信息与编码理论

王育民　梁传甲　编著

西北电讯工程学院出版社

高等学校教材

信息与编码理论

王育民 梁传甲 编著

西北电讯工程学院出版社

1986

内 容 提 要

本书重点介绍信息论的基本概念、信息量度、信道容量、信源和信道编码定理、信道编码（纠错码）理论以及多用户理论。

本书强调信息论近来的发展及其在工程实际中的应用，可作为通信、计算机、信息工程等专业高年级大学生和研究生的信息和编码理论课程的教材，也可供从事有关信息论和编码的教学、科研和工程技术人员参考。

高等学校教材

信息与编码理论

王育民 梁传甲 编著

西北电讯工程学院出版社出版

西北电讯工程学院印刷厂印刷

陕西省新华书店发行　　各地新华书店经售

开本787×1092 1/16　印张25 10/16　字数626千字

1986年3月第一版　1986年3月第一次印刷　印数1—5,000

统一书号：15322·34　　定价：5.00元

前　言

六十年代以来，信息论学科有了很大发展，在不少方面取得了新的成果。信道编码即纠错码理论已经形成，而且在通信和计算机工程实践中得到了广泛的应用，成为通信系统设计中的一项通用技术。作为数据压缩理论基础的信息速率失真理论在七十年代以来已经逐步形成，并且开始应用到数据压缩的实践中去。近十几年来，以计算机为中心的通信网迅速发展，使信息论由两端单向通信的经典理论向多用户信息论方向发展。另一方面，随着我国教育事业的迅速发展，大学有关专业比较普遍地开设了信息论课程，特别是各校广泛招收研究生以来，信息论几乎成了通信、雷达、计算机和应用数学等有关专业普遍选修的课程。在这种形势下，我们在两次编写信息论讲义的基础上，重新编写了这本教材，以期能适应科学技术和我国教育形势的新发展。

这本教材总结了我院自 1960 年创办信息论专业以来，为信息工程系各专业和其他工程专业开设信息论和编码理论课程的经验，特别是总结了近几年来为研究生开设此课的经验。书中较详细地讨论了信息论的基本理论，对信息论中新发展的若干重要课题（如率失真理论、多用户信息论等）都作了专题讨论，以反映信息论和编码理论这一学科的近况。这对于需要获得信息论基本知识的同志和想在这一领域从事研究工作的同志都将是有益的。

本书可作为有关专业大学生和研究生的教材。在给高年级大学生讲授时，可以只讲一些基本内容。书中标有＊号的章节主要供研究生阅读。若只想了解一般信息论的基本内容，第七、八章有关纠错码的部分可以从略。各章后面都附有一些难易程度不等的习题，可根据需要选用。书末附有较详尽的参考书目和参考文献，可供阅读时参考。

本书在编写中参考了一些有关著作，特别应当提到的有：Gallager[1968]，Fano[1961]，Golomb[1964]，Lin 和 Costello[1982]，McEliece[1977]，Viterbi 和 Omura[1979]，Kolesnik 和 Poltaler[1982]，Arimoto[1977]和周炯槃[1983]等教授的著作。在编写过程中得到西北电讯工程学院出版社图书编辑室的大力支持和鼓励，作者在此表示衷心感谢。

作　者

于西北电讯工程学院

1985 年 3 月

目　录

第一章 引 论

本章介绍通信系统模型、信息论研究的对象和基本方法。

§1.1 通信系统模型

“信息论”或者称为“通信的数学理论”，是研究信息的传输、存贮和处理的科学。通信的基本问题是在彼时(存贮情况)或彼地(通信情况)精确地或近似地再现此时或此地发出的消息。信息论研究的主要问题是在通信系统设计中如何实现有效性和可靠性。

各种通信系统(包括存贮系统)，如电报、电话、图象、计算机和雷达等系统，虽然它们的形式和用途各不相同，但从信息传输的角度来看，在本质上有许多共同之处。对有收发两端的单向传信系统，一般可概括为图 1.1.1 所示的模型。

