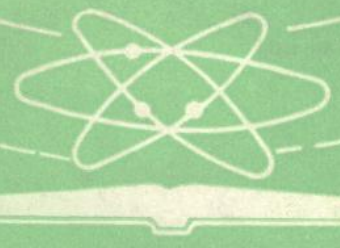
信息论与编码

高等学校教材

信息论与编码

吴伯修　归绍升　祝宗泰　俞槐铨　编



電子工業出版社

信息论与编码

吴伯修　归绍升　祝宗泰　俞槐铨　编

電子工業出版社

内 容 简 介

本书主要介绍狭义信息论的基本内容。包括信息的量度，离散信源和连续信源的信息量，离散信道和连续信道的信道容量，消息的信息传输速率，香农的信源编码定理和信道编码定理，信息率失真函数，纠错编码的代数基础，各种纠错码的编码和译码原理及实施技术。

本书可作为通信与电子系统等专业的研究生教材及无线电技术、通信、信息工程、生物医学工程、计算机科学与工程等专业的本科生教材，亦可供从事无线电技术和通信工程的技术人员参考。

信息论与编码

吴伯修 归绍升 祝宗泰 俞槐铨 编

责任编辑 王昌喜

电子工业出版社出版（北京海淀区万寿路）

新华书店北京发行所发行 各地新华书店经售

商务印书馆上海印刷厂印刷

*

开本：787×1092 毫米 1/16 印张：16 字数：369 千字

1987 年 5 月第 1 版 1987 年 5 月第 1 次印刷

印数：1—5,200 册 定价：2.65 元

统一书号：15290·492

出 版 说 明

根据国务院关于高等学校教材工作分工的规定，我部承担了全国高等学校工科电子类专业课教材的编审、出版的组织工作．从一九七七年底到一九八二年初，由于各有关院校，特别是参与编审工作的广大教师的努力和有关出版社的紧密配合，共编审出版了教材 159 种．

为了使工科电子类专业教材能更好地适应社会主义现代化建设培养人材的需要，反映国内外电子科学技术水平，达到"打好基础、精选内容、逐步更新、利于教学"的要求，在总结第一轮教材编审出版工作经验的基础上，电子工业部于一九八二年先后成立了高等学校《无线电技术与信息系统》、《电磁场与微波技术》、《电子材料与固体器件》、《电子物理与器件》、《电子机械》、《计算机与自动控制》，中等专业学校《电子类专业》、《电子机械类专业》共八个教材编审委员会，作为教材工作方面的一个经常性的业务指导机构．并制定了一九八二至一九八五年教材编审出版规划，列入规划的教材、教学参考书、实验指导书等共 217 种选题．在努力提高教材质量，适当增加教材品种的思想指导下，这一批教材的编审工作由编审委员会直接组织进行．

这一批教材的书稿，主要是从通过教学实践、师生反映较好的讲义中评选择优和从第一轮较好的教材中修编产生出来的．广大编审者，各编审委员会和有关出版社都为保证和提高教材质量作出了努力．

这一批教材，分别由电子工业出版社、国防工业出版社、上海科学技术出版社、西北电讯工程学院出版社、湖南科学技术出版社、江苏科学技术出版社、黑龙江科学技术出版社和天津科学技术出版社承担出版工作．

限于水平和经验，这一批教材的编审出版工作肯定还会有许多缺点和不足之处，希望使用教材的单位、广大教师和同学积极提出批评建议，共同为提高工科电子类专业教材的质量而努力．

电子工业部教材办公室

前　言

本教材由高等院校工科电子类《无线电技术与信息系统》教材编审委员会《无线通信》编审小组评选审定，并推荐出版.

信息论是在概率论随机过程和通信技术相结合的基础上发展起来的学科. 自香农在1948年发表奠定信息论基础的"通信的数学理论"一文以来，信息论有了迅猛的发展，已经渗透到许多学科. 就所涉及的内容和范围而言，可大致分为狭义信息论和广义信息论两种. 前者是在香农信息论的基础发展起来的，故称为经典信息论. 它仍然是当今信息论发展的主流. 后者是冲破香农信息论的框框建立起来的更为一般性的广义信息论，亦称为信息科学，它是以广义信息为主要研究对象，以扩张人类信息功能为目的的新兴边缘学科. 虽然这一学科还未成熟，但无疑地在即将来临的信息社会中一定会得到发展和应用，并将进一步推广到其它学科的领域.

