值，并且是等概分布的。这里为了进行限失真压缩，我们并不打算对 U^3 中的所有 8 种可能取值分别进行编码，而仅对其中的 2 个符号序列进行编码。编码方案如下：

$$\left.\begin{aligned}\boldsymbol{u}_1&=000\\ \boldsymbol{u}_2&=001\\ \boldsymbol{u}_3&=010\\ \boldsymbol{u}_4&=100\end{aligned}\right\}\rightarrow \boldsymbol{v}_1=000 \qquad \left.\begin{aligned}\boldsymbol{u}_5&=111\\ \boldsymbol{u}_6&=110\\ \boldsymbol{u}_7&=101\\ \boldsymbol{u}_8&=011\end{aligned}\right\}\rightarrow \boldsymbol{v}_2=111$$

经过上述编码后，$\boldsymbol{v}_1$、$\boldsymbol{v}_2$ 即可直接进入信道进行传输。假设信道是无噪无损的，这时对信道而言，在输入端只有 $M'=2$ 个不同的消息，因此，完全可以用下面的编码方案来传输序列 $\boldsymbol{v}_1$ 和 $\boldsymbol{v}_2$：

$$\boldsymbol{v}_1\rightarrow 0$$
$$\boldsymbol{v}_2\rightarrow 1$$

由此可见，通过这种简单的编码方案，就可把原来要传输的三个二元符号压缩成一个二元符号。这种编码器的信息传输率为

$$R'=\frac{\log M'}{3}=\frac{1}{3}\quad \text{比特/信源符号}$$

从接收端来看，当收到码字 0 或 1 时，就可译成对应的信源序列 $\boldsymbol{v}_1$ 和 $\boldsymbol{v}_2$，$\boldsymbol{v}_1$ 和 $\boldsymbol{v}_2$ 就是再现的信源符号序列 $\boldsymbol{u}_i(i=1, 2, \cdots, 8)$。这种编译码方法可用图 8-9 直观地表示出来。

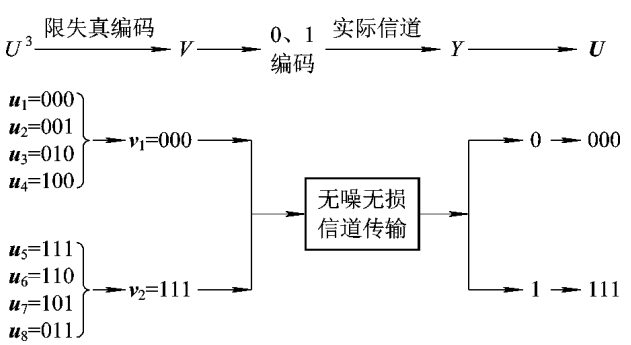

图 8-9　三次扩展信源 U 的编译码过程

显然，接收端再现的信源符号序列 $\boldsymbol{U}$ 与实际发送的信源符号序列 U 之间存在失真。这种失真是在发送端进行限失真压缩编码引起的。根据式(8.2)可计算得：

$$d(\boldsymbol{u}_1=000, \boldsymbol{v}_1=000)=d(\boldsymbol{u}_5=111, \boldsymbol{v}_2=111)=0$$

$$d(\boldsymbol{u}_2=001,\boldsymbol{v}_1=000)=d(\boldsymbol{u}_3=010,\boldsymbol{v}_1=000)$$
$$=d(\boldsymbol{u}_4=100,\boldsymbol{v}_1=000)=d(\boldsymbol{u}_6=110,\boldsymbol{v}_2=111)$$
$$=d(\boldsymbol{u}_7=101,\boldsymbol{v}_2=111)=d(\boldsymbol{u}_8=011,\boldsymbol{v}_2=111)=1$$

于是，可求得这种压缩编码的平均失真为

$$\overline{D}=\frac{1}{3}\sum_{i=1}^{8}P(\boldsymbol{u}_i)d(\boldsymbol{u}_i,\boldsymbol{v})=\frac{1}{3}\times\frac{1}{8}[0+1+1+1+0+1+1+1]=\frac{1}{4}$$

其中：

$$P(u_i)=\prod_{l=1}^{3}P(\boldsymbol{u}_{i_l})=\frac{1}{8}u_{i_l}\in U=\{0,1\}$$

可见，经过这种限失真压缩编码后信息传输率 $R'=\frac{1}{3}$ 比特/信源符号，产生的平均失真等于 $1/4$。

现在要讨论的问题是对于等概分布的二元信源 U 来说，在允许的平均失真 $D=1/4$ 时，这种压缩编码方法是否是最佳的呢？信源的输出信息率是否还可以进一步压缩呢？根据香农第三编码定理的含义可知，在允许的平均失真度 $D=1/4$ 的情况下，总可以找到一种压缩编码方法，使得信源的输出信息率压缩到极限值 $R(1/4)$，根据题意可求得：

$$R\left(\frac{1}{4}\right)=1-H\left(\frac{1}{4}\right)\approx 0.189 \quad \text{比特 / 信源符号}$$

显然，$R(1/4)<R'$。这意味着对于二元等概分布信源来说，当允许的失真限度 D 等于 $1/4$ 时，信源的输出信息率可以压缩到极限值(下限值)0.189 比特/信源符号。因此，上面讨论的这种压缩编码方法并不是最佳的，信源还可以进一步压缩。

由上面的分析可以知道，在允许一定限度失真的情况下，信源的信息率失真函数 $R(D)$ 可以用来作为衡量各种信源压缩编码方法性能的理论依据。

但是，与香农第一、二编码定理类似，香农第三编码定理同样也只是一个存在性定理，定理中并没有给出最佳的压缩编码方法。因此，有关理论的实际应用有待进一步的研究。在实际应用中，该理论主要存在着两大类问题。第一类问题是符合实际信源的 $R(D)$ 函数的计算非常困难。首先，需要对实际信源的统计特性有确切的数学描述；其次，符合主、客观实际的失真函数的定义也很困难。第二类问题是即便求出了符合实际信源的信息率失真函数，还需要去寻找最佳的编码方法，使得信源输出的信息率达到极限值 $R(D)$。目前，这两方面的工作都有进展，尤其是对实际信源(如语音信号、图像信号)的各种压缩编码方法有了较大的进展。

8.5 限失真信源编码的基本原理

8.5.1 限失真信源编码的必要性和可行性

由无失真信源编码定理和信道编码定理可知，只要信息传输率 R 大于信源的信息熵且小于信道的信道容量，即满足 $H(U)<R<C$，就总能够找到一种编码方法，使得经信道传输再现的信号中不产生失真。

但是，实际信源的输出常常是连续的消息(即取值是无限的、不可数的)，它的信息熵

$H(U)$为无穷大。如果要求无失真地传输连续信源的信息，则信息传输率 R 必为无穷大，因此要求信道的信道容量 C 也必须为无穷大。但是，显然这是无法实现的，因为对于实际的信道来说，它的带宽总是有限的，所以，其信道容量 C 总要受到限制，不可能为无穷大。因此，也就不可能实现无失真地传输信源的信息。

此外，随着信息技术的发展，由于数字系统在抗干扰、纠错、保密、适应性等方面有明显的优势，并能利用计算机技术进行信息的加工、处理和识别，因此数字系统得到了愈来愈广泛的应用。但是，将模拟信号数字化后带来的主要问题是数据量大大增加，使得占用的传输频带增宽，需要的存储空间更大。为了提高信息传输和存储系统的有效性，就必须通过对信源输出符号(序列)的编码处理实现高效的数据压缩。

实际信源输出的符号序列中，符号之间往往具有较强的相关性。例如，在模拟语音、图像信源输出的连续信号的数字化处理过程中，由于采样速率必须满足奈奎斯特定律的要求，因此采样速率一般远远高于模拟信号的变化，这使得离散数据序列中邻近数据的数值变化多是缓慢的。因此，实际离散信源输出的数据序列存在很强的相关性，可以将其近似地看做有记忆的 m 阶马尔可夫信源。由前面关于离散信源信息特性的分析可知，这种有记忆的 m 阶马尔可夫信源输出的数据序列中存在较大的信息冗余。为了提高信息系统的有效性，首先应该通过去相关处理去除信源符号之间的相关性，减小或消除信源的信息冗余，在保持信源输出信息不产生失真的前提下，通过信源编码实现数据压缩。由无失真信源编码定理可知，此时系统的信息传输率 R 可以达到的理论极限为信源的信息熵 $H(U)$。

另外，在人类社会活动和信息交流中，由于受到生理和心理特性的影响，人们往往并不要求完全无失真地再现信源输出的信息，而是要求在保证一定质量(保真度)的条件下近似地再现原来的信息，即在信息的传递和存储中允许存在一定限度的失真。例如，人耳能够感觉出的声音的频率范围为 20 Hz～20 kHz，对低于 20 Hz 的次音频信号及高于 20 kHz 的超声波，人耳是感觉不出来的。在语音通信中，如果要求高保真的 CD 音频质量，则必须传送 20 kHz 的频率分量；如果要求语音信号质量为广播质量，则需传送7 kHz 的频率分量；如果要求再现的语音信号质量为电话话音质量，则只需传送300～3400 Hz 的频率分量。显然，随着对音频数据质量要求的降低，对于占用通信信道资源的要求也将不断降低。对于图像和视频信号的传输与存储，考虑到人的视觉特性和心理特性的影响，也允许存在一定程度的失真。