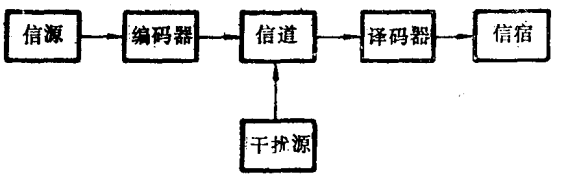


图 1.1.1

信源是产生消息的源。消息可以是文字、语言、图象等。它可以是离散的，也可以是连续的，但都是随机发生的，即在没有收到这些消息之前不可能确切地知道它们的内容，否则通信将失去意义。可以用随机变量或随机过程来描述消息。信源研究的主要问题是消息的统计特性和信源产生信息的速率。

编码器是将信源发出的消息变换成适于信道传送的信号的设备。一般包含三个部分，即**信源编码器**、**纠错编码器**和**调制器**。信源编码器是在一定的准则下，对信源的输出进行变换，目的在于求得有效性。纠错编码器是对信源编码器的输出进行变换，用以提高对于信道干扰的抗击能力。调制器将信道编码器的输出变成适合于信道传输要求(带宽、波段、功率、通信时间等)的信号形式。不一定每个系统的编码器都含有这三个部分，有的只有其中的两个或一个组成部分，也有的将其中的两个合并起来由一个组成部分实现。纠错编码器和调制器的组合又称作**信道编码器**，因为它们主要是针对信道情况进行设计的，目的在于充分利用信道的传信能力可靠地传送信息，参看图 1.1.2。

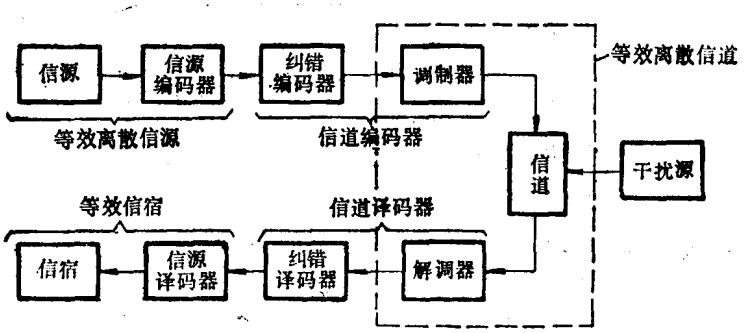


图 1.1.2

信道是将信号从发端传到收端的媒质或通道，它是包括收发设备在内的物理设施。信道的种类很多，如架空明线、电缆、表面波、声纳、光束、电离层反射、对流层散射、卡片、磁盘、磁鼓、书籍等都可看作是信道。

干扰源。为了分析方便，我们将整个通信系统中各部分引入的各种干扰，如衰落、多径、码间干扰、非线性失真、可加噪声等都集中于一个方框作用于信道。这种干扰源的统计特性是划分信道的重要因素，并且是决定信道传输能力的决定因素。信道的中心课题是研究信道的统计特性和它的传信能力，即**信道容量**。实际干扰可分成两大类。一类是由外界引入的随机干扰，如天电干扰、设备内部的噪声，它们与信道输入信号统计无关，信道的输出就是输入和干扰的和，所以这种干扰又称为**加性干扰**。另一类干扰是信号在传播过程中由于物理条件的变化(如温度、电离层位置随机变化等)引起信号参量(如频率色散、幅度衰减、相位偏移等)随机变化，此时信道的输出信号是输入信号与某些随机变量相乘的结果，所以又称这种干扰为**乘性干扰**。为了实现可靠通信，就要与各种干扰作斗争，这是通信系统设计中的一个基本问题。信息论要对干扰进行数学上的定量描述，以确定它们对传信能力影响的大小，从而给出有干扰下信道的传信能力。

译码器是编码的逆变换，它要从受干扰的信号中最大限度地提取出有关信源输出消息的信息，应尽可能精确地恢复信源的输出，并将它们送给信宿。中心问题是研究各种可实现的解调和译码方法。

信宿是信息的接收者，可以是人或物，且与信源处于不同地点或存在于不同时刻。它要对传过来的消息提出可接受条件，即提出一定的准则，发端将以此来确定对信源处理时所要保留的最小信息量。至于信宿本身的主观因素，在 Shannon 信息论中不加过问。