本书的内容主要是介绍经典信息论和提高通信可靠性的纠错编码理论. 全书分为两篇. 第一篇信息论基础，包括第一章至第五章，主要介绍信息传输方式和通信系统模型，信息的量度，离散信源及其熵，无扰离散信道的信道容量，信息传输速率，信源编码，香农第一定理，有扰离散信道的信道容量，香农第二定理，连续消息的信息量度，熵功率，连续信道的信道容量，香农信道容量公式，理想接收机，离散消息和连续消息的信息率失真函数等. 第二篇纠错编码，包括第六章至第十一章，主要介绍纠错编码的历史、现状及发展方向，纠错编码的代数基础，线性分组码的生成矩阵和校验矩阵，汉明距离和汉明重量，纠错能力，循环码的数学结构及其生成矩阵和校验矩阵，循环码的编码电路和译码方法，卷积码的特点和译码算法，几何码和算术码的特点及其编码和译码方法.

本书的第一篇由南京工学院吴伯修教授和祝宗泰合编、第二篇的第六章和第十一章由上海交通大学归绍升教授编，第七章至第十章由上海交通大学俞槐铨编. 华中工学院蔡德钧副教授担任主审. 他对本书内容提出了宝贵的修改意见. 编者在此向他表示诚挚的感谢.

限于编者的水平，加上时间比较侷促，书中难免还会有一些错误，殷切希望读者指正.

编者　1986年4月

目　录

第一篇　信息论基础

第二篇　纠 错 编 码

第一篇　信息论基础

第一章　通信的基础知识

第一节　引　　言

信息论是应用近代数理统计方法来研究信息的传输和处理的科学.

什么是信息呢?通常把客观存在的各个事物状态的表露,以及各个事物随时间所发生的变化的反映都称为**信息**. 可见,信息是依附于物质的,凡是有物质的地方就有信息存在. 人们总是通过现象去认识各个事物的,信息就是认识过程的中介物.

举凡讲话、写信、无线电定位、导弹的制导、电子计算机的运算,甚至生物的感觉和遗传过程,都有信息的传输和处理问题. 信息论是在信息可以量度的基础上,研究最有效、最可靠地传输和处理信息的理论. 由于这种理论对于许多学科都有指导意义,而且蕴藏着大量新的数学问题,因此近三十多年来得到许多数学家、科学家和工程师们的重视. 同时信息论的研究成就也已经在通信、电视、雷达、导航、制导、数据处理、计算机、自动控制、光学、生物学、心理学等领域中得到愈来愈广泛的应用.

最近三十多年来,信息论主要是沿着两条不同的途径发展的. 其一是在维纳(Wiener, N.)的"平稳时间序列的内插、外推和平滑方法"和"控制论"两本名著的基础上发展起来的一个信息论分支**微弱信号检测理论**. 其二是在香农(Shannon, C. E.)的"通信的数学理论"和"噪声中的通信"这两篇经典著作的基础上发展起来的另一个信息论分支**信号的设计和编码理论**.

虽然维纳和香农都认为信号和噪声均可以用规定集合的统计规律来描述,但是他们探讨的数学模型却有很大的不同. 维纳认为信号只有当它受到噪声干扰时,才需要进行处理. 而香农却认为信号在通过噪声信道的前后都需要进行处理. 两种模型的最终目的都是要在接收端尽可能逼真地重现原信号. 他们两人都研究噪声对于接收端重现原信号的影响,所不同的是:维纳着重研究自动控制过程中的信号预测问题,也就是主要研究重现负时延的原信号;而香农则着重研究重现正时延的原信号. 因此,维纳的研究工作及"微弱信号检测理论"着重研究在干扰作用下信号的最佳接收问题,它是通信、雷达、导航、遥测、遥控以及电子对抗等技术的理论基础. 而香农的研究工作及"信号设计和编码理论"着重研究信源和信道的统计特性及其编码方法,以提高信息传输的效率和可靠性. 由于香农信息论是通信理论的基础,所以这部分内容通常就称为信息论. 本书第一篇将着重介绍这部分内容. 但是,更加普遍的是把信息论看成是包括上述两个分支的科学.