于是，在去相关处理减小或消除信源信息冗余的基础上，可以进一步利用人的生理特性和心理特性，在信源编码中引入一定限度的失真，构成限失真的信源编码。信息率失真理论表明，对于给定的信源，如果允许的失真限度为 D，即满足 $\bar{d}\leqslant D$，则传输该信源输出信息时必须有的信息传输率为信息率失真函数 $R(D)$。根据香农第三编码定理，在满足失真限度 D 的条件下，高效信源编码的理论极限为信息率失真函数 $R(D)$。由于 $R(D)<H(U)$，甚至 $R(D)\ll H(U)$，因此如果采用限失真的信源编码，则能够以一定限度(即 D 内)的失真换取更加有效的数据压缩，这可大大提高信息传输和存储系统的有效性。

8.5.2　保真度准则

限失真编码的效率与失真的理论极限为给定信源的信息率失真函数 $R(D)$。但是，我们知道实际信源 $R(D)$的计算很困难。此外，符合主、客观特性的失真函数的定义也很困

难。因此，在实际工程中常用其他的主、客观评价准则来衡量限失真编码的效果。

1. 客观保真度准则

这种方法通过计算输入数据(原始数据)与输出数据(解码恢复数据)之间的误差，客观地评价两者之间的差异，以度量信源压缩编码所产生的失真大小，比较不同信源编码方法的失真性能。

计算并客观度量压缩编码失真大小的方法有多种，如均方根误差和均方根信噪比。

设信源输出数据为 $f(x)$，解码再现数据为 $g(x)$，其中：

$$x = 0, 1, \cdots, N-1$$

则可以使用均方根误差：

$$e_{\text{rms}} = \sqrt{\frac{1}{N}\sum_{x=0}^{N-1}[g(x) - f(x)]^2}$$

或均方根信噪比：

$$(\text{SNR})_{\text{rms}} = \sqrt{\frac{\sum_{x=0}^{N-1} g^2(x)}{\sum_{x=0}^{N-1}[g(x) - f(x)]^2}}$$

定量地度量编码方法产生的失真。

2. 主观保真度准则

在实际的信息传输、存储过程中，解码还原数据最终由人来使用。因此解码再现数据的失真大小、质量评价不仅与数据本身有关，而且与人的生理特性、观察者的心理状态、专业背景以及接收环境等多种因素有关。例如对于同一幅图像，不同的观察者、不同的时间、不同的地点可能会得出不同的质量评价结果。因此，对于压缩编码数据中的失真情况，由人做出主观的评价将更为合理、可靠，并且更具有权威性。

主观保真度准则是根据人们的生理和心理特性，由观察者对编码结果做出主观评价的一种方法。例如，对编码数据的主观评价可以使用品质尺度和妨碍尺度来描述。品质尺度是指由未受过训练的观察者描述数据的质量优劣感受，进行主观评价。妨碍尺度是指由训练有素的专业人员对数据进行主观质量评价，以发现数据差异的程度，评价其影响。

因为多数信息传输和存储处理的结果最终是由人接受的，这种压缩编码质量的主观评价准则也是压缩编码处理质量评价的主要方法。

8.5.3 预测编码方法

1. 基本原理

预测编码是一种直接利用数据间的相关性去相关、去冗余，实现数据压缩的信源编码技术。此时，编码传输或存储的并不是信源输出的数据本身，而是当前数据的预测值(或称为估计值)与实际值之间的差值。

设信源输出的数据 x_i 之间具有某种程度的相关性，利用这些相关性可以根据该时刻前面的某些数据做出当前数据 x_i 的预测值或估计值 $\hat{x}_i$，定义两者之间的误差：

$$e_i = x_i - \hat{x}_i$$

并将 e_i 称为预测误差或差值信号。

由于 $\hat{x}_i$ 是利用 x_i 与相邻数据之间的相关性进行估计(预测)得到的，因此，当前数据 x_i 与预测值 $\hat{x}_i$ 之间的预测误差所构成的差值信号序列 $\{e_i\}$ 中，数据之间的相关性将减小，$\{e_i\}$ 中的信息冗余将降低。

另外，由于预测值 $\hat{x}_i$ 是利用数据之间的相关性进行估计得到的，因此从统计观点来讲，多数预测值 $\hat{x}_i$ 将与当前数据的实际取值 x_i 比较接近。于是，虽然差值信号序列 $\{e_i\}$ 的数值分布区间扩大了一倍，但是预测误差 e_i 的取值将只集中在零附近的一个较小范围内，如图 8－10 所示。

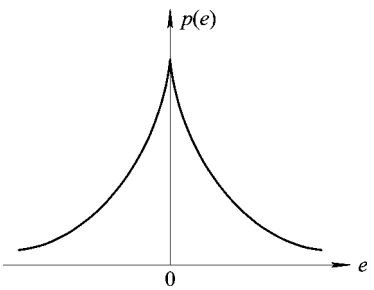

图 8－10　差值信号的概率分布

理论研究表明，e_i 的概率分布可近似为拉普拉斯(Laplace)分布，即

$$p(e)=\frac{1}{\sqrt{2}\sigma_e}\exp\left\{-\frac{\sqrt{2}}{\sigma_e}\mid e\mid\right\} \tag{8.64}$$

其中，e 为差值信号 e_i 的取值，σ_e 为差值信号的均方值。

由差值信号 e_i 的概率分布可以知道，由于其分布集中在 0 附近，因此差值信号的信息熵较小。若输入信号 x_i 已数字化，则对差值信号 e_i 可以不附加新的量化而直接编码，这便构成了一种信息保持型(无失真)的预测编码方法。但是，通常根据差值信号 e_i 的分布特点，利用人的生理特性和心理特性对 e_i 重新进行量化，以实现更大的数据压缩。由于量化产生的失真是不可逆的，因此在接收端再现的数据中将产生某种畸变，即产生了失真。将此失真控制在某一限度内，便构成了限失真的预测编码方法。

显然，这种限失真的编码方法不仅利用了数据间的统计特性，而且，在量化过程中可根据不同问题和应用要求充分利用人的生理和心理特性，实现更大的数据压缩，进一步提高系统的有效性。因此，限失真的预测编码方法是数据压缩的主要方法之一。

2. 差值脉冲编码调制

差值脉冲编码调制(Differential Pulse Code Modulation，DPCM)是一种最典型的限失真预测编码方法。DPCM 系统的原理框图如图 8－11 所示。

在接收端解码恢复的再现数据为 x_i'。于是，经过系统的传输，输入数据与输出数据之间产生的误差为

$$x_i-x_i'=x_i-(\hat{x}_i+e_i')=(x_i-\hat{x}_i)-e_i'=e_i-e_i'=q_i \tag{8.65}$$

可见，再现数据中产生的失真正是由发送端量化器中的量化误差造成的，与解码器完全无关。

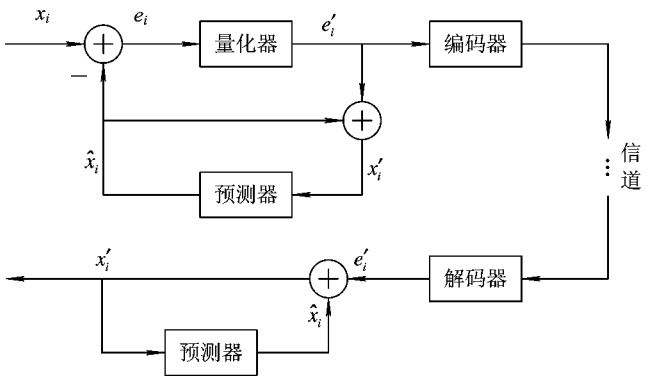

图 8-11 DPCM 系统的原理框图

在 DPCM 系统框图中，如果去掉量化器，使 $e_i=e_i'$，则 $q_i=0$，$x_i-x_i'=0$。这个结果表明，这类不带量化器的 DPCM 系统可以完全不失真地恢复原始信号 x_i，成为一个信息保持型的编解码系统。根据差值信号 e_i 的概率分布，可采用变长统计编码，降低平均编码比特数。这时传输每一预测误差所需的平均比特数的理论下限为差值信号序列 e_i 的熵值。

3. 线性预测

假定信源输出的随机变量序列$\cdots x_1\ x_2\cdots\ x_i\cdots$是一个平稳的 m 阶马尔可夫过程。由预测编码基本原理可知，预测是减小数据之间的相关性，去除信源信息冗余，提高编码效率的技术手段，也是预测编码中基本的处理步骤。因此，如何充分利用数据之间的相关性，由已经接收到的信源符号序列 $x_1x_2\cdots x_{N-1}x_N$ 计算 $\hat{x}_N=\sum_{i=1}^{k}a_ix_{N-i}$，得到信源符号 x_N 的准确的预测值，是预测编码技术的主要研究内容。

依据最佳线性预测理论，可以计算相邻数据 $x_i(i=1,2,\cdots,N-1)$的协方差，以均方误差(MSE)极小为准则，求解得到 $N-1$ 个最佳预测系数 a_i，进而得到 x_i 的最佳线性预测值 $\hat{x}_N$。显然，最佳线性预测能够充分利用数据之间的相关性，得到 x_i 的最佳线性预测值，是一种理想的线性预测方案。但是，由于此方法需要的计算量大，因此工程应用中需要结合具体问题和性能要求限定数据的相关长度，构成实用的线性预测方案。

