21世纪高等学校教材

王海燕

信息论基础

XINXILUN JICHU

东南大学出版社
SOUTHEAST UNIVERSITY PRESS

信息论基础

王海燕

东南大学出版社

内 容 提 要

本书主要介绍信息论中香农基本理论，即信息的度量、信道容量以及信源和信道编码理论等问题。全书共分7章，内容包括绪论、随机变量的信息度量、离散信源及其信息度量、离散信源的无失真编码、信道及信道容量、有噪信道编码、信息率失真函数和限失真信源编码等。每章后面附有习题。

本书可作为信息与计算科学、应用数学、通信工程、信息工程等专业本科生的教材，也可为其他相关专业同类课程所选用。

图书在版编目(CIP)数据

信息论基础/王海燕 .—南京：东南大学出版社，2003.11

ISBN 7-81089-379-3

Ⅰ.信… Ⅱ.王… Ⅲ.信息论-高等学校-教材 Ⅳ.G201

中国版本图书馆 CIP 数据核字(2003)第 096459 号

东南大学出版社出版发行

(南京四牌楼 2 号 邮编 210096)

出版人：宋增民

江苏省新华书店经销 姜堰市晨光印刷有限公司印刷

开本：700mm × 1000mm 1/16 印张：13 字数：255 千字

2003 年 12 月第 1 版 2003 年 12 月第 1 次印刷

印数：1-5000 定价：18.00 元

前　言

从1999年秋季开始，各高等院校开始按教育部颁布的经调整后的新专业目录进行招生，信息与计算科学专业是这次调整中新设置的一个理科专业。信息论基础是信息与计算科学专业的一门专业基础课，也是信息工程、通信工程等相关专业的专业基础课。

长期以来，信息论方面的著作和教材主要有两种类型，一种类型的教材主要致力于信息论的公理体系与更一般的抽象的数学模型，对信息论的基本定理给出较为一般的结果，比较注重理论上的严密和完整，因此在学习时比较抽象。另一种类型的教材侧重于工程技术中的应用，致力于信息可靠传输的可实现性，比较受工程技术人员的欢迎。这两类著作和教材各有特点和阅读对象。作为数学系信息与计算科学专业信息论基础课程的教材，本书在介绍香农基本理论的同时，尽量注重理论上的严格性和完整性，但同时避免太深奥的数学知识，以便工科相关专业的学生选读。

本书主要介绍信息论中香农基本理论，即信息的度量、信道容量以及信源和信道编码理论等问题。全书共分7章。第1章绪论主要介绍了通信系统的基本模型及信息论研究的对象、目的和内容。第2章是为后面各章作数学上准备的，主要介绍随机变量的各种信息度量，即离散和连续随机变量熵和互信息的概念及性质。第3章介绍信源的数学模型及分类，离散无记忆信源、离散平稳信源和马尔可夫信源的信息度量。第4章介绍了信源输出信息的有效表示问题，即信源的无失真编码问题，主要包括离散无记忆信源的等长和不等长编码定理，离散平稳信源和马尔可夫信源的编码定理以及离散无记忆信源的一些常用的不等长编码方法。第5章介绍了信道的数学模型及分类，主要包括离散无记忆信道容量的概念及各种计算方法、信道的组合方式以及某些特殊的连续信道的容量问题。第6章介绍了有噪信道的编码问题，包括译码规则、编码方法等对信息在信道传输的影响以及在有噪信道中实现信息可靠传输的有噪信道编码定理。第7章介绍了信息率失真函数的概念、计算方法以及在允许一定失真的情况下信源的编码问题。除第1章外，每章后面都附有习题。

本书是为信息与计算科学专业信息论基础这门课编写的教材，共48学时，也可作为其他相关专业如信息工程、通信工程等同类课程的教材。

本书的编写是在东南大学数学系和教务处的支持下进行的，东南大学教务处把信息论基础作为校级重点建设课程，并把本书作为“十五”校级重点规划教材，本书的出版也得到了东南大学出版社的大力支持，在此一并表示感谢。

限于水平，书中难免存在一些错误和不足，希望广大读者批评指正。

作者

2003年10月

目　录

1 绪 论

信息论是关于信息的理论，它有明确的研究对象和适用范围，但从信息论诞生的那时起人们就对它有不同的理解. 随着信息和信息科学对现代社会生活各方面影响的不断加大和深化，人们对信息论的意义的认识和价值的估计也不断变化. 本章将简要地从工程技术或技术科学的角度讨论什么是信息，什么是信息科学和信息论，并介绍信息论研究的对象、目的和内容.