系统的模型不是不变的，它根据实际情况而定。如图 1.1.2 将图1.1.1中的编、译码器作了更细致的划分，目的是使信源编码的研究主要和信源及信宿发生关系，此时信道编、译码器和信道的组合可等效为一个离散无扰信道。而信道编码的研究可和信源、信宿无关，而只与信道有关，此时信源和信源编码器组合成一个对于信道编码器来说的等效离散源，它的输出可近似地看成是无记忆、输出等概的数字序列。这种划分将使信源编码集中于解决在满足信宿要求下的有效性问题，而信道编码则集中于解决抗信道干扰和失真问题，即解决传输可靠性问题，从而简化了研究。这样划分是否会对发挥通信系统的传信潜力有根本性的限制呢？研究表明，在很一般的条件下，对大多数理论结果没有太大限制。这样划分也不一定总是合理的，有时将信源编码和信道编码统一考虑进行设计可能更有效些，有人已进行了这方面的探讨[Davisson 1973]。

图 1.1.1 和 1.1.2 给出的模型适用于两个用户(或终端)之间的单向通信情况。在网通信情况下，可能有很多分开的信源、信道和信宿，进行信息交换。为了研究网通信系统中的信息传输和处理问题，要对上面给出的单路系统进行修正，而引入**多用户通信系统**，并将单路通信的信息理论发展成为**多用户信息理论**，这是近十多年来信息论研究中的一个十分活跃的课题。本书第十章中将介绍其中的一些基本结果。

上面多次提到了信息、消息和信号，在结束本节时，我们想澄清这三个不同的但又密切相关的概念。

信息是一个抽象的概念，但它可以定量地描述。信息、物质和能量被认为是构成一切系统的三大要素，它是系统中传送(或存贮、处理)的对象，它包含在消息之中。消息是比较具

体的概念，但不是物理的，如语言、文字、数字、图象等。消息中载荷有信息，但同一信息可以由不同的消息载荷。例如，同一信息可以用语言表达，也可用文字表达。信号是表示消息的物理量，如电信号可通过幅度、频率、相位的变化表示不同的消息。

§1.2 Shannon信息论的中心问题

Shannon 信息论的基本任务是为设计有效而可靠的通信系统提供理论依据。通信的基本目的是在收端精确地或以给定的失真度重现信源的输出。信源编码器的作用是根据失真度准则对信源的输出进行划分，给每一类以不同的表示，即码字。信源译码器的任务是根据收到的信源表示恢复出信源所属的类。显然，精确度要求越高，即失真度要求越小，对信源的划分就要越细，因而为表示信源所需的信息量或码长就越大。在给定信源和失真度条件下，要多大信息速率才行？或对给定信源保留一定的信息速率下，可以达到的最小失真是多少？这是信息论所关心的信源编码问题，也是通信“可行性”的一个问题。另一个问题是如何实现这一理论结果，即找出实际可行的信源编码和译码方法。

信息论研究的另一个主要问题是信道编码问题。它和信源编码问题类似，但不是最有效地表示信源输出，而是在保证信息传输可靠性(如错误概率小于给定值)的条件下最有效地利用信道的传信能力。设送入信道的信息速率为 R，信道容量为 C，信道编码基本定理告诉我们，若 $R<C$ 则可以将速率为 R 的信息以任意高的可靠性送至收端；若 $R>C$，则不可能。这是信道编码和“可行性”问题。信道编码的另一个问题是寻找实际可行的编译码方法。

信息论在研究信源和信道编码定理时所用的方法都是随机编码方法，它是证明定理的重要工具，但为非构造性的，因此不能提供具体的可实现的编译码方法。由 Hamming[1951] 开始的可构造的编码理论在五十年代末和六十年代得到了很大的发展，成为信息论中的一个重要分支——纠错编码理论[Peterson 1961]。信源编码理论在七十年代有了很大发展，形成速率失真理论[Shannon 1959 和 Berger 1971]和数据压缩技术[Davisson 和 Gray 1976, Gilbert 1983]。

上述信息论的基本问题在 Shannon[1948]的早期著作中都已系统地提出并给出了启发式的证明。Shannon 信息论的最大特点是将概率统计的观点和方法引入到通信理论研究中，揭示了通信系统中传送的对象是信息，并对信息给出科学的、定量的描述，指出通信系统设计的中心问题是在随机噪声干扰下如何有效而可靠地传送信息，实现这一目标的途径是编码(信源编码和信道编码)，并且从理论上证明了可以达到的最佳性能限。