常用的预测方案有下面几种。

1) 前值预测

这种预测是一种最简单的预测方案。假定输入数据序列 $x_1x_2\cdots x_{N-1}x_N$ 是一个一阶的马尔可夫过程。数据 x_N 的预测值可由其最接近的前一个数据 x_{N-1}做出，即

$$\hat{x}_N=ax_{N-1}\quad(a\text{ 为待预测系数})$$

显然，前值预测仅利用了最邻近的一个数据的相关性，预测误差较大，数据压缩编码的效果较差。

2) 一维线性预测

假定输入数据序列 $x_1x_2\cdots x_{N-1}x_N$ 是一个 k 阶的马尔可夫过程，数据 x_N 的预测值可由其前面 k 个数据做出，即

$$x_N=\sum_{i=1}^{k}a_ix_{N-i}$$

$$\hat{x}_N = a_1 x_{N-1} + a_2 x_{N-2} + \cdots + a_k x_{N-k}$$

其中，阶数 k 常根据要求和设备能力决定；为避免求解最佳线性预测系数的繁杂运算，常由实验确定 a_i，$i=1, 2, \cdots, k$。

一维预测可用于语音数据序列，也可用于图像数据序列。对于图像数据，一维预测是利用同一扫描行中前面 k 个数据进行预测，由于它未利用不同扫描行之间数据的相关性，因此一维线性预测方案去除图像信源信息冗余的能力不够理想。

3）二维线性预测

对于静止图像和视频信源，为了更充分地利用数据间的相关性，应当考虑相邻行、列以及相邻帧数据之间的相关性，构成二维或三维的线性预测。

设输入图像为 $x(m, n)$，其线性预测值为

$$\hat{x}(m, n) = \sum_{(k, l) \in Z} a_{k, l} x(m-k, n-l)$$

相应的预测误差为

$$\begin{aligned} e(m, n) &= x(m, n) - \hat{x}(m, n) \\ &= x(m, n) - \sum_{(k, l) \in Z} a_{k, l} x(m-k, n-l) \end{aligned}$$

其中，Z 为像素 $x(m, n)$ 的相关点构成的集合。

从理论上讲，Z 应当包括与 $x(m, n)$ 相关性较大的像素点，如图 8-12 所示。但是在图像、视频编码中往往需要考虑像素出现的顺序及计算的因果性，即应由已接收再现的数据计算当前数据的预测值。为了满足实时处理的应用要求，应当将图 8-12(a)所示的非因果型相关邻域改变为图 8-12(b)所示的符因果型相关邻域。

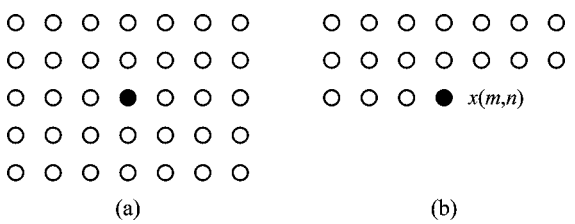

图 8-12　非因果型和符因果型相关邻域

在常用的线性预测方法中，预测系数 $a_{k, l}$ 可以在最小均方误差准则下求出。但为了减小计算量，多由实验确定得到。

4. 量化编码

前面已经指出，差值信号 e_i 呈拉普拉斯分布，利用此分布的特点，经过量化，编码便可使总的比特率减小，实现数据压缩。

若输入信号 x_i 已数字化，则差值信号 e_i 可能也是离散的。因此，可以通过对差值信号 e_i 直接进行无失真的编码(如霍夫曼编码)，以实现无失真的数据压缩。

有误差的量化、编码方法如下：

一般的图像、语音等信息传输系统对差值信号 e_i 重新进行量化后编码，并且往往依据人的生理特性和心理特性，在量化和编码中引入一定限度的失真，以实现更有效的数据压缩。常用的量化方法包括均匀量化、非均匀量化等。对于要求较高的 DPCM 系统，需采用

最佳量化器对 e_i 进行量化。

关于量化器的设计在此不作讨论。

8.5.4 正交变换编码方法

在信源编码数据压缩处理中，还有另一类基本方法，即正交变换编码。正交变换编码通过施加某种正交变换，使得变换域中数据的相关性减小或消除，从而减小信源的信息冗余，实现数据压缩。

1. 正交变换的物理意义

正交变换是信号分析与信息处理中常用的处理工具。此处针对正交变换在信源编码中的应用，以一个简单的例子对正交变换能够减小数据相关性的物理意义做一个定性分析。

设信源输出的数据为 0，1，…，7，信源输出的随机变量序列…x_1 x_2… x_i…中，相邻数据 x_1 和 x_2 构成一个联合事件(x_1, x_2)。联合事件(x_1, x_2)可以用图 8－13 所示坐标系中的点表示。显然，联合事件(x_1, x_2)的组合方式共有 $2^3 \times 2^3 = 64$ 种。

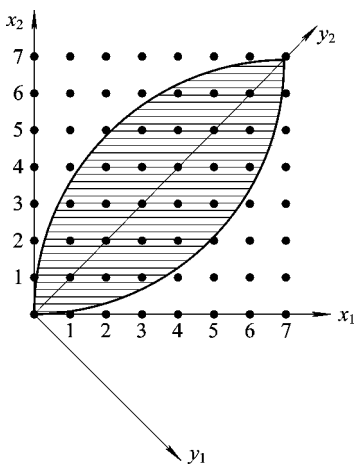

图 8－13 正交变换示意图

由于 x_1 和 x_2 是信源输出符号序列中的相邻数据，因此它们之间存在较大的相关性，即后一数据 x_2 与前一数据 x_1 的取值相近或相等的可能性较大。这一特点表明，图 8－13 所示阴影区中的点(联合事件)发生的概率较大。

对信源输出的数据实施正交变换，从几何关系上看相当于坐标系的变换。此处我们考虑将 x_1、x_2 坐标系旋转 45°，使其成为 y_1、y_2，坐标系如图 8－13 所示。

在 y_1、y_2 坐标系中，对于发生概率较大的联合事件(阴影区中的点)，前一数据 y_1 的取值在 0 附近的可能性大，而其后相邻数据 y_2 的取值在 0～7 之间的可能性都比较大。可见，坐标旋转后，联合事件(y_1, y_2)中，分量 y_2 的取值对分量 y_1 的取值大小的依赖性变小了。

由此可以看出，通过坐标旋转变换，减小了相邻数据之间的相关性，即将相关性强的数据(x_1, x_2)转换成了相关性较弱的数据(y_1, y_2)。还可看出，虽然坐标旋转变换前后各数据的方差总和可能不变，即

$$\sigma_{x_1}^2 + \sigma_{x_2}^2 = \sigma_{y_1}^2 + \sigma_{y_2}^2$$

但是坐标旋转变换前与旋转变换后相比，坐标轴上的方差分布却有很大的差别，即

$$\sigma_{x_1}^2 \approx \sigma_{x_2}^2$$

而且

$$\sigma_{y_2}^2 \gg \sigma_{y_1}^2$$

由于方差是数据随机性的一种表征，而随机性是信息的一个基本属性，因此方差大表明数据的随机性强，方差大的随机变量载荷的信息量大。因此，经过简单的坐标旋转变换，在 y_1、y_2 坐标系中数据的随机性变得不均等，能量分布趋于集中。

数据方差分布的不均匀性是正交变换编码能够实现数据压缩的基本条件之一。

2．正交变换压缩编码的一般原理

设信源输出的具有某种程度相关性的时间域或空间域数据可以表示为列向量：

$$\boldsymbol{f} = [f(1),\ f(2),\ \cdots,\ f(N)]$$

施加正交变换之后，其变换域系数序列为

$$\boldsymbol{F} = [F(1),\ F(2),\ \cdots,\ F(N)]$$

正交变换编码能够实现数据压缩，是因为正交变换具有下面三个特点：

(1) 能量保持。经正交变换后，时间域或空间域中的总能量在变换域中得到保持，即若

$$[f(x)] \leftrightarrow [F(u)] \quad x,\ u = 0,\ 1,\ \cdots,\ N-1$$

则由巴塞瓦公式知：

$$\sum_{x=0}^{N-1} f^2(x) = \sum_{u=0}^{N-1} F^2(u)$$

(2) 能量重新分布与集中。相对于相关的输入数据序列 $f(1)$，$f(2)$，…，$f(N)$，变换域中的能量多集中在与直流和低频分量相对应的少数变换系数中。

(3) 去相关。在变换域中，变换域系数 $F(1)$，$F(2)$，…，$F(N)$之间的相关性为零或较小。

因此，正交变换所具有的去相关、能量保持和能量重新分布与集中三个基本特性为数据压缩创造了有利的条件。在对信源输出的时间域或空间域数据施加正交变换之后，可以使用较少的变换域系数替代原空间域数据，实现数据压缩。于是，我们可以得到正交变换编码数据压缩的一般方法。

此方法首先对输入数据序列施加某种形式的正交变换，在变换域中设计某种形式的滤波器以取出信源信息中的主要部分，而将次要部分舍弃，从而实现数据压缩。

变换域区域编码方法的原理框图如图 8－14 所示。

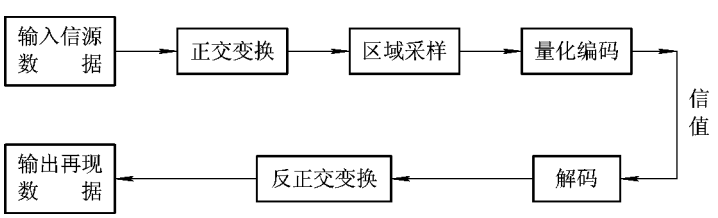

图 8－14　变换域区域编码方法的原理框图

信源的压缩编码中可以使用多种形式的正交变换。其中，K－L 变换具有变换域系数之间的相关性为 0 和变换域中的能量分布最集中的最佳性能，因此被称为最佳变换。但是由于 K－L 变换需要的计算量大，难以付诸实用，因此工程中常使用具有快速算法的准最