1.1 信息、信息科学和信息论

信息(information)在牛津英文字典里给出的解释是“某人被通知或告知的内容、情报、消息”，在这样的解释中，信息一词不是作为科学名词或技术术语来定义的. 实际上在不同的字典中对信息一词有不同的解释，更不用说不同领域的人们对信息有不同的理解了. 尽管信息一词的含义模糊和难以捉摸，但人人都感受到它的存在，每时每刻我们都通过对周围世界的观察去获取它，并且通过一定的方法把它传送给别人、进行交换或把它存储起来留作以后使用. 这种目前尚未明确定义的信息称为广义意义上的信息.

信息作为技术术语广泛使用是在计算机特别是微处理器得到广泛应用以后的事. 在计算机发展的早期，计算机处理的对象仍沿用过去的名词，如数据、记录、报表、文字等等. 但随着计算机的不断发展，无论在计算机学术界还是工业界都产生一种明显的倾向，即希望有一个名称能把所有这些处理对象统统包含在内，信息这一名称恰好符合这一要求，因为只有这样一个含糊的术语才能对多种多样且在不断涌现的对象得到一个统一的、全面的、不需时时改变的表达. 信息作为一个可以用严格的数学公式定义的科学名词首先出现在统计数学中，随后又出现在通信技术中. 无论是统计数学中还是在通信技术中定义的信息都是一种统计意义上的信息，可以称之为统计信息. 统计信息是非常明确的，同时其适用范围要比广义信息狭隘得多. 本书中讨论的信息论正是关于这种统计信息的理论.

信息科学作为一个名词，最早出现在图书馆学中，主要研究图书文献的检索. 在计算机出现以后，信息科学被赋予新的含义，但在不同国家中它的含义不尽相同. 在日本信息科学的含义和美国的计算机科学的含义相似，主要研究科学计算的理论和方法. 而在美国信息科学原先主要指科学计算以外，如商业、服务业、管理统计部门等所需要的涉及大量数据但计算比较简单的数据处理问题. 20 世纪

80年代以来，信息科学的含义不断扩大，不但逐渐把计算机科学的内容统一包含在内，而且有把信息技术涉及的所有科学理论统统包含在内的趋势.

信息论是信息科学的基本理论. 从历史上看信息论的形成是两部分人共同努力的结果，一部分是通信工程方面的学者，另一部分是统计数学家. 这两部分人虽然研究的是同一领域的问题，但他们感兴趣的方面和侧重点是有差异的. 这种情况从信息论产生时起一直保持到现在，今天从事信息论研究工作的人仍然由这两部分人组成.

1.2 信息论研究的对象、目的和内容

信息论研究的核心问题是信息的传输和处理，我们把各种通信系统中具有共同特点的部分抽象出来，概括成一个统一的理论模型(如图1.1所示)，通常称它为通信系统的基本模型.

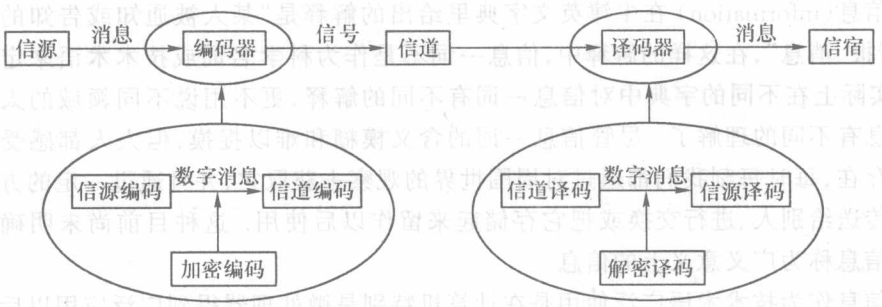

图1.1 通信系统基本模型

信息论的研究对象正是这种统一的通信系统模型，人们通过系统中消息的传输和处理来研究信息传输和处理的共同规律.

这个模型由以下五个部分组成：

(1) 信源. 信源是产生消息或消息序列的源. 消息通常是符号序列或时间函数，消息取值服从一定的统计规律，所以信源的数学模型可以是一个离散的随机序列或连续的随机过程.

(2) 编码器. 编码是消息变换成信号的措施，编码器输出的是适合信道传输的信号，信号携带着消息，它是消息的载荷者.