除了 Shannon 以外，在差不多同一时候还有几位学者也明确地提出用统计观点研究信息问题。如统计学家 R.A.Fisher 从古典统计理论出发给出了信息的定义[Kullback 1959]。N.Wiener[1948, 1949]从控制论和噪声中提取信息的最佳滤波器设计角度研究信息。在通信理论中，Wiener 的过滤接收理论被看作是通信理论中的一个重要分支。V.A.Kotelnikov [1956]也用概率统计的方法研究接收信号的检测与估值、模拟调制和数字调制信号的最佳解调问题，这一方向被称作是最佳接收理论。Shannon 的信息论是从整个通信系统的最佳化来研究信息的传送和处理问题的。

在同一时期有这样多的学者都以概率统计的观点研究信息的产生、传送和处理，发表了具有划时代意义的著作，这绝非偶然的事，它是通信工程和通信理论发展的必然结果。有关

信息论产生和发展的情况可参阅 Cherry[1951], Pierce[1973], Slepain[1973], Viterbi [1973], Wyner[1974], Wolf[1973], Kotz[1966], Dobrushin[1961, 1972], Price[1984], 池田止戈夫和广田修[1983]等。

信息论常被理解为包括更广的领域，如语义学、语言学、神经生理学、心理学和组织学等，并被用于更广的边缘学科中，这常被称作是“广义”信息论。本书所讨论的 Shannon信息论则被称作是“狭义”信息论或经典信息论。当然，在信息时代中，人们对于信息的理解是非常广泛的，如语义信息、生物信息(含遗传信息)、经济信息、管理信息、自然信息等等，它们远远超出了 Shannon 信息论的讨论范围，这要求进一步认识和发展信息概念和信息理论。目前虽有些讨论，但还远未成熟。因此本书只限于讨论在通信科学中已建立了完整理论并取得重大技术成就的经典信息论，这对于正确理解信息和信息理论，以及进一步发展信息论是必需的和极有帮助的。

第二章　信息量和熵

为了定量地研究通信系统，首先要建立信息量的概念，以便度量各种通信系统中最本质的东西，即信息量的大小。信息的量度与我们最熟悉的能量的情况很类似。大家知道，能量可以从一种形式转换成另一种形式，在很多情况下可以存贮和传输，且有用能量常常因为热损耗而减少，但绝对不可能增加。与此类似，在通信系统中信息的概念也是很重要的，信息可以处理、存贮和传输，但由于干扰的影响，不管对接收到的信息怎样处理，信息只会减少，绝不可能增加。

本章首先给出离散情况下的信息量、平均信息量和熵的定义，讨论它们的性质和各种关系式，而后考虑连续变量及随机过程等情况下的信息量和熵。

§2.1　离散变量的非平均信息量

通信系统模型的每个方框都是对输入作某种变换，其共同特点是每个框的输出都和输入发生一定的关系，因此输出都可在一定程度上限定其输入。通信的目的是在接收端精确地或以给定的失真重现发送的消息。

我们首先讨论输入、输出均为离散的情况，它们可用离散概率空间描述。令 X 和 Y 分别表示输入离散事件集和输出离散事件集。其中 $X=\{x_k,\ k=1,2,\cdots,K\}$，对每个事件 $x_k\in X$，相应概率为 $q(x_k)$，简记作 q_k，且

$$\begin{aligned} &q_k\geqslant 0 \qquad k=1,2,\cdots,K \\ &\sum_{k=1}^{K} q_k=1 \end{aligned} \tag{2.1.1}$$

以 $\{X,\ q(x)\}$ 表示输入概率空间。类似地有 $Y=\{y_j,j=1,2,\cdots,J\}$，对每个事件 $y_j\in Y$，相应概率为 $\omega(y_j)$，简记作 ω_j，且

$$\begin{aligned} &\omega_j\geqslant 0 \qquad j=1,2,\cdots,J \\ &\sum_{j=1}^{J} \omega_j=1 \end{aligned} \tag{2.1.2}$$