佳正交变换，如离散傅立叶变换(DFT)、离散余弦变换(DCT)、沃尔什-哈德玛变换(W-HT)等。

3. 方差准则

由于正交变换使得能量在变换域中重新分布与集中，因此依据正交变换编码的基本原理和处理框图可以知道，实现高效的正交变换数据压缩的关键在于：如何选择样本(变换域系数)才能取出信源信息中的主要部分，而使舍弃的系数中含有的信息量最少，即丢失(损失)的信息量最少。在此我们给出一个变换域系数选择中应当遵循的方差准则。

首先分析使用最佳变换(K-L 变换)时的情况。设

$$[f(x)]_N \leftrightarrow [F(u)]_N$$

为 N 点 K-L 变换，经过 K-L 变换后，变换域系数的协方差矩阵是对角线矩阵，即

$$[C_F] = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & 0 \\ & & \ddots & \\ & 0 & & \lambda_{N-1} \\ & & & & \lambda_N \end{bmatrix}$$

并且满足：

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_N$$

希望从 $[F(u)]_N$ 中选出一个由 M 个分量组成的子集($M<N$)，当删去剩下的 $N-M$ 个分量，仅用 $[F(u)]_N$ 中所保留的 M 个分量重建数据 $[f(x)]_N$ 时，所产生的均方误差最小。

由 K-L 变换的能量集中特性可知，应当选出具有最大特征值的 M 个系数并构成保留子集，即

$$F(1),\ F(2),\ \cdots,\ F(M)$$

而舍弃其他变换域系数

$$F(M+1),\ F(M+2),\ \cdots,\ F(N)$$

特征值 λ_i 是 $[C_F]$ 主对角线上的元素，它们实际上是变换系数 $F(u)$ 的方差，即

$$\sigma_F^2(i) = \lambda_i \qquad i = 1,\ 2,\ \cdots,\ N$$

对于离散傅立叶变换(DFT)、离散余弦变换(DCT)、离散沃尔什-哈德玛变换等常用准最佳变换，协方差矩阵 $[C_F]$ 一般不是严格的对角线矩阵。因此，经准最佳正交变换处理后，变换域系数之间仍保留一些剩余相关性，但是对于准最佳变换，变换域中的能量仍集中于方差大的变换域系数中。于是我们可以得出正交变换编码中选择变换域系数应当遵循的方差准则：保留具有最大方差的 M 个系数，其余($N-M$)个系数可以舍弃。

4. 正交变换编码的基本方法

设 N 点数据组成的列向量 $[f(x)]$、$[F(u)]$ ($x,\ u=0,\ 1,\ \cdots,\ N-1$)为一对正交变换，即对于信源数据列向量：

$$\boldsymbol{f} = [f(1),\ f(2),\ \cdots,\ f(N)]$$

施加正交变换后，变换域系数列向量为

$$\boldsymbol{F} = T[f(x)] = [F(1),\ F(2),\ \cdots,\ F(N)]$$

逆变换为

$$\boldsymbol{f} = T^{-1}[\boldsymbol{F}]$$

由上面的讨论知道，正交变换信源编码是在变换域内进行样值选择和量化编码。变换域内的样值选择和量化编码主要有两种方法。

1）区域采样法

区域采样法为利用变换域中能量集中于低频区域的特点，设计一个二维低通滤波器，即令变换域抽样函数 $\psi(u)$ 为

$$\psi(u) = \begin{cases} 1 & \text{低通取样域内} \\ 0 & \text{低通取样域外} \end{cases}$$

于是低频分量的变换域系数被保留并编码传输，而其他分量被舍弃。

在接收端，解码得到变换域系数，按照 $\psi(u)$ 的关系适当排放并令低通取样区域外的系数为零，重建变换域系数阵列，通过反变换重建空间域数据：

$$f'(x) = T^{-1}[F(u)]$$

由于在变换域做了区域采样，因此依逆变换关系得到的重建数据为

$$f'(x) = \sum_u F(u)\psi(u)T^{-1}(x;\ u)$$

显然，因为区域采样舍弃了部分变换域系数（丢掉了部分交流信息），所以解码再现数据将存在一定程度的失真。

可以证明，重建数据 $f'(x)$ 与输入数据 $f(x)$ 之间的均方误差为

$$\varepsilon = \frac{1}{N}\sum_u F^2(u)[1-\psi(u)]$$

显然，信源输出数据和解码再现数据之间的均方误差的大小与正交变换形式在变换域中的能量集中程度有关。

2）区域编码法

在区域采样的基础上，依变换域系数方差的分布特点，将低通区域分为若干子区域，对不同区域内的变换域系数用不同的比特数进行量化和编码，这样就构成了区域编码。

区域编码法可进一步节省数码率，实现更有效的数据压缩。

8.5.5　矢量量化编码方法

1. 问题的提出

限失真编码中，需要对信源输出符号或去相关处理后的数据进行量化，主要的量化方式包括标量量化（对单个符号进行量化处理）和矢量量化（对随机变量序列进行量化处理）。

前面讨论信源数据压缩编码中采用的是标量量化方法，即对信源进行去相关处理后，对其输出的单个符号数据（标量）进行量化和编码。标量量化编码方法处理过程简单，但是存在下面两方面的问题：

（1）标量量化不能有效利用数据之间的剩余相关性。

标量量化的基本假设是数据相互独立。然而，对于具有一定相关性的信源，经 DFT、DCT 等准最佳正交变换或预测处理后，数据之间仍然存在一定的剩余相关性。由于矢量量化是对某数据序列进行编码，编码中考虑了数据序列的整体变化关系，因此可以更充分地利用数据序列中各分量之间的剩余相关性，进一步提高系统的有效性。

另外，香农第一、第三编码定理表明：对扩展信源输出的 p 个信源符号组成的随机矢量进行编码，其编码效率优于标量的编码，并且 p 愈大，编码效率愈高。

可见，信源输出数据的矢量量化编码可以进一步提高数据压缩的有效性。

(2) 标量量化的失真控制策略不够合理。

在限失真编码中，失真大小是衡量编码方法优劣的重要指标。标量量化可以使每一随机变量的量化失真最小。但是，这种失真控制策略不能保证量化结果与数据序列的变化规律一致，从而使得信源数据序列的总体量化失真较大。

2. 矢量量化的基本原理

设信源数据 X 取值于一维的实数空间，且 $X\in\{x_1, x_2, \cdots, x_L\}$。

在信源输出的随机变量序列 $X_1X_2\cdots X_i\cdots$ 中，将 p 个数据划分为一组并构成 p 维的随机矢量：

$$\boldsymbol{X}^p = X_1X_2\cdots X_p$$

其中：

$$X_i \in X \quad i = 1, 2, \cdots, p$$

于是，可以将信源看做一个输出 p 维随机矢量 $\boldsymbol{X}$ 的扩展信源 $\boldsymbol{X}^p$。扩展信源的输出 $\boldsymbol{X}$ 有 L^p 种可能的取值组合形式(样矢量)，每一个样矢量 $\boldsymbol{X}_i(i=1, 2, \cdots, L^p)$ 对应于 p 维实空间中的一个点。

定义 p 维随机矢量的失真函数(如样矢量之间的距离)。如果某些点之间的距离较小，则认为它们的样矢量比较相似(接近)，可以将这些样矢量(点)划分为一类，于是，可以将 p 维实空间划分为 N 类($N\ll L^p$)子空间。在每一类子空间中总可以找到一个点(矢量)，使得该子空间中的样矢量距该点的距离(失真)均比较小。定义此点为该类样矢量的类中心矢量(也称典型矢量、代表矢量)。

于是，信源输出的 L^p 种 p 维样矢量可以由 N 种类中心矢量来表示并组成矢量量化码本，码矢地址可以用每一典型矢量的序号表示。

在信源输出信息的传输或存储系统中，仅对典型矢量的序号 i(码矢地址)进行编码。在解码端则依据码矢地址查找典型矢量，恢复信源输出的数据。如果典型矢量序号 i 用等长码表示，则传输一个样矢量只需 $\log N$ 比特。于是，实现的压缩比为

$$\log\frac{L^p}{\log N}$$

显然，矢量量化可以实现高效的数据压缩。

注意：将每一样矢量用其类中心矢量表示时，重建信号将产生少量的失真。因此，矢量量化编码是一种限失真的信源编码。

3. 矢量量化的处理步骤

首先依据某种准则(如失真度量、距离、概率分布)和系统的性能要求，将样矢量 $\boldsymbol{X}^p$ 构成的 p 维空间进行划分，设计矢量量化码本。

设 p 维矢量 $\boldsymbol{X}^p$ 可以划分为 N 类，如将 p 维空间划分为 $S_1, S_2, \cdots, S_N(N\ll L^p)$ 等 N 个子空间，N 个子空间两两互不相交，且子空间之和为 p 维矢量定义的全空间，则这样的子空间是 p 维空间中的一个超多面体。