编码可分成两种，即信源编码和信道编码. 信源编码是对信源输出的消息进行适当的变换和处理，把信源产生的消息变换成数字序列，目的是为了提高信息传输的效率，在不允许编码失真的情况下，保证能从其输出数字序列无错地恢复输出消息序列的前提下，减少输出数字序列的速率，也就是在保证不失真的条件下对输出消息序列进行压缩；在允许编码失真的情况下，信源编码的目的是对给定信源，

在保证消息平均失真不超过某给定允许值的条件下，尽量减少输出数字序列的速率. 信道编码是为了提高信息传输的可靠性而对消息进行变换和处理，即把信源编码输出的数字序列变换成适合于信道传输的，由信道入口符号组成的序列. 信道编码器最主要的作用是要对其输出序列提供保护，以抵抗信道噪声和干扰.

(3) 信道. 信道在实际通信系统中是指传输信号的媒介或通道. 在狭义的通信系统中，信道有明线、电缆、波导、光纤、人造卫星、无线电传播空间等，这些都是属于传输电磁波能量的信道. 对广义的通信系统来说，信道还可以是其他的传输媒介. 信道除了传送信号以外，还有存储信号的作用，如书写通信方法.

在信道中引入噪声和干扰是一种简化的表达方式. 为了分析方便起见，把在系统其他部分产生的干扰和噪声都等效地折合成信道干扰，看成是由一个噪声源产生的，它将作用于所传输的信号上，这样，信道输出的是已叠加了干扰的信号. 由于干扰或噪声往往具有随机性，所以信道的特性也可以用概率空间来描述. 而噪声源的统计特性又是划分信道的依据.

(4) 译码器. 译码是把信道输出的编码信号(已叠加了干扰) 进行反变换. 译码器也可分成信源译码器和信道译码器.

(5) 信宿. 信宿是消息传送的对象，即消息的接收者，例如人或机器.

图 1.1 给出的模型只适用于收发两端单向通信的情况，它只有一个信源和一个信宿，信息传输也是单向的. 更一般的情况是信源和信宿各有若干个，即信道有多个输入和多个输出，另外信息传输方向也可以双向进行. 例如广播通信是一个输入，多个输出的单向传输的通信；而卫星通信网则是多个输入，多个输出和多向传输的通信. 要研究这些通信系统，只需对两端单向通信系统模型作些适当修正即可.

近年来，以计算机为核心的大规模信息网络、尤其是互联网的建立和发展，对信息传输的质量要求更高了，不但要求既快速有效又能可靠地传递信息，而且还要求信息传递过程中保证信息的安全保密，不被伪造和窜改. 因此，在编码器这一环节中还需加入加密编码，相应地在译码器中加入解密译码.

信息论研究的目的就是要找到信息传输过程的共同规律，以提高信息传输的可靠性、有效性、保密性和认证性，达到信息传输系统最优化.

所谓可靠性高就是要使信源发出的消息经过信道传输以后，尽可能准确地、不失真地再现在接收端. 而所谓有效性高，就是经济效果好，即用尽可能短的时间和尽可能少的设备来传送一定数量的信息. 然而，提高可靠性和提高有效性常常会发生矛盾，这就需要统筹兼顾，例如为了兼顾有效性，有时就不一定要求绝对准确地在接收端再现原来的消息，而是可以允许一定的误差或一定的失真，或者说允许近似地再现原来的消息. 所谓保密性就是隐蔽和保护通信系统中传送的消息，使它只能被授权接收者获取，而不能被未授权者接收. 所谓认证性是指接收者能正

确判断所接收消息的正确性，验证消息的完整性，而不是伪造的和被窜改的．有效性、可靠性、保密性和认证性四者才构成现代通信系统对信息传输的全面要求．

信息传输系统模型不是不变的，它根据信息传输的要求而定．研究信息传输有效性时，可只考虑信源与信宿之间的信源编(译)码，将其他部分都看成一无干扰信道；研究信息传输可靠性时，将信源、信源编码和加密编码都等效成一个信源，而将信宿、信源译码和解密译码都等效成一信宿；研究信息传输的保密性和认证性时，将信源和信源编码等效成一信源，将信道编码、信道、噪声源和信道译码等效成一无干扰信道，而将信源译码和信宿等效于信宿．

信息论研究的内容有以下三种层次：

(1) 狭义信息论(或称经典信息论)．它主要研究信息的度量、信道容量以及信源和信道编码理论等问题，这部分内容是信息论的理论基础，又称香农(Shannon)基本理论．