以 $\{Y,\omega(y)\}$ 表示输出概率空间。X 和 Y 的联合空间

$$XY=\{x_k y_j;\ x_k\in X,\ y_j\in Y,\ k=1,2,\cdots,K;\ j=1,2,\cdots,J\}$$

与每组事件 $x_k y_j\in XY$ 相应的概率为 $p(x_k y_j)$，且

$$\sum_{k=1}^{K}\sum_{j=1}^{J}p(x_k y_j)=1$$

$$q_k=\sum_{j=1}^{J}p(x_k y_j) \tag{2.1.3}$$

$$\omega_j=\sum_{k=1}^{K}p(x_k y_j)$$

以$\{XY,p(xy)\}$表示联合概率空间。一般在给定事件x_k下，事件y_j出现的条件概率为

$$p(y_j|x_k)=\frac{p(x_k y_j)}{q(x_k)} \qquad q(x_k)>0 \tag{2.1.4}$$

类似地，在事件y_j出现条件下事件x_k出现的概率为

$$p(x_k|y_j)=\frac{p(x_k y_j)}{\omega(y_j)} \qquad \omega(y_j)>0$$

当事件x_k和y_j彼此独立时有

$$p(x_k y_j)=q(x_k)\omega(y_j) \tag{2.1.5}$$

若上式对所有k和j均成立，则集合X和Y为统计无关或统计独立，否则称为统计相关。

一、非平均互信息

在引入信息量定义之前，首先举例说明输出事件和输入事件之间的概率关系。

例 2.1.1 设输入空间$X=\{x_1, x_2, \cdots, x_8\}$，$q(x_k)=1/8$，$k=1,\cdots,8$。将各消息$x_k$分别以三位二元数字表示，并作为输出事件。我们来看通过对输出事件的观察如何推测输入的消息。假定系统的输入消息为x_4，则输出为011，称这三个二元数字为码字。作为观察者，我们只知道各x_k出现的概率相等，而不知道输入是哪个消息。当观察到输出的二元数

表 2.1.1

输入消息	码字(输出)	消息先验概率	消息后验概率		
			收到0后	收到01后	收到011后
x_1	000	1/8	1/4	0	0
x_2	001	1/8	1/4	0	0
x_3	010	1/8	1/4	1/2	0
x_4	011	1/8	1/4	1/2	1
x_5	100	1/8	0	0	0
x_6	101	1/8	0	0	0
x_7	110	1/8	0	0	0
x_8	111	1/8	0	0	0

字后，用后验概率公式易于算出各消息 x_k 的后验概率，如表 2.1.1 所示。由表 2.1.1 可以看出，每收到一个数字之后，各消息出现的后验概率都作相应变化，这有助于我们对输入发生的事件进行猜测。在接收 011 三个数字的过程中，消息 x_4 出现的后验概率逐步增加，最终达到 1，而其它消息出现的后验概率都先后减小到 0，从而完全确定出输入消息。

当输入消息的先验概率不等时，后验概率的变化情况有所变化，如表 2.1.2 所示，但总的趋势仍然是某个消息出现的后验概率逐步增加到 1，而其它消息的后验概率最终变为 0，从而完全确定了输入端发生的事件。但变化情况不同。输入等概时，x_4 出现的概率从 1/8 变为 1；而输入不等概时，x_4 出现的概率从 1/4 变为 1。对观察者来说，同样观察事件 011，但输入消息等概情况下“收获”要大些，即得到的“信息”要多些。 #

表 2.1.2

输入消息	码字(输出)	消息先验概率	消息后验概率		
			收到 0 后	收到 01 后	收到 011 后
x_1	000	1/8	1/6	0	0
x_2	001	1/4	1/3	0	0
x_3	010	1/8	1/6	1/3	0
x_4	011	1/4	1/3	2/3	1
x_5	100	1/16	0	0	0
x_6	101	1/16	0	0	0
x_7	110	1/16	0	0	0
x_8	111	1/16	0	0	0

例 2.1.2 设某系统的输入空间为 $X=\{x_1, x_2\}$，分别以二元数字组 000 和 111 表示。若系统变换过程中的转移概率为 $p(0|0)=p(1|1)=1-p, p(1|0)=p(0|1)=p$，则不难算出当观察到输出数字为 010 的过程中输入消息 x_1 和 x_2 的后验概率变化，如表 2.1.3 所示。