随着自动控制、电子计算技术、系统工程、仿生学、人工智能及其它边缘科学的发展,它们和信息论结合,成为新兴的**信息科学**,所包括的内容就更加广泛. 信息科学的主要任务是应用现代数理方法来研究信息的性质,研究利用机器来检测、变换、传输和控制各种信息的

基本理论和技术，研究实现这些功能所需要的设备和系统的原理．信息科学的基本理论是信息论和控制论，但它比信息论的研究范围更广阔，涉及的内容更深刻、更复杂，因此需要更好的工具，这就是信息论需要与电子计算机相结合的原因．信息论与电子计算机的结合为信息系统提供了代数运算、贮存、记忆和逻辑演绎的能力，把信息处理技术提高到一个更高的水平．此外，信息论与仿生学和智能理论相结合，将为信息系统的发展开辟一个广阔的新方向．自然界各种生物都有它们独特的信息识别和信息处理的能力．从生物识别和处理信息的机理中吸取技术思想，并用电子技术的各种手段在信息系统中实现，从而可以明显地改善机器的信息处理能力．人脑的思维活动是比较完善又比较精密的高级信息处理系统．若能从人脑处理信息的机理中吸取技术思想来丰富和提高机器的能力，那末机器就能实现逻辑演绎、推理和形式逻辑的思维．具有这种能力的系统称为智能信息系统，它是信息科学研究的基本对象．

可以预期，随着信息科学的发展，将为人类提供最有效、最可靠的信息传输、信息处理和信息控制的手段．它将对现代科学技术的发展产生更为深远的影响．

第二节　信息的传输

一、通信系统的模型

通信系统是传输消息的系统．电报通信系统传输的电文、电话通信系统传输的话音、广播系统播送的音乐和话音、电视系统传输的图象和伴音、雷达系统获得的目标参数、遥测系统测量得到的数据、遥控系统传输的指令等都是消息．尽管这些系统的结构和所传输的消息各不相同，但它们都属于广义通信系统的范畴．通信系统的模型可用图 1-1 来表明．我们将利用这个模型来探讨通信系统在传输消息的过程中带有普遍性的规律．下面简单介绍通信系统模型中各个组成部分的作用．

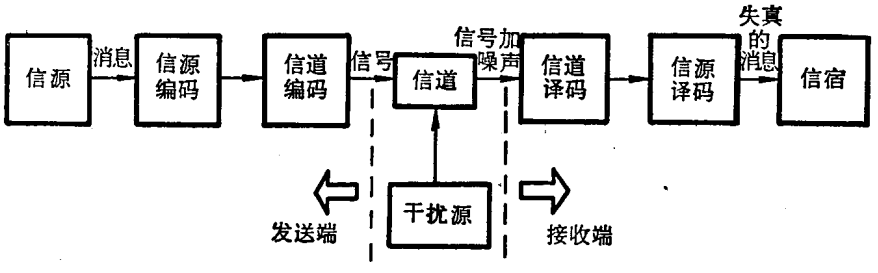


图 1-1　通信系统的模型

信源是向通信系统提供消息的人或机器．信源发出的消息可以有多种形式，但可以归纳成两类：一类是**离散消息**，如字母、符号、文字和数字等；另一类是**连续消息**，如话音、图象、时间上连续的电参量等．

信宿是消息传输的对象，也就是接收消息的人或机器．

信道是指通信系统传输消息的媒介．它可以是一对导线、一条同轴电缆、传输电磁波的空间、一条光导纤维等等．

信源编码器、信道编码器是通信系统发送端的组成部分．它们的共同任务是把信源发出的消息变换成适宜于在信道上传输的信号．两者的区别是信源编码的目的在于提高传输

消息的有效性，而信道编码的目的在于提高传输消息的可靠性.

信源译码器、信道译码器是通信系统接收端的组成部分. 它们的共同任务是把信道输出的信号变换成信宿需要的消息形式. 同时信源译码是信源编码的逆过程，信道译码是信道编码的逆过程.

干扰源是整个通信系统上各种干扰的集中反映. 消息在传输过程中会遭受到各种噪声和干扰(如器件热噪声、信道的热骚动、人为干扰等)的作用. 为简化起见，我们把各种干扰的影响折合为直接作用在信道上的一个干扰源. 干扰源产生的干扰和信道上的传输信号一起成为接收端接收的信号. 对于任何通信系统，干扰是限制系统性能的基本因素.

在这里，需要把信息、消息和信号这三个名词的含义说明一下. **信息**是指各种事物状态和变化的反映；**消息**是指包含信息的事物(例如语言、文字、图象等)；**信号**是消息的运载工具. 为了传输消息，就必需把消息载入具有某种物理能量(例如电、光、热能量)的信号上去.

二、通信目的

通信的目的是使信源发出的消息能被信宿接受. 或者说，通信的目的是使得信源发出的消息能在接收者认为方便的地方恢复成接收者所要求的消息形式.

关于通信目的，还要说明两个问题. 其一是接收者并不一定需要在接收端完整无缺地恢复信源发出的消息，接收者需要的仅是符合一定保真度准则的重现消息. 这就是说接收者需要的消息通常允许有一定的失真. 其二是接收者需要的消息形式不一定与信源发出的消息形式相同.

