普通高等院校**电子信息类**应用型规划教材

信息论简明教程

梁栋 张兴 编著

XINXILUN JIANMING JIAOCHENG

北京邮电大学出版社
www.buptpress.com

普通高等院校电子信息类应用型规划教材

信息论简明教程

梁栋　张兴　编著

北京邮电大学出版社
·北京·

内 容 简 介

本书是专为应用型高等院校电子信息类专业本科生撰写的一本教材。考虑到读者的实际情况，本书选取了信息论中的核心内容和基础内容作为讲述的重点，对于信息论中比较艰深的概念和数学推导做了适当的删减。全书分为正文和附录两个部分。正文分为 7 章，内容包括：数学基础回顾、信息的度量、信源与信源熵、信道与信道容量、信源编码初步、信道编码初步等；附录分为 5 个部分，附录 A 给出了各章习题的详细解答，附录 B~E 给出了若干定理的证明以及部分算法的源程序。与同类信息论书籍相比，本书阅读门槛较低，习题解答详细，可供独立学院学生、自考生、成教生和广大自学爱好者使用。

图书在版编目(CIP)数据

信息论简明教程/梁栋，张兴编著. —北京：北京邮电大学出版社，2009

ISBN 978-7-5635-2044-2

Ⅰ. 信… Ⅱ. ①梁…②张… Ⅲ. 信息论—教材 Ⅳ. TN911.2

中国版本图书馆 CIP 数据核字(2009)第 119606 号

书　　名：信息论简明教程

作　　者：梁栋　张兴

责任编辑：李欣一

出版发行：北京邮电大学出版社

社　　址：北京市海淀区西土城路 10 号(邮编：100876)

发 行 部：电话：010-62282185　传真：010-62283578

E-mail：publish@bupt.edu.cn

经　　销：各地新华书店

印　　刷：北京源海印刷有限责任公司

开　　本：787 mm×1 092 mm　1/16

印　　张：11

字　　数：245 千字

印　　数：1—3 000 册

版　　次：2009 年 8 月第 1 版　2009 年 8 月第 1 次印刷

ISBN 978-7-5635-2044-2　　定　价：18.00 元

前　言

信息论是运用概率论与数理统计的方法研究信息、信息熵、通信系统、数据传输、密码学、数据压缩等问题的应用数学学科。目前各级各类高等院校电子、信息、通信和计算机及其相关专业的本科生、研究生都开有信息论这门专业基础课。

随着高等教育的发展，许多依托重点高等院校举办的独立学院也纷纷开设了信息论这门课程。独立学院的教学与研究型重点大学存在很多不同，例如在培养目标方面，独立学院更加侧重于应用；在生源方面，独立学院学生的基础普遍比较薄弱，尤其是数学基础不太牢固。正是基于这些不同，作者在多年教学和科研实践的基础上，同时参考借鉴了众多国内外优秀的信息论教材及参考书后，撰写了这本专门为应用型本科生使用的教材。

针对本书主要面向读者的特点，本书仅纳入了信息论中的核心内容和基础内容，对于比较艰深的概念和数学证明都做了适当舍弃，对各章的习题都给出了详细的解答，力争在降低难度的同时，使学生在阅读本书后能对信息论的基本概念和理论有清晰的理解。基于本书入门门槛低、内容删繁就简、习题解答详细等特点，本书也可供自考生、成教生和广大自学爱好者自学。

本书分为正文和附录两个部分。正文分为 7 章，第 1 章绪论介绍香农信息论的研究对象、目的、内容和发展史等；第 2 章对信息论需要用到的数学基础作了简单的回顾，基础较薄弱的读者应详细阅读本章；第 3 章介绍关于信息度量的几个重要概念：自信息、平均自信息和熵、联合熵、条件熵、互信息与平均互信息以及熵的链规则等基本概念，这是信息论的基础知识；第 4 章研究信源熵的问题，包括离散单符号信、离散多符号信源和连续信源，其中离散多符号信源是本章的重点和难点；第 5 章研究信道容量的问题，包括离散单符号信道、离散多符号信道、组合信道和连续信道等，其中关于信道容量的计算方法是本章的重点和难点；第 6 章介绍了信源编码的概念、分类和无失真信源编码，其中信源编码定理的理解是本章的难点，霍夫曼编码是本章的重点；第 7 章简单介绍了信道编码的概念、有噪信道编码定理及其逆定理以及若干种信道编码方法。附录分为 5 个部分，