表 2.1.3

输入消息	二元数字表示	消息先验概率	消息后验概率		
			收到 0 后	收到 01 后	收到 010 后
x_1	000	1/2	$1-p$	1/2	$1-p$
x_2	111	1/2	p	1/2	p

在收 010 的过程中，消息出现的可能性，即后验概率也在不断地变化，但变化的趋势不再象例 2.1.1 那样单调地变化，而是有起伏的，且最后并未达到 1 或 0。这就是说，我们观察到 010 之后不能断定是那个消息出现了。但是由观察结果计算出来的某个消息出现的后验概率大于 1/2 或小于 1/2，使我们可比未观察前较有把握地推断出消息的出现，因而多少得到了一些有关消息出现的“信息”。 #

从上述两个系统可以看出，在一个系统中我们所关心的输入是哪个消息的问题，只与事件出现的先验概率和经过观察后事件出现的后验概率有关。这就是说信息量应当是先验概率和后验概率的函数，即

$$I(x_k;\ y_j)=f(q(x_k),\ p(x_k|y_j)) \tag{2.1.6}$$

其中，x_k 表示系统可能的输入消息，如例 2.1.1 中的 x_1，x_2，…，x_8；而 y_j 表示系统可能的输出事件，如例 1 中的 8 个码字。

如果我们令收到的三个二元数字分别以 y_{j1}，y_{j2} 和 y_{j3} 表示，则每接收一个数字得到的信息量应分别为

$$\left.\begin{aligned}
&I(x_k;\ y_{j1})=f(q(x_k),\ p(x_k|y_{j1}))\\
&I(x_k;\ y_{j2}|y_{j1})=f(p(x_k|y_{j1}),\ p(x_k|y_{j1}y_{j2}))\\
&I(x_k;\ y_{j3}|y_{j1}y_{j2})=f(p(x_k|y_{j1}y_{j2}),p(x_k|y_{j1}y_{j2}y_{j3}))
\end{aligned}\right\} \tag{2.1.7}$$

而且应当满足

$$I(x_k;\ y_j)=I(x_k;\ y_{j1})+I(x_k;\ y_{j2}|y_{j1})+I(x_k;y_{j3}|y_{j1},y_{j2}) \tag{2.1.8}$$

即满足

$$\begin{aligned}
f(q(x_k),\omega(y_j))=&f(q(x_k),p(x_k|y_j)+f(p(x_k|y_{j1}),p(x_k|y_{j1}y_{j2}))\\
&+f(p(x_k|y_{j1}y_{j2}),p(x_k|y_{j1}y_{j2}y_{j3}))
\end{aligned} \tag{2.1.9}$$

选择什么样的函数才能满足(2.1.6)式和(2.1.8)式的要求呢？可以证明，对数函数才能满足这些要求。这可从例 2.1.1 说明。当某个可能的消息，例如 x_4 输入后，我们可按下述方式进行猜测事件发生的情况。假定出现的是前 4 个消息中的某一个，猜对了（由接收的第1个数字 0 判定）；再假定是前两个中的一个出现，猜错了(由接收的第二个数字 1 来判断)；则一定是 x_3 或 x_4 中之一出现，再猜一次就能完全肯定是 x_4 出现了（由第三个数字为 1 判定）。在 8 个消息等概下，我们猜三次就足以判定消息的出现。在 16 个等概消息时，四次猜测就可判定。一般，n 次猜测可以确定 $M=2^n$ 个等可能事件中某一事件的出现。我们每猜一次得到一定的信息量，猜 n 次应得到 n 倍于一次猜测的信息量。这样，在等可能条件下，信息量是事件数目的对数函数，即

$$I(x_k;\ y_j)=\log_2 M=-\log\frac{1}{2M} \tag{2.1.10}$$

其中 $1/M$ 是事件 x_k 出现的先验概率，1 是经过 n 次猜测后事件的后验概率。这样，信息量就被表示成为事件的后验概率与事件的先验概率之比的对数函数了。下面给出一般情况下信息量的定义。