(2) 一般信息论．主要也是研究信息传输和处理问题，除了香农理论以外，还包括噪声理论、信号滤波和预测、统计检测和估计理论、调制理论、信息处理理论以及保密理论．后一部分内容主要的贡献是维纳(N.Wiener)和柯尔莫哥洛夫(A.Kolomogorov)等人的．

(3) 广义信息论．不仅包含上述两方面的内容，而且包括所有与信息有关的自然和社会领域，如模式识别、计算机翻译、心理学、遗传学、神经生理学、语言学、语义学甚至包括社会学中有关信息的问题．

综上所述，信息论是一门应用概率论、随机过程、数理统计和高等代数的方法来研究信息传输、提取和处理系统中一般规律的学科；它的主要目的是提高信息系统的可靠性、有效性、保密性和认证性，以便达到系统最优化；它的主要内容包括香农理论、编码理论、维纳理论、检测和估计理论、信号设计和处理理论、调制理论、随机噪声理论和密码学理论．

限于篇幅，本书主要介绍香农基本理论．

2 随机变量的信息度量

本章主要介绍了随机变量的各种信息度量，为后面各章作一些数学上的准备. 在2.1节引入了自信息、条件自信息、互信息和条件互信息的概念；在2.2节介绍了离散随机变量的平均自信息即熵的概念；在2.3节介绍了离散随机变量的平均互信息的概念；在2.4节讨论了离散随机变量熵和互信息的凸性；在2.5节介绍了连续随机变量的互信息和微分熵.

2.1 自(互)信息和条件自(互)信息

如果随机试验的所有可能的结果，可以用一个变量 X 来表示，在一次试验中，X 取什么数值不能事先确定，它随着试验结果的不同而变化，也就是说 X 是样本点 w 的一个函数，即 $X = X(w)$，如果它取各个值的可能性可以用概率描述，则称这个定义在样本空间 Ω 上的单值实函数 $X = X(w)$ 为随机变量，一般用大写英文字母 X, Y, Z 等表示. 随机变量 X 的某个取值称为事件，一般用小写英文字母 x，y, z 等表示，$p(x), p(y), p(z)$ 等表示这些事件发生的概率.

2.1.1 自信息

定义 2.1.1 对随机变量 X，事件 x 的自信息定义为

$$I(x) = \log \frac{1}{p(x)} = -\log p(x) \tag{2.1}$$

自信息 $I(x)$ 是事件 x 的不确定性即事件 x 发生的可能性的一种度量，表示事件 $X = x$ 发生时，事件 x 所含有或所提供的信息量. $p(x)$ 越小，则 $I(x)$ 越大，因为一事件发生的可能性越小，则当其出现时所带来的信息就越多，这与我们的感觉相符.

自信息的单位取决于对数所选取的底，如果取以2为底，则所得的自信息的单位为比特(bit，binary unit 的缩写)；如果采用以e为底自然对数，则所得的自信息的单位为奈特(nat，nature unit 的缩写)；如果采用以10为底，则所得的自信息的单位为哈特(Hart，Hartly unit 的缩写). 底的改变仅仅改变了计量的尺度，以后如不特别说明，一般都采用以2为底的对数，且为了书写简洁，把底数2略去不写.

例 2.1.1 从英文字母中任意选取一个字母时所给出的信息量是多少呢？

因为有26个英文字母，任取一个字母的概率为 $\frac{1}{26}$，所以

$$I(x) = -\log\frac{1}{26} \approx 4.7\text{bits}$$

这是任意选择一个英文字母所给出的信息量.

例 2.1.2 设随机选择一个 m 位数字的二进制数,该数的每一位可从两个不同的数字{0,1}中任取一个,因此共有 2^m 个等概率的可能组合,所以

$$I(x) = -\log\frac{1}{2^m} = m\text{bits}$$

容易证明,自信息具有以下性质:

性质 1 自信息 $I(x)$ 是非负的.

性质 2 当 $p(x) = 0$ 时,$I(x) = \infty$.

性质 3 当 $p(x) = 1$ 时,$I(x) = 0$.

性质 4 当 $p(x) > p(y)$ 时,$I(x) < I(y)$,即 $I(x)$ 是 $p(x)$ 的单调递减函数.

由于 x 是一个随机量,因此自信息 $I(x)$ 也是一个随机量. 小概率事件所包含的不确定性大,其自信息大;而出现概率大的随机事件所包含的不确定性小,其自信息小. 性质 2 和 3 正好是两种极端情况,即不可能事件一旦发生,带来的信息量非常大;而必然事件由于不包含任何不确定性,因此也不含任何信息量.