附录 A 给出了各章习题的详细解答；附录 B～E 给出了若干定理的证明以及算法的源程序

全书 1～7 章及附录由梁栋统稿，张兴编写了附录中部分内容。

在本书编写过程中，参阅了国内外一些经典著作和习题集（列于书后参考文献中），在此向这些作者表示感谢，其中特别感谢《信息论基础教程》的作者李梅和李亦农。在此一并表示感谢！

由于作者能力所限，难免有疏漏之处，敬请读者赐教。

作　者

2009 年 6 月

目　录

本书中标注 * 和 ** 的内容不作为基本内容，供学时充裕的读者选读，其中标注 * * 的内容较难掌握，供学有余力的读者参考。

第1章 绪 论

1.1 信息的概念及香农信息论的发展史

信息论是运用概率论与数理统计的方法研究信息、信息熵、通信系统、数据传输、密码学、数据压缩等问题的应用数学学科。

在日常生活中，存在两个概念：消息和信息，通常人们认为消息是信息的载体。例如，当人们收到一封电报，电报可以认为是消息，电报中包含的对人来说有意义的内容可以认为是信息。信息的概念给出后，如何对信息进行度量，是信息论中首要的问题。

首先，人们收到消息后，如果消息告诉了人们很多原来不知道的新内容，人们会感到获得了很多的信息，而如果消息是人们基本已经知道的内容，人们得到的信息就不多，所以信息应该是可以度量的。第二，常识告诉人们，一个可能性极小的事件发生后带来的信息量要远大于一个可能性极大的事件发生后带来的信息量。例如连续抛掷 10 次硬币，正面全朝上这个事件会让人觉得很惊奇，获得了很多信息；而抛掷 10 次硬币 5 次朝上的事件则让人觉得很自然，没有多少信息。因此，信息的度量应该是概率的减函数。第三，信息的度量应该满足可加性，即两个独立事件同时发生带来的信息量应该等于两个事件各自信息量之和，即信息的度量函数 $I(X)$ 应该满足

$$I(XY)=I(X)+I(Y) \tag{1.1}$$

显然，满足上述 3 个要求的信息度量函数 $I(X)$ 应该具有类似对数函数的因素，1928 年，哈特莱(Hartley)首先提出了用对数来度量信息的概念，即一个消息所含有的信息量用它的所有可能取值的个数的对数来表示。比如，抛掷一枚硬币可能有两种结果：正面和反面，所以当人们得知抛掷结果后获得的信息量是 $\log_2 2=1$ bit。而一个十进制数字可以表示 0～9 中的任意一个符号，所以一个十进制数字含有 $\log_2 10=3.321\,9$ bit 的信息量。这里对数取以 2 为底，信息量的单位为 bit。

香农继承了哈特莱的工作，他进一步注意到消息的信息量不仅与它的可能值的个数有关，还与消息本身的不确定性有关。例如，抛掷一枚偏畸硬币，如果正面向上的可能性

是90%，那么当人们得知抛掷结果是反面时得到的信息量会比得知抛掷结果是正面时得到的信息量大。

一个消息之所以会含有信息，正是因为它具有不确定性，一个不具有不确定性的消息是不会含有任何信息的，而通信的目的就是为了消除或部分消除这种不确定性。比如，在得知硬币的抛掷结果前，对于结果会出现正面还是反面是不确定的，通过通信，人们得知了硬币的抛掷结果，消除了不确定性，从而获得了信息。因此，**信息**是对事物运动状态或存在方式的不确定性的描述。这就是香农信息的定义。

用数学的语言来讲，不确定性就是随机性，具有不确定性的事件就是随机事件。因此，可运用研究随机事件的数学工具——概率——来测度不确定性的大小。在信息论中，人们把消息用随机事件表示，而发出这些消息的信源则用随机变量来表示。比如，抛掷一枚硬币的试验可以用一个随机变量来表示，而抛掷结果可以是正面或反面，这个具体的消息则用随机事件表示。