定义 2.1.1 对给定的两个离散事件集$\{X,q(x_k)\}$和$\{Y,\omega(y_j)\}$。事件 $y_j\in Y$ 的出现给出关于事件 $x_k\in X$ 的**信息量** $I(x_k;\ y_j)$定义为

$$I(x_k;\ y_j)\triangleq\log_a\frac{p(x_k|y_j)}{q(x_k)} \tag{2.1.11}$$

上式中对数的底 $a(>1)$可任意地选择，不同的底决定不同的信息量单位。最常用的底是“2”和“e”。以 2 为底时信息的单位称作**比特**(bit)，即二进制单位。实际中，大多数逻辑电路器件均为两个状态，一比特信息正是我们对等可能的两个事件进行猜测时所需的信息，所以又称作**是否信息**。以“e”为底时信息的单位称作**奈特**(nat)，它在分析时比较方便。不同单位之间可通过换算式

$$\log_a x = \log_a b \cdot \log_b x \tag{2.1.12}$$

计算，可得出 1 bit = 0.693 nat，1 nat = 1.443 bit。

类似于 $I(x_k;\ y_j)$可定义事件 $x_k \in X$ 出现给出的关于事件 $y_j \in Y$ 的信息量为

$$I(y_j;\ x_k) = \log \frac{p(y_j|x_k)}{\omega(y_j)} \tag{2.1.13}$$

而由

$$p(x_k y_j) = q(x_k)p(y_j|x_k) = \omega(y_j)p(x_k|y_j)$$

不难得出

$$I(x_k;\ y_j) = I(y_j;\ x_k) = \log \frac{p(x_k y_j)}{q(x_k)\omega(y_j)} \tag{2.1.14}$$

这就是说，事件 x_k 出现给出的关于事件 y_j 的信息量等于事件 y_j 出现给出的关于事件 x_k 的信息量。因此，$I(x_k;\ y_j)$被称作是事件 x_k 与事件 y_j 之间的**互信息量**。式 (2.1.14) 正反映了互信息的**对称性**。

事件之间所以有互信息是因为两个事件 x_k 和 y_j 之间统计相关。若事件 x_k 和事件 y_j 彼此独立，即 $p(x_k y_j) = q(x_k)\omega(y_j)$，则 $I(x_k;\ y_j) = \log 1 = 0$。我们不能从对一个事件的观察获得关于另一事件的任何信息。

若事件 y_j 的出现有助于肯定事件 x_k 的出现，即 $p(x_k|y_j) > q(x_k)$，则 $I(x_k;\ y_j) > 0$。反之，若事件 y_j 的出现告诉我们事件 x_k 出现的可能性减小了，即 $p(x_k|y_j) < q(x_k)$，则 $I(x_k;\ y_j) < 0$。所以两个事件之间的互信息量可正、可负、也可能为 0。

二、条件互信息与联合事件的互信息

可以将两个概率空间中事件之间的互信息概念推广到三个概率空间中事件之间的互信息。设有乘积空间 $U_1U_2U_3 = \{u_1u_2u_3;\ u_1 \in U_1,\ u_2 \in U_2,\ u_3 \in U_3\}$，其中 U_1、U_2 和 U_3 都是离散事件集，且有下述概率关系式：

$$\sum_{u_1}\sum_{u_2}\sum_{u_3} p(u_1u_2u_3) = 1 \tag{2.1.15}$$

$$\begin{aligned}
p(u_1u_2) &= \sum_{u_3} p(u_1u_2u_3) \\
p(u_1u_3) &= \sum_{u_2} p(u_1u_2u_3) \\
p(u_2u_3) &= \sum_{u_1} p(u_1u_2u_3)
\end{aligned} \tag{2.1.16}$$

$$\begin{aligned}
p(u_1) &= \sum_{u_2} p(u_1u_2) = \sum_{u_3} p(u_1u_3) \\
p(u_2) &= \sum_{u_1} p(u_1u_2) = \sum_{u_3} p(u_2u_3) \\
p(u_3) &= \sum_{u_1} p(u_1u_3) = \sum_{u_2} p(u_2u_3)
\end{aligned} \tag{2.1.17}$$