三、模拟通信和数字通信

如前所述，信源发出的消息可分成连续消息和离散消息两类. 由连续消息变换成的连续信号真实地反映连续消息随时间变化的过程，它在时间上是连续的，幅度取值也是连续的. 因为这种连续信号能模拟原消息的整个变化过程，故又称为**模拟信号**. 另一类是由离散消息变换成的**离散信号**. 此外，应用取样定理对连续信号进行取样，也可把它变成时间离散的信号，然后再通过幅度量化过程，使其幅度取值由连续取值变为所处量化等级的量化值，那么它就变成数字信号了.

通信系统可以分成模拟通信和数字通信两类. **模拟通信**是指用连续信号作为传输信号的通信系统，如广播、电视、载波长途电话、市内电话等. **数字通信**是指用数字信号作为传输信号的通信系统，如电报、数据传输、数字电话、计算机通信等. 与模拟通信相比，数字通信具有较多优点，故近年来数字通信已有迅速的发展. 数字通信的优点有：抗干扰能力强；在远距离中继通信中干扰的影响不会累积；可以用纠错技术提高可靠性；适宜于与计算机通信网相结合等. 但数字通信也有明显的缺点，即它所需要的系统通频带比模拟通信系统大得多(如数字电话系统的通频带通常为模拟电话的八倍). 但随着通信系统工作频段的提高，特别是光纤通信的发展，通信系统获得的通频带正在增加，可以满足数字通信的需要.

四、信息论研究的基本问题

信息论研究的基本问题是有关信源、信宿和信道的统计特性，以及信源编码和信道编码等问题.

关于信源，信息论研究了信源所包含的信息量(熵)，以及在单位时间内信源发出的信息量(时间熵). 关于信宿，信息论研究了在无扰信道和有扰信道上信宿能收到的信息量的多

少. 关于信道，信息论研究了信道传输信息量的能力(信道容量)，并叙述了当信息传输速率小于信道容量时，可以以很小的误码率传输消息(有扰离散信道的信道编码定理). 关于信源编码，信息论的信源编码理论指出：通过信源编码器的编码过程可以使信源发出的消息变换成其码长度与信源各消息的概率分布达到匹配的代码组，以提高传输消息的有效性. 关于信道编码，信息论的纠错编码理论指出了各种纠错编码方式的数学理论、纠错能力和实施方案. 本书第二篇将专门讨论信道编码的理论和实施技术.

信息论在理论上指出了建立具有最佳编码、最佳调制和最佳接收方法的最佳系统的理论原则，它对通信体制和通信系统的研究具有指导意义. 根据本书的编写目的，本书仅介绍信息论的基本内容.

第二章　信息的量度

第一节　自信息量和条件自信息量

在上一章第一节引言中已指出信息论是在信息可以量度的基础上，研究最有效、最可靠地传输和处理信息的科学．可见信息的量度是建立信息论的基础，是一个十分重要的概念．本章将介绍量度信息量的方法．

一、自信息量

任意随机事件的**自信息量**定义为用其出现概率的对数的负值来量度．若随机事件 x_i 的出现概率为 $P(x_i)$，那么它的自信息量 $I(x_i)$ 定义为

$$I(x_i)\triangleq -\log P(x_i) \tag{2.1}$$

自信息量的单位与所用的对数底有关．通常取对数的底为 2，信息量的单位为**比特**(Bit)．若 $P(x_i)=\frac{1}{2}$，$I(x_i)=1$ 比特，即该随机事件 x_i 具有 1 比特的自信息量．比特是信息论中最常用的信息量单位．此外，从数学运算方便出发，取自然对数(对数底为 e)计算信息量，此时信息量的单位为**奈特**(Nat)．还有以 10 作为对数底，此时信息量的单位为**哈脱来**(Haitely)．这三个信息量单位之间的换算关系为

$$1\text{ 奈特}=\log_2 e\text{ 比特}\simeq 1.443\text{ 比特}$$

$$1\text{ 哈脱来}=\log_2 10\text{ 比特}\simeq 3.322\text{ 比特}$$

通常规定一个以等概率出现的二进制码元 (0 或 1) 所包含的自信息量为 1 比特．即当 $P(0)=P(1)=\frac{1}{2}$ 时，它们各自的自信息量分别为

$$I(0)=I(1)=-\log_2\frac{1}{2}=\log_2 2=1\text{ 比特}$$

这里要引入随机事件的**不确定性**(或称**不肯定性**)的概念．首先要指出，由于各种随机事件的出现概率不同，它们所包含的不确定性也有大小的差别．一个随机事件的出现概率接近于 1，说明该事件发生的可能性很大，它所包含的不确定性就很小．反之，出现概率很小的随机事件，它所包含的不确定性就很大．例如若有教师告诉一位学生："明天是星期日，你可以不来上课。"这句话包含的不确定性很小．但如果有教师对学生讲："明天是星期六，你们不要到学校来上课。"那么学生一定会问："为什么明天不要来上课呀？"教师就要把这个星期六不上课的原因解释一番．可见，"在一般情况下星期六不上课"这一事件所包含的不确定性很大．又倘若你事先已经确知明天是星期日，而有人来告诉你："明天是星期日。"那么，你没有得到任何信息．这是因为它是确定性事件，出现概率等于 1，而所含的不确定性为零．