按照某种失真准则，通过优化可以在每一子空间 S_i 内确定一个类中心矢量 $\boldsymbol{y}_i(i=1, 2, \cdots, N)$。在该子空间中，将每一个样矢量用类中心矢量表示时所产生的平均失真最小。如果 $\boldsymbol{X}^p \in S_i$，则将该样矢量 $\boldsymbol{X}^p$ 用类中心矢量 $\boldsymbol{y}_i$ 表示，即子空间 S_i 中所有样矢量都被量化为类中心矢量 $\boldsymbol{y}_i$。

矢量量化码本由 N 个子空间中的类中心矢量(p 维)构成，即

$$\{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N\}$$

N 个类中心矢量分别以数字 1、2、…、N 为码矢量地址并编码表示。

在发送端，每一 p 维样矢量被量化为类中心矢量，以其序号的编码在信道中传输。

在接收端，解码恢复某一数字，在码本中找到相应的类中心矢量，还原 p 维矢量，重建信源数据序列。

矢量量化处理的一般步骤如下：

(1) 将信源输出的随机变量序列 $X_1X_2\cdots X_i\cdots$ 分组，每一组为一个 p 维随机矢量 $\boldsymbol{X}^p = X_1X_2\cdots X_p$。

(2) 确定码本的长度 N，指定一组初始码矢量：

$$\boldsymbol{Y}_0 = \{\boldsymbol{y}_1^{(0)}, \boldsymbol{y}_2^{(0)}, \cdots, \boldsymbol{y}_N^{(0)}\}$$

(3) 依某种准则迭代处理，优化 p 维空间的划分和类中心矢量。通过迭代优化码本 $\boldsymbol{Y}_i$，使 $\boldsymbol{Y}_i$ 所产生的失真度满足迭代门限要求，得到优化的码本 $\boldsymbol{Y}=\{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_N\}$ 和优化的子空间划分 $S_1, S_2, \cdots, S_N$。

(4) 矢量量化。对于信源输出的每一个样矢量 $\boldsymbol{X}^p$，通过与码本 $\boldsymbol{Y}$ 匹配(依某种失真度量)来确定该样矢量 $\boldsymbol{X}^p$ 的类别归属。若 $\boldsymbol{X}^p \in S_i$，则 $\boldsymbol{X}^p$ 被量化为 $\boldsymbol{y}_i$，即 $\boldsymbol{X}^p$ 被 $\boldsymbol{y}_i$ 取代表示。

(5) 编码。对 $\boldsymbol{y}_i$ 的序号 i 作等长或变长编码，进入传输系统。在接收端，收到的码字经解码恢复序号 i，在码本找出相应的码矢量 $\boldsymbol{y}_i$，以 $\boldsymbol{y}_i$ 作为 $\boldsymbol{X}^p$ 的解码，重建信源数据序列。

习　题　8

8.1　已知信源 $[u, P(u)]=\begin{bmatrix} u_1 & u_2 & u_3 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$ 和失真函数矩阵 $\boldsymbol{D}=\begin{bmatrix} 0 & 1 & 3 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{bmatrix}$，在下面不同的信道中计算平均失真 $\bar{d}$ 和互信息 $I(u; v)$。

(1) $[P(v|u)]=\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$；

(2) $[P(v|u)]=P(v)$ 且 $P(v_1)=\frac{1}{3}$，$P(v_2)=\frac{1}{6}$，$P(v_3)=\frac{1}{2}$；

(3) $[P(v|u)]=P(v)$ 且 $P(v_1)=P(v_3)=0$，$P(v_2)=1$。

由计算结果可得到什么结论？

8.2　已知 $[U, P(u)]=\begin{bmatrix} u_1 & u_2 \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}$，接收符号集为 $V=\{v_1, v_2\}$，失真矩阵为 $\boldsymbol{D}=\begin{bmatrix} \frac{1}{8} & 1 \\ 1 & \frac{1}{2} \end{bmatrix}$，

计算 $D_{\min}$ 和 $D_{\max}$，依 $R(D)$ 的性质做出大致的函数曲线。

8.3 已知二元信源 X：$\begin{bmatrix} X \\ P \end{bmatrix}=\begin{bmatrix} 0, & 1 \\ p, & 1-p \end{bmatrix}$ 以及失真矩阵 $[d_{k_j}]=\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$，试求 $D_{\min}$、$D_{\max}$ 和 $R(D)$。

8.4 若某无记忆信源 $[U, P(u)]=\begin{bmatrix} -2 & 0 & 2 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$，$V=\{-1, 1\}$，根据实际情况的要求和信道错误的风险定义失真函数 $\boldsymbol{D}=\begin{bmatrix} 1 & 4 \\ 2 & 1 \\ 4 & 1 \end{bmatrix}$，求传输此信源信息时的最小失真限度 $D_{\min}$ 和最大失真限度 $D_{\max}$，并求出如何选择信道可达到 $D_{\min}$ 和 $D_{\max}$。

8.5 已知 $[U, P(u)]=\begin{bmatrix} 0 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}$，$V\in\{0, 1\}$，失真矩阵 $\boldsymbol{D}=\begin{bmatrix} 0 & a \\ a & 0 \end{bmatrix}(a>0)$，计算给定信源 U 的信息率失真函数 $R(D)$ 并验证信道的存在性。

8.6 一个四进制等概率信源 $[U, P(u)]=\begin{bmatrix} 0 & 1 & 2 & 3 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$，接收符号集 $V=\{0, 1, 2, 3\}$，其失真矩阵 $\boldsymbol{D}=\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$，计算信源的 $R(D)$，并画出其曲线（取 4～5 个点）。

8.7 设已知离散无记忆信源在给定失真量度 $d(k, j)(k=1,2,\cdots,K;\ j=1,2,\cdots,J)$ 下的信息率失真函数为 $R(D)$，现定义新的失真量度 $d'(k, j)=d(k, j)-g_k$。试证：在新的失真量度下，信息率失真函数 $R'(D)$ 为 $R'(D)=R(D+G)$，其中 $G=\sum_k P(a_k)g_k$。

8.8 令某离散无记忆信源 U 的信息率失真函数为 $R(D)$，而失真函数为 $\{d(u, v): u\in U, v\in V\}$。研究另一种测度：

$$\{d'(u, v)=d(u, v)-\min_{v\in V} d(u, v): u\in U, v\in V\}$$

它相当于失真矩阵中某一行各元素减去该行中的最小值，这时在失真测度 $d'(u, v)$ 下的信息率失真函数为 $R'(D)$。证明：

$$R'(D)=R(D+D_{\min})$$

其中：$D_{\min}=\sum_U P(u_i)\min d(u_i, v_j)$。

附录 A　凸函数与颜森(Jensen)不等式

A.1　一元函数的凸性

凸性是反映函数变化关系的一种基本性质。在信息理论的讨论中，凸性是描述信息度量关系的一个很有意义的性质。下面首先由一元函数的几何意义，观察和了解函数的凸性。

设有一元函数 $f(x)$如图 A－1 所示，由 $f(x)$的几何关系可以看出，此函数是下凸的。我们需要找出函数凸性的一般的描述方式。

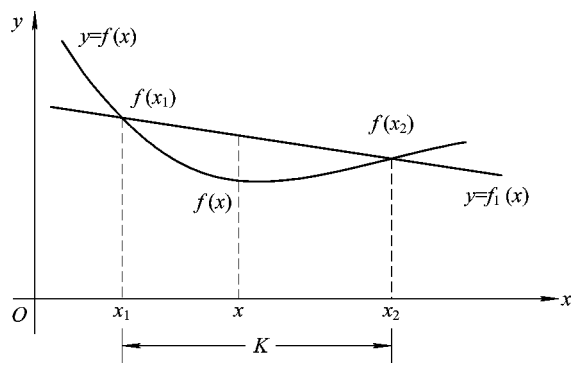

图 A－1　一元下凸函数

若 K 是 $f(x)$定义域中一个连续的子区间，则在 K 中任取两点 x_1 和 x_2(不妨设 $x_1<x_2$)，对于任意的 $0\leqslant t\leqslant 1$，它们的线性组合 $x=tx_1+(1-t)x_2\in K$，且满足 $x_1\leqslant x\leqslant x_2$。为了观察函数 $f(x)$的凸性，由 x_1、x_2 处的函数值 $f(x_1)$、$f(x_2)$做 $f(x)$的弦(割线)$f_1(x)$(如图 A－1 中的直线)。此弦的方程 $f_1(x)$可以由两点式写出，即

$$\frac{f_1(x)-f(x_1)}{x-x_1}=\frac{f(x_2)-f(x_1)}{x_2-x_1}$$

由此得到：

$$\begin{aligned}f_1(x)&=\frac{f(x_2)-f(x_1)}{x_2-x_1}(x-x_1)+f(x_1)\\&=\frac{x_2-x}{x_2-x_1}f(x_1)+\frac{x-x_1}{x_2-x_1}f(x_2)\end{aligned}$$

作变量替换，令 $t=\dfrac{x_2-x}{x_2-x_1}$，由于 $x_1\leqslant x\leqslant x_2$，因此 $0\leqslant t\leqslant 1$。因 $\dfrac{x-x_1}{x_2-x_1}=1-\dfrac{x_2-x}{x_2-x_1}=1-t$，