人们把某个消息 x_i 出现的不确定性的大小定义为该消息的**自信息**，用这个消息出现的概率的对数的负值来表示：

$$I(x_i) = -\log\ p(x_i) \tag{1.2}$$

自信息同时表示这个消息所包含的信息量，也就是最大能够给予收信者的信息量。如果消息能够正确传送，收信者就能够获得这么大小的信息量。

香农在单个消息的信息量计算公式的基础上，进一步提出了**信源熵**的概念，信源熵被定义为信源可能发出的多个消息所包含的信息量的加权平均值，即

$$H(X) = -\sum_{i=1}^{q} p(x_i)\log\ p(x_i) \tag{1.3}$$

式中，q 表示信源消息的个数。式(1.3)的物理意义可以理解为：①信源随机发出一个消息前，这个消息所包含的平均不确定性；②信源随机发出一个消息后，该消息所含有的平均信息量；③信源随机发出一个消息后，平均看来能够消除的不确定性。在信源熵的概念中，信源用随机变量来定量描述。

在信源熵的基础上，香农进一步提出了两个信源的联合熵、条件熵和互信息的概念。联合熵表示两个信源联合后所包含的平均信息量；条件熵表示已知一个信源后，另一个信源仍包含的信息量；互信息表示一个信源所给出的关于另一个信源的平均信息量。

在上述概念的基础上，香农进一步提出了信源压缩和信道容量的概念，信源所包含的信息熵给定后，通过适当的编码方式，可以用最少的比特数来传输信息，这称为信源编码，它用来提高信息传输的有效性；信息通过信道的传输后，由于噪声和干扰的影响可能会造成损失，信源和信宿的互信息的最大值被称为信道容量，通过适当的编码方式，可以降低信息传输的错误概率，逼近信道容量，这称为信道编码，它用来提高信息传输的可靠性。

香农对于信息论的若干结论，被总结为香农三大定理，这些定理被收录于香农的两篇经典论文中：《A Mathematical Theory of Communication》、《Communication Theory of Secrecy Systems》，其中第一篇论文发表于1948年，被认为是信息论的奠基性著作。

1.2 香农信息论的研究对象、目的和内容

香农信息论的研究对象是广义的通信系统，它把所有的通信系统抽象为一个统一的模型，如图1.1所示。

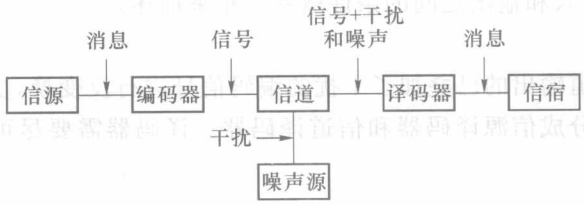

图1.1 通信系统的香农模型

该模型主要分成5个部分：

(1) 信源

顾名思义，信源是产生消息和消息序列的源。信源可以是人、生物、机器或其他事物。比如，各种气象状态是信源，能够产生独特的气味吸引蜜蜂来采花蜜的花朵是信源，人脑的思维活动也是一种信源。信源的输出是消息或消息序列。

消息有着各种不同的形式，例如：文字、符号、语言、图片、图像、气味等。**消息**以能被通信双方所理解的形式，通过通信进行传递和交换。消息携带着**信息**，是信息的载体。信源输出的消息是随机的、不确定的，但又有一定的规律性，因此用随机变量或随机矢量等数学模型来表示信源。

(2) 编码器

编码就是把消息变成适合在信道传输的物理量，这种物理量称为**信号**(如电信号、光信号、声信号、生物信号等)。信号携带着消息，它是消息的载体。

编码器可分为信源编码器和信道编码器。信源编码的目的是压缩信源的冗余度(即多余度)，提高信息传输的效率，这是为了提高通信系统的有效性。信源编码又可分为无失真信源编码和限失真信源编码。信道编码是为了提高信息传输的可靠性而有目的地对信源编码器输出的代码组添加一些监督码元，使之具有纠、检错能力。比如，老师讲课需要把知识进行加工和提炼，以提高信息传输的有效性，而为了让学生听得明白，有时又需要适当地重复，这是为了提高信息传输的可靠性。

在实际的通信系统中，可靠性和有效性常常是相互矛盾的，提高有效性必须去掉信源符号的冗余部分，但是这会导致可靠性的下降，而提高可靠性就需要增加监督码元，这又降低了有效性。有时为了兼顾有效性，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失真，也就是说允许近似地再现原来的消息。