由此得出结论：出现概率小的随机事件所包含的不确定性大，也就是它的自信息量大．出现概率大的随机事件所包含的不确定性小，也就是它的自信息量小．在极限情况，出现概率为 1 的确定性事件，其自信息量为零．可见上述自信息量的定义是合理的．

随机事件的不确定性在数量上等于它的自信息量，也可用(2.1)式计算，同时两者的单位也相同.

信宿收到从信道传输来的消息所携带的信息量后，可以全部消除或部分减小信宿对信源发出消息的不确定性. 当信宿收到的信息量等于信源发出消息包含的不确定性时，就可以唯一地确定信源发出的消息. 例如当随机事件 x_i 的出现概率为 $P(x_i)=\frac{1}{8}$ 时，它包含3比特的不确定性. 当信宿能收到3比特信息量时，就能唯一地确定信源发出的是消息 x_i.

在二维联合集 XY 上元素对 x_iy_j 的自信息量可与(2.1)式相类似，定义为

$$I(x_iy_j)\triangleq -\log P(x_iy_j) \tag{2.2}$$

其中 $P(x_iy_j)$ 为元素对 x_iy_j 的联合概率.

二、条件自信息量

条件自信息量定义为用其条件概率的对数的负值来量度. 若随机事件 x_i 的条件概率为 $P(x_i|y_j)$，那么它的条件自信息量 $I(x_i|y_j)$ 定义为

$$I(x_i|y_j)\triangleq -\log P(x_i|y_j) \tag{2.3}$$

随机事件的条件自信息量可以理解为：能在规定条件下唯一地确定该事件必须提供的信息量.

由于每一个随机事件的概率在 0～1 范围内，所以自信息量和条件自信息量都是非负值.

第二节　互信息量和条件互信息量

一、互信息量

设有两个离散的符号消息集合 X 和 Y，X 表示信源发出的符号消息集合，Y 表示信宿接收到的符号消息集合. 由于信宿事先不知道信源发出的是哪一个符号消息，所以每个符号消息相当于一个随机事件. 信源发出的符号消息通过信道传送给信宿，如图 2-1 所示. 这是最简单的通信系统模型. 有时我们也把信源发出的消息说成信道的输入消息，同时把信宿收到的消息说成信道的输出消息. 通常信宿预先知道信源 X 集合包含的各个符号消息以及它们的概率分布，也就是预先知道信源 X 集合的概率空间 X，$P(x_i)$ 的情况. 各个消息 $x_i(i=1, 2, \cdots\cdots)$ 的概率 $P(x_i)$ 称为**先验概率**. 当信宿收到 Y 集合中的一个消息 y_j 后，接收者重新估计的关于信源各个消息的概率分布就变成条件概率. 如对消息 x_i 而言，就有条件概率 $P(x_i|y_j)$. 这种条件概率又称为**后验概率**.

图 2-1　最简单的通信系统模型

互信息量定义为后验概率与先验概率比值的对数. 用数学公式表示，互信息量 $I(x_i; y_j)$ 定义为

$$I(x_i; y_j)\triangleq \log\frac{P(x_i|y_j)}{P(x_i)} \tag{2.4}$$

互信息量的单位与自信息量一样，取决于对数的底数. 当底数分别为 2、e、10 时，互信息量的单位分别为比特、奈特和哈脱来.

二、互信息量的性质

1．互信息量有互易性

互信息量的互易性可用下式表示：

$$I(x_i; y_j) = I(y_j; x_i) \tag{2.5}$$

可见，互信息量表明了两个随机事件 x_i 与 y_j 之间的统计约束程度，同时互易性也说明了互信息量名称的由来.

当信宿收到 y_j 后，若 x_i 的后验概率 $P(x_i|y_j)$ 大于 x_i 的先验概率 $P(x_i)$，则互信息量 $I(x_i; y_j)$ 为正值，说明信宿获得了有关 x_i 的信息量，也就是 y_j 提供了有关 x_i 的信息量. 由互易性可知，由 y_j 提供的有关 x_i 的信息量等于由 x_i 提供的有关 y_j 的信息量.