定义 2.1.2 对随机变量 X 和 Y,在二维联合集 XY 中的事件 xy 的联合自信息定义为

$$I(xy) = \log\frac{1}{p(x,y)} = -\log p(x,y) \tag{2.2}$$

其中,$p(x,y)$ 为元素 xy 的二维联合概率,即 $0 \leqslant p(x,y) \leqslant 1$,$\sum\sum p(x,y) = 1$.

当随机变量 X 和 Y 相互独立时,联合自信息有以下性质:

性质 5 当 X 和 Y 相互独立,即 $p(x,y) = p(x)p(y)$ 时,

$$I(xy) = I(x) + I(y)$$

说明两个随机事件相互独立时,同时发生得到的自信息等于这两个随机事件各自独立发生得到的自信息之和.

自信息的性质 1 ~ 5 保证了自信息的表达式(2.1)在相差一个常数的意义下是惟一的,即有以下定理.

定理 2.1.1 若函数 $I(x)$ 满足性质 1 ~ 5,则

$$I(x) = -C\log p(x) \tag{2.3}$$

其中 C 为常数.

为了证明定理 2.1.1,只需先证以下引理[1].

引理 如果实函数 $f(x)$,对 $1 \leqslant x < +\infty$ 满足:(1) $f(x) \geqslant 0$,(2) $f(x)$ 是严格单调增函数,即 $x < y$ 时,$f(x) < f(y)$,(3) $f(xy) = f(x) + f(y)$,则 $f(x)$

$= C\log x$.

证 对 $1 \leqslant x, y < +\infty$ 与任意自然数 k，总存在非负整数 n，使

$$y^n \leqslant x^k < y^{n+1} \tag{2.4}$$

取对数并除以 $k\log y$，得

$$\frac{n}{k} \leqslant \frac{\log x}{\log y} < \frac{n+1}{k} \tag{2.5}$$

另一方面，由条件(3)可得

$$f(x^k) = f(xx^{k-1}) = f(x) + f(x^{k-1}) = \cdots = kf(x) \tag{2.6}$$

因此由条件(2)及式(2.4)和(2.6)可得

$$nf(y) \leqslant kf(x) < (n+1)f(y)$$

又由条件(1)，当 $f(y) \neq 0$ 时

$$\frac{n}{k} \leqslant \frac{f(x)}{f(y)} < \frac{n+1}{k} \tag{2.7}$$

所以由式(2.5)和(2.7)有

$$\left| \frac{f(x)}{f(y)} - \frac{\log x}{\log y} \right| \leqslant \frac{1}{k}$$

由 k 的任意性，得

$$\frac{f(x)}{f(y)} = \frac{\log x}{\log y}$$

因此

$$\frac{f(x)}{\log x} = \frac{f(y)}{\log y} = C$$

即

$$f(x) = C\log x$$

在引理中取 $f\left(\frac{1}{p(x)}\right) = I(p(x))$，即得定理 2.1.1.

2.1.2 条件自信息

若 $p(x|y)$ 表示事件 y 发生时，发生事件 x 的条件概率，$p(y|x)$ 表示事件 x 发生时，发生事件 y 的条件概率.

定义 2.1.3 事件 y 发生时，发生事件 x 的条件自信息定义为

$$I(x|y) = -\log p(x|y) \tag{2.8}$$

事件 x 发生时，发生事件 y 的条件自信息定义为

$$I(y|x) = -\log p(y|x) \tag{2.9}$$

条件自信息也是随机变量，其值随着 x 和 y 的变化而变化，且条件自信息也有非负性和单调递减性.

由于 $p(x,y) = p(x)p(y|x) = p(y)p(x|y)$，因此联合自信息、自信息和条件自信息之间满足以下关系式：

$$I(xy) = I(x) + I(y|x)$$
$$= I(y) + I(x|y) \tag{2.10}$$

2.1.3 互信息

定义 2.1.4 对随机变量 X 和 Y，X 的事件 x 与 Y 的事件 y 之间的互信息定义为

$$I(x;y) = \log \frac{p(x|y)}{p(x)} \tag{2.11}$$

互信息的单位与自信息一样，取决于对数的底.

互信息具有以下性质：

性质 1 $I(x;y) = I(y;x)$.