即

$$x = tx_1 + (1-t)x_2$$

于是，弦 $f_1(x)$ 的直线方程为

$$f_1(x) = tf(x_1) + (1-t)f(x_2)$$

由图 A-1 所示的函数曲线的变化关系可以知道，在 $x_1 \leqslant x \leqslant x_2$ 中，$f(x)$ 上任意两点 x_1、x_2 处所做的弦都在 $f(x)$ 的上方，即

$$f_1(x) = tf(x_1) + (1-t)f(x_2) \geqslant tf(x) + (1-t)f(x) = f(x)$$

故而有 $f(x) \leqslant f_1(x)(x \in K)$。

A.2 一元函数凸性的定义

对于定义域中任意两点 x_1、x_2，如果 $x_1 \leqslant x \leqslant x_2$，曲线 $f(x)$ 的弦 $f_1(x)$ 都在 $f(x)$ 的上方，即满足：

$$f(tx_1 + (1-t)x_2) \leqslant tf(x_1) + (1-t)f(x_2)$$

则称函数 $f(x)$ 在区间 $[x_1, x_2]$ 内是下凸的，仅当不等式成立时为严格下凸。若不等号反向，则 $f(x)$ 是上凸或严格上凸的。

A.3 函数凸性的判别

定义在 K 上的一元函数 $f(x)$ 其凸性可以由其一阶、二阶导数加以判别。

(1) 若 $f(x)$ 的一阶导数 $f'(x)$ 在 K 上存在且非递减，则 $f(x)$ 下凸；若 $f'(x)$ 递增，则 $f(x)$ 严格下凸。

反之，$f(x)$ 上凸或严格上凸。

(2) 若 $f(x)$ 的二阶导数 $f''(x)$ 在 K 上存在且满足 $f''(x) \geqslant 0$，则 $f(x)$ 下凸。只有不等式成立时，$f(x)$ 严格下凸。

反之，$f''(x) \leqslant 0$ 时 $f(x)$ 上凸或严格上凸。

(3) 定理　设函数 $f(x)$ 在区间 $[a, b]$ 上可导，则下面条件等价：

① 为 $[a, b]$ 上的上凸函数；

② 为 $[a, b]$ 上的增函数；

③ 对 $[a, b]$ 上的任意两点 x_1、x_2，有：

$$f(x_2) \geqslant f(x_1) + f'(x_1)(x_2 - x_1)$$

若 $f(x)$ 为区间 $[a, b]$ 上的下凸函数，则不等号反向。

A.4 Jensen 不等式

设函数 $f(x)$ 为区间 $[a, b]$ 上的上凸函数，则对任意 $x_i \in [a, b]$，$\lambda_i > 0$，$\sum_{i=1}^{n} \lambda_i = 1$，$i = 1, 2, \cdots, n$，有 Jensen 不等式：

$$f\Big(\sum_{i=1}^{n}\lambda_i x_i\Big)\leqslant\sum_{i=1}^{n}\lambda_i f(x_i)$$

等号当且仅当 $x_1=x_2=\cdots=x_n$ 时成立。若 $f(x)$为区间$[a,\ b]$上的下凸函数，则不等号反向。

证明：假设 $x_0=\sum_{i=1}^{n}\lambda_i x_i$，则 $x_0\in[a,\ b]$。根据上述定理，对于任意 $x_i\in[a,\ b]$，有

$$f(x_i)\geqslant f(x_0)+f'(x_0)(x_i-x_0)$$

在不等号两边同乘以 $\lambda_i>0$，有

$$\lambda_i f(x_i)\geqslant\lambda_i f(x_0)+\lambda_i f'(x_0)(x_i-x_0)$$

对不等号两边求和，可得：

$$\begin{aligned}\sum_{i=1}^{n}\lambda_i f(x_i)&\geqslant\sum_{i=1}^{n}\lambda_i f(x_0)+\sum_{i=1}^{n}\lambda_i f'(x_0)(x_i-x_0)\\&=f(x_0)\sum_{i=1}^{n}\lambda_i+f'(x_0)\Big(\sum_{i=1}^{n}\lambda_i x_i-x_0\sum_{i=1}^{n}\lambda_i\Big)\\&=f(x_0)+f'(x_0)(x_0-x_0)\\&=f\Big(\sum_{i=1}^{n}\lambda_i x_i\Big)\end{aligned}$$

即

$$\sum_{i=1}^{n}\lambda_i f(x_i)\geqslant f\Big(\sum_{i=1}^{n}\lambda_i x_i\Big)$$

证毕。

A.5　凸域和凸函数

上面关于一维欧氏空间中函数凸性的结论可以推广到 n 维欧氏空间中。

1）凸域

凸域表示欧氏子空间的一种性质，其定义如下：

设有 n 维欧氏空间的子空间 K，对于 K 中任意两个矢量(点)$\boldsymbol{x}_1=(x_1^1,\ x_1^2,\ \cdots,\ x_1^n)$，$\boldsymbol{x}_2=(x_2^1,\ x_2^2,\ \cdots,\ x_2^n)$，若它们的线性组合矢量：$\boldsymbol{x}=t\boldsymbol{x}_1+(1-t)\boldsymbol{x}_2$ 仍在 K 中，则称子空间 K 是凸状的或凸域。

2）凸函数

n 维空间中凸函数的定义可以仿照一维函数凸性的定义推广给出。

对于 n 维空间中的实函数 $f(\boldsymbol{x})$，如果 K 是 $f(\boldsymbol{x})$定义域上的一个凸的子集，那么对于任意的 $\boldsymbol{x}_1$，$\boldsymbol{x}_2\in K$，$t\in[0,\ 1]$，若满足：

$$f(t\boldsymbol{x}_1+(1-t)\boldsymbol{x}_2)\leqslant tf(\boldsymbol{x}_1)+(1-t)f(\boldsymbol{x}_2)$$

则称 $f(\boldsymbol{x})$为下凸函数(杯型)。若只有不等式成立，则称 $f(\boldsymbol{x})$是严格下凸的。

若不等号反向，则 $f(\boldsymbol{x})$为上凸函数(帽型)或严格上凸函数。

此处将 $f(\boldsymbol{x})$的定义域局限在凸域，这是为了保证矢量 $\boldsymbol{x}=t\boldsymbol{x}_1+(1-t)\boldsymbol{x}_2$ 仍在域 K 中。

3）凸函数的性质

(1) 若 $f(\boldsymbol{x})$是下凸函数，则$-f(\boldsymbol{x})$是上凸的，反之也成立。

(2) 设 $f_1(\boldsymbol{x})$，$f_2(\boldsymbol{x})$，…，$f_m(\boldsymbol{x})$均为 K 中的下凸函数，若 C_1，C_2，…，C_m 为正实数，则它们的线性组合 $f(\boldsymbol{x}) = \sum_{i=1}^{m} C_i f_i(\boldsymbol{x})$ 也是下凸函数。

(3) 若 $f_1(\boldsymbol{x})$，$f_2(\boldsymbol{x})$，…，$f_m(\boldsymbol{x})$中至少有一个是严格下凸的，则 $f(\boldsymbol{x})$严格下凸；反之，当 $f_1(\boldsymbol{x})$，$f_2(\boldsymbol{x})$，…，$f_m(\boldsymbol{x})$均为上凸函数时，$f(\boldsymbol{x})$是上凸的。

(4) 若 $f_1(\boldsymbol{x})$，$f_2(\boldsymbol{x})$，…，$f_m(\boldsymbol{x})$中至少有一个是严格上凸的，则 $f(\boldsymbol{x})$严格上凸。

A.6 凸域中的 Jensen 不等式

设 K 是 n 维欧氏空间的凸域，并设有随机矢量 $\boldsymbol{x}$。若随机矢量 $\boldsymbol{x}$ 的数学期望 $E[\boldsymbol{x}]$存在，$f(\boldsymbol{x})$在 K 内是下凸函数，则有：

$$f(E[\boldsymbol{x}]) \leqslant E[f(\boldsymbol{x})]$$

若 $f(\boldsymbol{x})$是 K 中的上凸函数，则不等号反向。

例如，对于随机变量 X，设它有 m 种取值可能，即 x_1，x_2，…，$x_m \in X$，已知 x_i 发生的概率为 $P(x_i)(i=1, 2, \cdots, m)$，且 $0 \leqslant P(x_i) \leqslant 1$，$\sum_{i=1}^{m} P(x_i) = 1$。设有随机变量的函数 $f(X)$，如果 $f(X)$在凸域 K 内是下凸的，则由 Jensen 不等式有：

$$f\left(\sum_{i=1}^{m} P(x_i)x_i\right) \leqslant \sum_{i=1}^{m} P(x_i)f(x_i)$$

附录 B　线性代数基础

B.1　群、域与环

1. 群

1）**定义**

设 G 是非空的集合，在 G 中定义一种代数运算，符号为“□”，若 G 中的元素满足下述公理：

(1) 有封闭性。对任意的 a，$b\in G$，恒有 $a\square b\in G$。

(2) 结合律成立。对任意的 a，b，$c\in G$，有 $(a\square b)\square c=a\square(b\square c)$。

(3) G 中有一恒等元 e 存在。对任意的 $a\in G$，有 $e\in G$，使得 $a\square e=e\square a=a$。

(4) 对于任意的 $a\in G$，存在有逆元 $a^{-1}\in G$，使得 $a\square a^{-1}=a^{-1}\square a=e$。

则称 G 构成一个群。

2）**加群和乘群**

一般而言，若群 G 中的运算“□”是加法，则简称 G 为加群；若群 G 中的运算“□”是乘法，则简称 G 为乘群。

3）**有限群和无限群**

若群中的元素个数有限，则称为有限群；否则，称为无限群。

4）**阿贝尔群**

若群 G 中，对于任何的 a，$b\in G$，有 $a\square b=b\square a$，则称 G 为交换群、可换群或阿贝尔群(Abel **群**)。