互信息量的互易性证明如下：

在(2.4)式右边的分子和分母上同时乘以 $P(y_j)$，得

$$I(x_i; y_j) = \log \frac{P(x_i|y_j)P(y_j)}{P(x_i)P(y_j)} = \log \frac{P(x_iy_j)/P(x_i)}{P(y_j)}$$

$$= \log \frac{P(y_j|x_i)}{P(y_j)} = I(y_j; x_i)$$

2．当 x_i 与 y_j 相互独立时，互信息量为零

当 x_i 与 y_j 相互独立时，由概率论的公式知 $P(x_iy_j) = P(x_i) \cdot P(y_j)$. 在这种情况下，互信息量为

$$I(x_i; y_j) = \log \frac{P(x_iy_j)}{P(x_i)P(y_j)} = \log \frac{P(x_i)P(y_j)}{P(x_i)P(y_j)} = \log 1 = 0$$

可见，当 x_i 和 y_j 相互独立时，它们的互信息量为零，也表示 x_i 和 y_j 之间没有统计约束关系.

3．互信息量可正可负

如前所述，当后验概率大于先验概率时，互信息量大于零，为正值. 但是，当后验概率小于先验概率时，互信息量就是负值.

关于互信息量，下面举两个例子说明.

[例 1] 某人 A 预先知道他的三位朋友 B、C、D 中必定有一人于某天晚上要到他家中来，并且这三人来的可能性相同. 我们用 $P(B)$、$P(C)$、$P(D)$ 分别表示 B、C、D 三人到 A 家来的概率，即先验概率 $P(B) = P(C) = P(D) = \frac{1}{3}$. 但这天上午，$A$ 接到 D 的电话，说他晚上有会议，不能来了. 我们把上午的这次电话作为事件 E，那么有后验概率 $P(D|E) = 0$，$P(B|E) = P(C|E) = \frac{1}{2}$. 这天下午，$A$ 又接到 C 的电话，说他因晚上要去看电影，不能来 A 家. 我们把下午这一次电话作为事件 F，那么就有后验概率 $P(C|EF) = P(D|EF) = 0$，$P(B|EF) = 1$.

在接到上午的电话后，A 获得的关于 B，C，D 的互信息量为

$$I(B; E) = \log \frac{P(B|E)}{P(B)} = \log_2 \frac{\frac{1}{2}}{\frac{1}{3}} = \log_2 1.5 = 0.585 \text{ 比特}$$

$$I(C; E) = I(B; E) = 0.585 \text{ 比特}$$

因为 $P(D|E) = 0$，即在 E 事件条件下，不出现 D 事件，所以不必考虑 D 事件与 E 事件之间的互信息量.

在接到两次电话后，A 获得的关于 B, C, D 的互信息量为

$$I(B;EF)=\log\frac{P(B|EF)}{P(B)}=\log_2\frac{1}{\frac{1}{3}}=\log_2 3=1.585\text{ 比特}$$

因为其它两个条件概率都为零，所以不必考虑 C 事件、D 事件与 EF 事件之间的互信息量。

[例 2] 已知信源 U 包含八个数字消息 0, 1, 2, 3, 4, 5, 6, 7. 为了在二进制信道上传输，我们用信源编码器把这八个十进制数编成三位二进制代码组. 信源各消息的先验概率及相应的代码组见表 2-1 的前三列. 消息的传输情况可用图 2-2 表示，这是带有信源编码器和信源译码器的通信系统模型.

表 2-1 各消息的后验概率分布情况

信源消息	二进制代码组	先验概率 $P(u_i)$	译码器收到 0 后的后验概率 $P(u\|x_0)$	译码器收到 01 后的后验概率 $P(u\|x_0y_1)$	译码器收到 011 后的后验概率 $P(u\|x_0y_1z_1)$
$0(u_0)$	$000(x_0y_0z_0)$	1/4	1/3	0	0
$1(u_1)$	$001(x_0y_0z_1)$	1/4	1/3	0	0
$2(u_2)$	$010(x_0y_1z_0)$	1/8	1/6	1/2	0
$3(u_3)$	$011(x_0y_1z_1)$	1/8	1/6	1/2	1
$4(u_4)$	$100(x_1y_0z_0)$	1/16	0	0	0
$5(u_5)$	$101(x_1y_0z_1)$	1/16	0	0	0
$6(u_6)$	$110(x_1y_1z_0)$	1/16	0	0	0
$7(u_7)$	$111(x_1y_1z_1)$	1/16	0	0	0