证 $$I(x;y) = \log \frac{p(x|y)}{p(x)}$$
$$= \log \frac{p(x,y)}{p(x)p(y)}$$
$$= \log \frac{p(y|x)}{p(y)} = I(y;x)$$

性质 2 当 x 与 y 统计独立时，$I(x;y) = 0$.

证 由于 x 与 y 统计独立，即 $p(x,y) = p(x)p(y)$，则

$$I(x;y) = \log \frac{p(x,y)}{p(x)p(y)} = \log 1 = 0$$

性质 3 互信息可正可负.

当 $p(x|y) > p(x)$，即事件 y 的发生有助于事件 x 的出现时，$I(x;y) > 0$；而当 $p(x|y) < p(x)$，即事件 y 的发生使事件 x 出现的可能性小时，$I(x;y) < 0$.

性质 4 $I(x;y) \leqslant I(x)$，$I(x;y) \leqslant I(y)$.

证 由于 $p(x|y) \leqslant 1$，所以

$$I(x;y) = \log \frac{p(x|y)}{p(x)} \leqslant \log \frac{1}{p(x)} = I(x)$$

同理

$$I(x;y) = I(y;x) \leqslant \log \frac{1}{p(y)} = I(y)$$

性质 5 $I(x;y) = I(x) - I(x|y) = I(y) - I(y|x)$.

证 $$I(x;y) = \log \frac{p(x|y)}{p(x)} = -\log \frac{1}{p(x)} - \log \frac{1}{p(x|y)}$$
$$= I(x) - I(x|y)$$

同理可证另一等式.

互信息 $I(x;y)$ 与前面的联合自信息 $I(xy)$ 不同，它们满足以下关系：

性质 6 $I(x;y) = I(x) + I(y) - I(xy)$

证 $$I(x;y) = \log \frac{p(x,y)}{p(x)p(y)}$$

$$= -\log p(x) - \log p(y) + \log p(x,y)$$
$$= I(x) + I(y) - I(xy)$$

2.1.4 条件互信息

定义 2.1.5 对随机变量 X、Y 和 Z，在给定事件 z 的条件下，事件 x 与事件 y 之间的条件互信息定义为

$$I(x;y|z) = \log \frac{p(x|yz)}{p(x|z)} \tag{2.12}$$

事件 x 与联合事件 yz 之间的互信息定义为

$$I(x;yz) = \log \frac{p(x|yz)}{p(x)} \tag{2.13}$$

性质 $I(x;yz) = I(x;z) + I(x;y|z)$.

证
$$I(x;yz) = \log \frac{p(x|yz)}{p(x)}$$
$$= \log\left[\frac{p(x|yz)}{p(x)} \cdot \frac{p(x|z)}{p(x|z)}\right]$$
$$= \log \frac{p(x|z)}{p(x)} + \log \frac{p(x|yz)}{p(x|z)}$$
$$= I(x;z) + I(x;y|z)$$

性质表明：事件 x 与联合事件 yz 之间的互信息等于事件 x 与事件 z 之间的互信息加上在给定事件 z 的条件下事件 x 与事件 y 之间的互信息.

2.2 离散随机变量的平均自信息(熵)

如果随机变量 X 的取值只有有限个或可列举的无穷多个，则称 X 为离散随机变量. 设 X 的所有可能取值为 $x_1, x_2, \cdots, x_k, \cdots$，称 $\{x_k | k = 1,2,\cdots\}$ 为离散事件的集合，如果还给出了随机变量 X 取每一个值 x_k 的概率 $p(x_k) = P\{X = x_k\}$，$k = 1,2,\cdots$，则 X 的取值特征就可被完整地描述，称序列 $\{p(x_k)\}$，$k = 1,2,\cdots$ 为离散随机变量 X 的概率分布，记为

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_k & \cdots \\ p(x_1) & p(x_2) & \cdots & p(x_k) & \cdots \end{bmatrix}$$

其中 $p(x_k)$ 满足：(1) $p(x_k) \geqslant 0, k = 1,2,\cdots$；(2) $\sum\limits_k p(x_k) = 1$.

2.2.1 离散随机变量平均自信息的定义

自信息 $I(x)$ 是随机变量 X 中事件 x 所含的信息量，事件不同，它们所含的信息量也不同. 由于 $I(x)$ 也是一个随机变量，为了度量整个随机变量 X 的信息，通过数学期望，引入平均自信息.

定义 2.2.1 随机变量 $I(x)$ 的数学期望定义为离散随机变量 X 的平均自信