2. 环

1）**定义**

若在非空集合 R 中定义加和乘两种代数运算，且满足下述公理：

(1) 集合 R 关于加法构成阿贝尔群。

(2) 集合 R 中的乘法具有封闭性。对任意的 a，$b\in R$，恒有 $ab\in R$。

(3) 乘法结合律成立，且加法和乘法之间有分配律，即对任意的 a，$b\in R$，有

$$\begin{cases} a(b+c)=ab+ac \\ (b+c)a=ba+ca \end{cases} \tag{B.1}$$

则称 R 是一个环。

2）有单位元环和剩余类环

在环中，对于乘法，无恒等元(单位元)，也无逆元。若环中有单位元，则称其为有单位元环。

整数全体关于模整数 m 的剩余类构成的环，称为剩余类环，用 Z_m 表示。

3. 域

若在非空集合 F 中定义加和乘两种代数运算，且满足下述公理：

(1) F 关于加法构成阿贝尔群。其加法恒等元记为“0”。

(2) F 中非零元素对乘法构成阿贝尔群。其乘法恒等元(单位元)记为“1”。

(3) 加法和乘法间有如下分配律：

$$\begin{cases} a(b+c) = ab + ac \\ (b+c)a = ba + ca \end{cases} \tag{B.2}$$

则称 F 是一个域。

B.2 有限域基础

1. 有限域和无限域

若域中的元素个数有限，则称为有限域；否则，称为无限域。

有限域中元素的个数称为域的阶，q 阶有限域亦称伽罗华域(Galois 域)，记为 $\mathrm{GF}(q)$ 或 F_q。

2. p 阶有限域

设 p 为素数，则整数全体关于模 p 的剩余类 $\overline{0}$，$\overline{1}$，$\overline{2}$，…，$\overline{p-1}$，在模 p 相加和相乘运算下构成的域，称为 p 阶有限域，记为 F_p 或 $\mathrm{GF}(p)$。

3. 二进制域

若在域中定义模 2 加和乘法两种运算的代数，则这种只有 0、1 两个元素的域称为二进制域，记为 $\mathrm{GF}(2)$。

B.3 有限域上的线性代数

1. 即约多项式

设 $f(x)$ 是次数大于零的多项式，若除了常数和常数本身的乘积之外，再不能被域 $\mathrm{GF}(p)$ 上的其他多项式除尽，则称 $f(x)$ 为域 $\mathrm{GF}(p)$ 上的即约多项式。

2. p^n 阶有限域

若 n 次($n>1$)首一多项式 $f(x)$ 在域 $\mathrm{GF}(p)$ 上即约，则由 $f(x)$ 为模所组成的多项式剩余类环是一个有 p^n 个元素的有限域，记为 $\mathrm{GF}(p^n)$。

$\mathrm{GF}(2)$ 是 $\mathrm{GF}(2^m)$ 域的一个子域，$\mathrm{GF}(2^m)$ 称为 $\mathrm{GF}(2)$ 的扩域。

3. 本原域元素和本原多项式

若 α 为 $\mathrm{GF}(q)$ 中的 n 级元素，即 $\alpha^n=1$，则称 α 为 n 次单位原根。若在 $\mathrm{GF}(q)$ 中某元素

α 的级为 $q-1$，则称 α 为本原域元素。

系数取自 GF(p)上的以 GF(p^n)中本原域元素为根的最小多项式，称为本原多项式。

有限域 GF(2^m)中的 2^m 个元素由一个 m 次的本原多项式 $p(x)$产生。我们引入一个新符号 α，并设 $p(\alpha)=0$。然后分别计算得到 2^m-1 个非零元素，再加零元素，总共有 2^m 个域元素。域元素可用幂表示式、多项式表示式和矢量表示式表示。

例如，计算 $p(x)=x^4+x+1$ 产生的域元素。设 $p(\alpha)=0$，则 $p(\alpha)=\alpha^4+\alpha+1=0$，$\alpha^4=\alpha+1$。经运算得到 16 个域元素的三种表示式如表 B-1 所示。

表 B-1　16 个域元素的三种表示式

幂表示式	多项式表示式	矢量表示式
0	0	0000
1	α^0	0001
α	α	0010
α^2	α^2	0100
α^3	α^3	1000
α^4	$\alpha+1$	0011
α^5	$\alpha^2+\alpha$	0110
α^6	$\alpha^3+\alpha^2$	1100
α^7	$\alpha^3+\alpha+1$	1011
α^8	α^2+1	0101
α^9	$\alpha^3+\alpha$	1010
α^{10}	$\alpha^3+\alpha^2+1$	0111
α^{11}	$\alpha^3+\alpha^2+\alpha$	1110
α^{12}	$\alpha^3+\alpha^2+\alpha+1$	1111
α^{13}	$\alpha^3+\alpha^2+1$	1101
α^{14}	α^3+1	1001

4. m 序列

1) m 序列的定义与构造

根据 GF(2)上的 n 次本原多项式组成的线性反馈移位寄存器所产生的序列，称为最大周期线性移位寄存器序列(或最长 n 级线性移位寄存器序列)，简称 m 序列。

在最长序列线性移位寄存器电路中，每右移一次相当于乘 α，因此可以应用这种电路进行 GF(2^n)上域元素之间的相乘和相除运算。

如果生成多项式用下式表示：

$$f(x)=c_nx^n+\cdots c_1x+c_0 \quad c_n=c_0=1 \tag{B.3}$$

则相对应的线性反馈移位寄存器如图 B-1 所示。

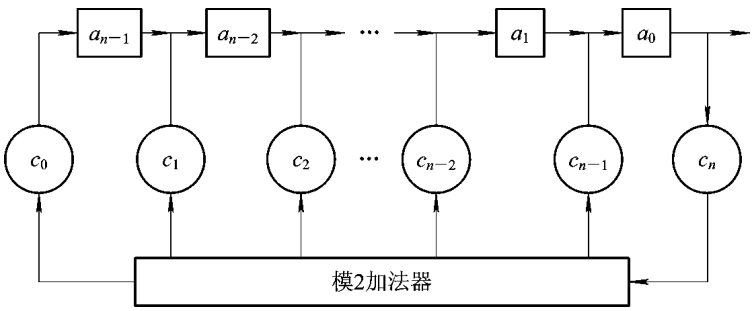

图 B-1 n 级线性反馈移位寄存器

2) m 序列的性质

(1) m 序列的伪随机性质。

① 在每个周期 $p=2^n-1$ 内，“0”出现 $2^{n-1}-1$ 次，“1”出现 2^{n-1} 次，“1”比“0”多出现一次。

② 在每个周期内，共有 2^{n-1} 个元素游程，其中 0 游程和 1 游程的数目各占一半。对于 $n>2$，当 $1\leqslant k\leqslant n-2$ 时，长度为 k 的游程占游程总数的 $1/2^k$，其中 0 游程和 1 游程各占一半；长度为 $n-1$ 的游程只有一个，为 0 游程；长度为 n 的游程也只有一个，为 1 游程。

③ m 序列$\{a_k\}$与其位移序列$\{a_{k-\tau}\}$的模 2 和仍是 m 序列的另一个位移序列$\{a_{k-\tau'}\}$，即

$$\{a_k\}+\{a_{k-\tau}\}=\{a_{k-\tau'}\}$$

(2) m 序列的相关性质。m 序列的自相关函数满足如下关系：

$$R(\tau)=\begin{cases}1 & (\tau=0)\bmod p\\ -\dfrac{1}{p} & (\tau\neq 0)\bmod p\end{cases} \tag{B.4}$$

(3) m 序列优选对。m 序列优选对是指 m 序列集中，互相关函数最大值的绝对值$|R_{\max}|$最接近或达到互相关值下限(最小值)的一对 m 序列。

如果 A 是对应于 n 级本原多项式 $f(x)$所产生的 m 序列，B 是对应于 n 级本原多项式 $g(x)$所产生的 m 序列，则当它们的互相关函数$|R_{ab}(k)|$满足：

$$|R_{ab}(k)|=\begin{cases}2^{\frac{n+1}{2}}+1 & n\text{ 为奇数}\\ 2^{\frac{n+2}{2}}+1 & n\text{ 为偶数且不为 4 的倍数}\end{cases} \tag{B.5}$$

时，$f(x)$和 $g(x)$所产生的 m 序列 A 和 B 构成一对优选对。

经过计算机搜索，部分优选对见表 B-2。

表 B－2　部分优选对码表

级数	基准本原多项式	配对本原多项式			
7	211	217	235	277	325
		203	357	301	323
	217	211	235	277	325
		213	271	357	323
	235	211	217	277	325
		313	221	361	357
	367	277	203	313	345
		221	361	271	375
9	1021	1131		1333	
	1131	1021	1055	1225	1725
	1461	1743	1541	1853	
10	2415	2011	3515	3177	
	2641	2517	2213	3045	
11	4445	4005	5205	5337	5263
	4215	4577	5747	6765	4563