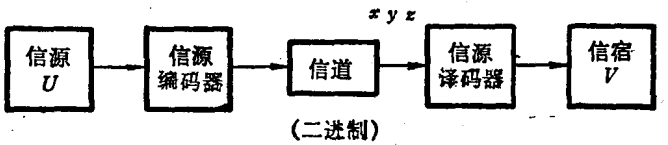


图 2-2 带有信源编码器的通信系统模型

当译码器收到第一个码元为 0(即 x_0)后，因为消息 $u_4\sim u_7$ 的代码组的第一个码元是 1，说明这四个消息的后验概率必定为零. 消息 u_0 的后验概率 $P(u_0|x_0)$ 为

$$P(u_0|x_0)=\frac{1}{4}\Big/\left(2\times\frac{1}{4}+2\times\frac{1}{8}\right)=\frac{1}{3}$$

消息 u_1 的后验概率 $P(u_1|x_0)$ 为

$$P(u_1|x_0)=\frac{1}{4}\Big/\left(2\times\frac{1}{4}+2\times\frac{1}{8}\right)=\frac{1}{3}$$

消息 u_2 和 u_3 的后验概率为

$$P(u_2|x_0)=P(u_3|x_0)=\frac{1}{8}\Big/\left(2\times\frac{1}{4}+2\times\frac{1}{8}\right)=\frac{1}{6}$$

这些后验概率列入表 2-1 的第四列.

当译码器又收到第二个码元 1(即 y_1)后，由于连续收到两个码元 01(即 x_0y_1)，所以消息 u_0 和 u_1 的后验概率为零. 而消息 u_2 和 u_3 的后验概率为

$$P(u_2|x_0y_1)=P(u_3|x_0y_1)=\frac{1}{6}\Big/\left(2\times\frac{1}{6}\right)=\frac{1}{2}$$

最后，当译码器又收到第三个码元 1(即 z_1)后，由于先后收到 011(即 $x_0y_1z_1$)三个码元，

所以只有消息 u_8 的后验概率 $P(u_8|x_0y_1z_1)=1$，其它消息的后验概率都为零.

计算出后验概率之后，就能计算互信息量. 各个消息 u_i 与收到的第一个码元 x_0 之间的互信息量为

$$I(u_i;\ x_0)=\log\frac{P(u_i|x_0)}{P(u_i)}$$

例如消息 u_3 与 x_0 之间的互信息量为

$$I(u_3;\ x_0)=\log\frac{P(u_3|x_0)}{P(u_3)}=\log_2\frac{1/6}{1/8}=\log_2\frac{4}{3}=0.415\text{ 比特}$$

这说明译码器收到第一个码元 0 之后提供有关消息 u_3 的信息量为 0.415 比特.

三、条件互信息量

条件互信息量 $I(x_i;\ y_j|z_k)$ 定义为在 XYZ 联合集中，在给定 z_k 条件下，x_i 与 y_j 之间的互信息量. 用数学公式表示，条件互信息量的定义为

$$I(x_i;\ y_j|z_k)\triangleq\log\frac{P(x_i|y_jz_k)}{P(x_i|z_k)}\tag{2.6}$$

在 XYZ 联合集上还有 x_i 与 y_jz_k 之间的互信息量，其定义为

$$I(x_i;\ y_jz_k)\triangleq\log\frac{P(x_i|y_jz_k)}{P(x_i)}\tag{2.7}$$

若在上式的分子和分母上都乘以 $P(x_i|y_j)$，可得

$$\begin{aligned}I(x_i;\ y_jz_k)&=\log\left[\frac{P(x_i|y_jz_k)}{P(x_i|y_j)}\cdot\frac{P(x_i|y_j)}{P(x_i)}\right]\\&=\log\frac{P(x_i|y_j)}{P(x_i)}+\log\frac{P(x_i|y_jz_k)}{P(x_i|y_j)}\\&=I(x_i;\ y_j)+I(x_i;\ z_k|y_j)\end{aligned}\tag{2.8}$$

从而得知：一对 y_jz_k 事件出现后所提供的有关 x_i 的信息量等于 y_j 事件出现后所提供的有关 x_i 的信息量，加上在给定 y_j 的条件下再出现 z_k 事件后所提供的有关 x_i 的信息量.