定义 2.1.2 对于三个离散事件集的联合概率空间$\{U_1U_2U_3,\ p(u_1u_2u_3)\}$，在给定事件

$u_3 \in U_3$ 条件下。事件 $u_1 \in U_1$ 和事件 $u_2 \in U_2$ 之间的**条件互信息量**定义为

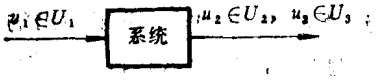
$$I(u_1; u_2|u_3) = \log \frac{p(u_1|u_2u_3)}{p(u_1|u_3)} = \log \frac{p(u_1u_2|u_3)}{p(u_1|u_3)p(u_2|u_3)} \tag{2.1.18}$$

条件互信息量的定义和无条件互信息量的定义之间的差别仅在于它的先验概率和后验概率均为某种特定条件下的取值。这个定义可以推广到任意有限多个空间情况。对于 N 个空间 $U_1, U_2, \cdots, U_N$ 中的事件 $u_1, u_2, \cdots, u_N$，可以考察 $u_1, u_2, \cdots, u_{N-2}$ 已知条件下 u_{N-1} 和 u_N 之间的条件互信息量为

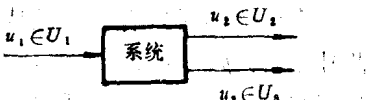
$$I(u_N; u_{N-1}|u_1u_2\cdots u_{N-2}) = \log \frac{p(u_N|u_1u_2\cdots u_{N-1})}{p(u_N|u_1u_2\cdots u_{N-2})} \tag{2.1.19}$$

条件互信息量也具有对称性，即

$$I(u_N; u_{N-1}|u_1u_2\cdots u_{N-2}) = I(u_{N-1}; u_N|u_1u_2\cdots u_{N-2}) \tag{2.1.20}$$



(a)



(b)

图 2.1.1

若将 U_1 作为一个系统的输入空间，而 U_2 和 U_3 作为系统的输出空间，其中 U_2 和 U_3 可为并行或按时间前后的串行，如图 2.1.1 所示。现在要问，知道事件 $u_2 \in U_2$，$u_3 \in U_3$ 后总共给出的有关 $u_1 \in U_1$ 的信息量是多少？它与 u_2 和 u_3 单独提供给 u_1 的信息量有什么关系？由互信息的定义有

$$\begin{aligned} I(u_1; u_2u_3) &= \log \frac{p(u_1|u_2u_3)}{p(u_1)} \\ &= \log \frac{p(u_1|u_2)p(u_1|u_2u_3)}{p(u_1)p(u_1|u_2)} \\ &= \log \frac{p(u_1|u_2)}{p(u_1)} + \log \frac{p(u_1|u_2u_3)}{p(u_1|u_2)} \\ &= I(u_1; u_2) + I(u_1; u_3|u_2) \end{aligned} \tag{2.1.21}$$

这意味 u_2u_3 联合给出的关于 u_1 的信息量等于 u_2 给出的关于 u_1 的信息量与 u_2 已知条件下 u_3 给出的关于 u_1 的信息量之和。这是互信息量的另一个重要性质，称其为**可加性**。它与直观概念一致。

不难推出

$$I(u_1; u_2u_3) = I(u_1; u_3) + I(u_1; u_2|u_3) \tag{2.1.22}$$

由(2.1.21)和(2.1.22)可推出对于 u_2 和 u_3 的对称关系式

$$I(u_1; u_2u_3) = \frac{1}{2}[I(u_1; u_2) + I(u_1; u_3) + I(u_1; u_3|u_2) + I(u_1; u_2|u_3)] \tag{2.1.23}$$

由互信息的对称性有

$$I(u_1; u_2u_3) = I(u_2u_3; u_1) = I(u_2; u_1) + I(u_3; u_1|u_2) \tag{2.1.24}$$

即事件 u_1 给出的关于联合事件 u_2u_3 的信息量等于 u_1 给出的关于 u_2 的信息量加上 u_2 已知条件下 u_1 给出的关于 u_3 的信息量。

上述关系式易于推广到任意有限维空间的情况。可加性对于表述多个事件之间的互信息量非常方便，特别是在多用户信息论中十分有用。

例 2.1.3 由例 2.1.1 中表 2.1.2 的数据可以算出：