部分习题参考答案

第 2 章

2.1 1(该信息量单位)=2.33 比特；1(该信息量单位)=1.61 奈特。

2.2 1 比特，2 比特，3 比特，3 比特；1.75 比特/符号。

2.3 $I(x=点)=0.415$ 比特，$I(x=画)=2$ 比特；$H(X)=0.811$ 比特/符号。

2.5 2.341 比特/符号。

2.6 (1) 5.17 比特，2.85 比特，4.17 比特；(2) 3.274 比特/符号。

2.8 $I_2(X_3)=0.15$ 比特，$I_0(X_3)=2$ 比特，$I_1(X_3)=1.85$ 比特。

2.9 $I_1=1$ 比特，$I_2=2$ 比特；

$I_{1总}=4$ 比特，$I_{2总}=3.24$ 比特；

$\bar{I}_X=1$ 比特/符号$=H(X)$，$\bar{I}_Y=0.81$ 比特/符号$=H(Y)$。

2.10 $H(S')=H(S)-(1-\varepsilon)\log(1-\varepsilon)-\varepsilon\log\varepsilon$。

2.12 $\begin{cases} H(Y)=H(X)+\log a & a>0 \\ H(Y)=-H(X)-\log a & a<0 \end{cases}$。

2.13 (1) $H(X)=1+\ln\lambda$；

(2) $H(X)=1+\ln(2\lambda)$；

(3) $H(X)=\ln(\lambda^2)-\ln(\lambda)+\frac{1}{2}$。

第 3 章

3.4 (1) $H(X)=1.9527$ 比特/符号，$H(X|Y)=1.851$ 比特/符号，$H(Y|X)=1.8692$ 比特/符号。

(2) $I(x_1;y_1)=-0.8480$ 比特，$I(x_1;y_2)=0.7370$ 比特，$I(x_2;y_1)=0.1520$ 比特，$I(x_2;y_2)=-0.2630$ 比特，$I(X;Y)=0.1017$ 比特/符号。

3.5 (1) $I(x=0;y=1)=-1.3785$ 比特；(2) $I(x=1;Y)=0.072\,06$ 比特；

(3) $I(X;Y)=0.2087$ 比特/符号。

3.6 实际求传输系统的平均互信息量：$I(X;Y)=0.093$ 比特/符号。

3.7 (1) $I(x_1)=0.737$ 比特，$I(x_2)=1.322$ 比特；

(2) $I(x_1;y_1)=0.0589$ 比特，$I(x_1;y_2)=-0.2630$ 比特，$I(x_2;y_1)=-0.0931$ 比

特，$I(x_2;\ y_2)=0.3219$ 比特；

(3) $H(X)=0.9710$ 比特/符号，$H(Y)=0.7219$ 比特/符号；

(4) $H(X|Y)=0.9635$ 比特/符号，$H(Y|X)=0.7145$ 比特/符号；

(5) $I(X;\ Y)=0.0075$ 比特/符号。

第 4 章

4.1 $C=0.0817$ 比特/符号；输入为等概率分布。

4.2 准对称信道 $C=0.398$ 比特/符号。

4.3 $C=H\left(\frac{1}{2}-\varepsilon,\ \frac{1}{2}-\varepsilon,\ \varepsilon,\ \varepsilon\right)-H(\bar{p}-\varepsilon,\ p-\varepsilon,\ 2\varepsilon,\ 0)$。

4.5 (2) $C=\frac{1}{2}\log\left(1+\frac{p}{\sigma^2}\right)$，当输入是方差为 p 的高斯分布时达到信道容量。

第 5 章

5.2 (1) $G=\begin{bmatrix}1 & 0 & 1 & 1 & 1\\ 0 & 1 & 1 & 0 & 1\end{bmatrix}$；

(2) 译码表如下：

伴随式	陪集首			
000	00000	01101	10111	11010
001	00001	01100	10110	11011
010	00010	01111	10101	11000
011	00011	01110	10100	11001
100	00100	01001	10011	11110
101	01000	00101	11111	10010
110	10001	11100	00110	01011
111	10000	11101	00111	01010
译码结果	00000	01101	10111	11010

(3) $p_e=(1-\varepsilon)^5+5\varepsilon(1-\varepsilon)^4+2\varepsilon^2(1-\varepsilon)^3$。

第 6 章

6.1 (1) $H(X^2)\approx 1.6226$ 比特/符号序列；

(2) $H(X^3)\approx 2.4338$ 比特/符号序列。

6.2 $H_0(X)=2$ 比特/符号，$H_1(X)\approx 1.319$ 比特/符号，相对熵 $\eta\doteq 0.6595$，冗余度 $\gamma=0.3405$。

6.3 (1) $H(X)=1.5$ 比特/符号；

(2) $H(X_2|X_1)\approx 0.9363$ 比特/符号，$H(X_1X_2)=2.4363$ 比特/符号。

6.4 (1) $H_1(X)=H(1/4,1/2,1/4)=1.5$ 比特/符号，此时冗余度 $\gamma=0.0536$；

(2) $H_2(\boldsymbol{X})=1.2181$ 比特/符号，此时 $\gamma\approx0.2315$。

6.6 (1) 是平稳信源；

(2) $H(X^2)=1.942$ 比特/符号序列，$H(X_3|X_1X_2)=0.971$ 比特/符号，$\lim\limits_{N\to\infty}H_N(X)=0.971$ 比特/符号；

(3) $H(X^4)=3.884$ 比特/符号序列。

6.7 $H_\infty=1.4388$ 比特/符号。

6.8 (1) $P(0)=P(1)=P(2)=\dfrac{1}{3}$；

(2) $H_\infty=-\bar{p}\log\bar{p}-p\log p+p$；

(3) $H(X)\approx1.585$ 比特/符号；

(4) 当 $p=0$ 时，$H_\infty=-\bar{p}\log\bar{p}-p\log\dfrac{p}{2}=0$；

当 $p=1$，时 $H_\infty=-\bar{p}\log\bar{p}-p\log\dfrac{p}{2}=1$ 比特/符号。

6.9 (1) $P(0)=P(1)=P(2)=1/3$；

(2) 一阶马尔可夫信源的熵 $H_\infty=H(p)$；

(3) 当 $p=0$ 或 $p=1$ 时，$H_\infty=0$。

6.10 (1) $H(X)=0.881$ 比特/符号；

(2) $H_2=0.554$ 比特/符号。

第 7 章

7.1 (1) 信息熵 $H(X)=8$ 比特；

(2) $R=6.4\times10^4$ 比特/秒；

(3) 3 个数符；

(4) 序列长度为 24。

7.2 (1) $\eta_1=0.469$；

(2) $\bar{N}=0.645$，$\eta_2=0.727$；

(3) $\bar{N}=0.533$，$\eta_3=0.880$。

7.3 (1) 码 A、B、C、E 是唯一可译码；

(2) 码 A、C、E 是即时码；

(3) $\bar{N}_A=3$，$\bar{N}_B=2\dfrac{1}{8}$，$\bar{N}_C=\bar{N}_B=2\dfrac{1}{8}$，$\bar{N}_E=2$；

(5) 存在字首码；

(6) 不存在满足字首性的编码。

7.4 (1) $\eta_{HX}=0.959$，$\eta_{HY}=0.993$；

(2) $\eta_{FX}=0.952$，$\eta_{FY}=0.993$。

7.5 (1) 编码效率 $\eta=0.907$；

(2) 编码效率 $\eta=0.979$。

7.6 (1) $N_0=18\ 840$；

(2) $0.99\times 2^{143\ 425}\leqslant M_\varepsilon\leqslant 2^{16\ 226.5}$。

7.7 (1) $l=18$；

(2) $\varepsilon>0.1909$；

(3) **按编码考虑** $p_e\approx 0.0017$，**按契比雪夫不等式考虑** $p'_e\approx 0.08$。

7.9 (1) $H(s)=0.469$ **比特/符号，剩余度** $\gamma=53.1\%$；

(2) **平均码长为** $\bar{L}=1$ **码符号/信源符号**；

(3) **当** $N=2$ **时，平均码长** $\left(\frac{\bar{L}_N}{N}\right)\approx 0.645$ **码符号/信源符号**；

当 $N=3$ **时，平均码长** $\left(\frac{\bar{L}_N}{N}\right)\approx 0.533$ **码符号/信源符号**；

当 $N=4$ **时，平均码长** $\left(\frac{\bar{L}_N}{N}\right)\approx 0.493$ **码符号/信源符号**；

当 $N=\infty$ **时，** $\lim\limits_{N\to\infty}\frac{\bar{L}_N}{N}\approx 0.469$ **码符号/信源符号**。

7.10 (4) $v=(a-1)2^{k+1}$，$u=n-v=(2-a)2^k$，

$$\bar{L}=\frac{v}{a2^k}(k+1)+\frac{a2^k-v}{a2^k}k=k+\frac{2(a-1)}{a}。$$

第 8 章

8.1 (1) $\bar{d}=0$，$I(U;V)=1.585$ **比特/符号**；

(2) $\bar{d}=\frac{10}{9}$，$I(U;V)=0$ **比特/符号**；

(3) $\bar{d}=\frac{2}{3}$，$I(U;V)=0$。

8.2 $D_{\min}=\frac{3}{8}$，$D_{\max}=\frac{2}{3}$。

8.3 $D_{\min}=0$，$D_{\max}=\min\{p,1-p\}$，$R(D)=H(p)-H(D)$，

其中 $H(D)=-D\ \ln D-(1-D)\ln(1-D)$，$0\leqslant D\leqslant D_{\max}$。

8.4 $D_{\min}=1$，$D_{\max}=2$。

8.5 $$R(D)=\begin{cases}0.918-H\left(\frac{D}{a}\right) & 0\leqslant D\leqslant\frac{a}{3}\\ 0 & D>\frac{a}{3}\end{cases}。$$

8.6 $$R(D)=\begin{cases}2-D\ \log 3-H(D) & 0\leqslant D\leqslant\frac{3}{4}\\ 0 & D>\frac{3}{4}\end{cases}。$$