结合前面的例 2，可求出在给定 x_0 条件下，各消息与 y_1 之间的条件互信息量. 例如消息 u_3 与 y_1 之间的条件互信息量为

$$\begin{aligned}I(u_3;\ y_1|x_0)&=\log\frac{P(u_3|x_0y_1)}{P(u_3|x_0)}=\log\frac{1/2}{1/6}\\&=\log_2 3=1.585\text{ 比特}\end{aligned}$$

进一步，还可求出在给定 x_0y_1 条件下，消息 u_3 与 z_1 之间的条件互信息量为

$$\begin{aligned}I(u_3;\ z_1|x_0y_1)&=\log\frac{P(u_3|x_0y_1z_1)}{P(u_3|x_0y_1)}\\&=\log\frac{1}{1/2}=\log_2 2=1\text{ 比特}\end{aligned}$$

最后可以求得整个代码组 $x_0y_1z_1$ 出现后所提供的有关消息 u_3 的信息量为 $I(u_3;\ x_0)$ 信上面两个信息量之和，即

$$\begin{aligned}I(u_3;\ x_0y_1z_1)&=I(u_3;\ x_0)+I(u_3;\ y_1|x_0)+I(u_3;\ z_1|x_0y_1)\\&=0.415+1.585+1=3\text{ 比特}\end{aligned}$$

$x_0y_1z_1$ 与 u_3 之间的互信息量也可以用下式直接计算：

$$I(u_3;\ x_0y_1z_1)=\log\frac{P(u_3|x_0y_1z_1)}{P(u_3)}=\log\frac{1}{1/8}=\log_2 8=3\text{ 比特}$$

在(2.8)式中，y_j 和 z_k 的出现次序可以交换，即可得

$$I(x_i;\ y_jz_k)=I(x_i;\ z_k)+I(x_i;\ y_j|z_k) \tag{2.9}$$

由(2.8)式和(2.9)式，又可得

$$I(x_i;\ y_jz_k)=\frac{1}{2}[I(x_i;\ y_j)+I(x_i;\ z_k)+I(x_i;\ z_k|y_j)+I(x_i;\ y_j|z_k)] \tag{2.10}$$

根据互信息量的互易性，即有

$$I(x_i;\ y_jz_k)=I(y_jz_k;\ x_i) \tag{2.11}$$

$$I(y_jz_k;\ x_i)=I(y_j;\ x_i)+I(z_k;\ x_i|y_j) \tag{2.12}$$

$$I(y_jz_k;\ x_i)=I(z_k;\ x_i)+I(y_j;\ x_i|z_k) \tag{2.13}$$

在结束本节之前，还要说明一下自信息量和互信息量的对应关系，即自信息量相当于后验概率为1时的互信息量.

第三节 通 信 熵

一、平均自信息量(熵)

前面已讲过，要唯一地确定信源的一个消息(或符号)x_i 所需要的信息量是它的自信息量 $I(x_i)$. 实际上，信源往往包含着多个消息，而且各个消息的出现概率按信源的概率空间分布. 当各个消息的出现为相互独立时，这种信源称为**无记忆信源**. 无记忆信源的平均自信息量是各消息自信息量的概率加权平均值(统计平均值)，即**平均自信息量** $H(X)$ 定义为

$$H(X)\triangleq\sum_X P(x)I(x) \tag{2.14}$$

用(2.1)式代入上式，得

$$\begin{aligned}H(X)&=\sum_X P(x)[-\log P(x)]\\&=-\sum_X P(x)\log P(x)\end{aligned} \tag{2.15}$$

所以平均自信息量 $H(X)$ 也可直接定义为

$$H(X)\triangleq-\sum_X P(x)\log P(x) \tag{2.16}$$

上式中的 $H(X)$ 是唯一地确定信源 X 中任意一个消息所需要的平均信息量.

(2.16)式是信息论的一个基本的重要公式. 因为此式与统计热力学中"熵"的表示形式相同，因此往往把平均自信息量 $H(X)$ 称为**熵**. 由(2.16)式可知，$H(X)$ 是 $P(x)$ 的函数.

根据(2.14)式，由于自信息量 $I(x)$ 为非负量，且 $0\leqslant P(x)\leqslant 1$，所以熵 $H(X)$ 也是非负量. $P(x)\log P(x)$ 仅当 $P(x)=0$ 和 $P(x)=1$ 时才等于零. 当 $P(x)=0$ 时，说明消息 x 并不出现，所以信息量为零. 当 $P(x)=1$ 时，这说明信源只有一个消息，且该消息是必然出现的，所以传输这个消息并没有提供任何信息量.

定理2.1 熵满足不等式

$$H(X)\leqslant\log M \tag{2.17}$$

式中 M 是信源 X 中的消息数目. 当且仅当信源 X 中各消息的出现概率 $P(x)$ 都等于 $\frac{1}{M}$ 时，上式取等号.

证明：此定理可借助下列不等式来证明：