(3) 信道

信道是指通信系统把载荷消息的信号从发送端送到接收端的媒介或通道，是包括收

发设备在内的物理设施。信道除了传播信号以外，还有存储信号的作用。在狭义的通信系统中，实际信道有明线、电缆、光缆、无线电波传播空间、磁盘、光盘等，这些都属于传输电磁波能量的信道。对于广义的通信系统来说，信道还可以是其他的传输媒介。

在信道中引入噪声和干扰，这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，信道输出的已是叠加了干扰的信号。噪声源的统计特性是划分信道的依据，并且是信道传输能力的决定因素。由于干扰或噪声往往具有随机性，所以信道用输入和输出之间的条件概率分布来描述。

(4) 译码器

译码就是把信道输出的已叠加了干扰的编码信号进行反变换，变成信宿能够理解的消息。译码器也可分成信源译码器和信道译码器。译码器需要尽可能准确地再现信源输出的消息。

(5) 信宿

信宿是消息传送的对象，即接收消息的人、机器或其他事物。

研究香农信息论的目的是从数学上对通信系统进行描述和分析，首先研究通信系统的若干性能极限，例如信源的压缩极限，即信源熵、信道的传输极限，也即信道容量；其次是致力于达到这些极限，即提高信息传输的有效性和可靠性，其具体实现方法可以是信源编码和信道编码等。

香农信息论研究的内容是关于通信系统的最根本、最本质的问题。例如：

① 什么是信息？如何度量信息？

② 怎样确定信源的输出中含有多少信息量？

③ 对于一个信道，它传输信息量的最高极限(信道容量)是多少？

④ 为了能够无失真地传输信源信息，对信源编码时所需的最少的码符号数是多少？这是无失真信源编码，即香农第一定理的内容。

⑤ 在有噪信道中有没有可能以接近信道容量的信息传输率传输信息而错误概率几乎为零？这是有噪信道编码，即香农第二定理的内容。

⑥ 如果对信源编码时允许一定量的失真，所需的最少的码符号数又是多少？这是限失真信源编码，即香农第三定理的内容。

1.3 信息论的发展

香农创立信息论至今的60多年来，信息论获得了巨大的发展。首先从香农信息论本身来看，最初的信息论仅考虑了单个信源、信宿的点对点通信模型。随着通信技术的发展，出现了组网通信、广播通信、卫星通信等多种场景，这些场景下同时存在多个信源和信宿并行通信，这就需要把传统的点对点通信模型做适当修正，由此发展出了多用户网络信息理论。

又比如香农提出信道容量后，在最初的几十年里信道编码虽然屡有进展，但是香农所提出的极限却被认为是几乎不可能达到的，随着20世纪末Turbo码的提出以及LDPC码的实用化，信道编码理论顿时成为研究的热点，很多学者提出了宝贵的创造性成果。

信息论除在通信领域获得巨大发展外，还引申到了其他众多领域，包括经济学、语言学、神经学、心理学等，其中在经济学的应用尤为突出。目前信息论的研究范围一般有3种理解。

(1) 狭义信息论：又称香农信息论。主要通过数学描述与定量分析，研究通信系统从信源到信宿的全过程，包括信息的测度、信道容量以及信源和信道编码理论等问题，强调通过编码和译码使收、发两端联合最优化，并且以定理的形式证明极限的存在。这部分内容是信息论的基础理论。

(2) 一般信息论：也称工程信息论。主要也是研究信息传输和处理问题，除香农信息论的内容外，还包括噪声理论、信号滤波和预测、统计检测和估计、调制理论、信息处理理论以及保密理论等。

(3) 广义信息论：也称信息科学，不仅包括上述两方面内容，而且包括所有与信息有关的自然和社会科学领域，如模式识别、机器翻译、心理学、遗传学、神经生理学、语言学、语义学、金融学，甚至包括社会学中有关信息的问题。

本课程将主要研究香农信息论的内容，在不加特别说明时，将香农信息论简称为信息论。

1.4 如何学好信息论

学好信息论，应特别注意以下几个方面：

(1) 打好数学基础。信息论是用数学来研究通信系统的应用科学，无论是定理的证明还是实际问题的计算都需要较好的数学知识，特别是概率论的相关知识。如果读者在数学基础方面比较薄弱，建议先复习该部分知识再学习本课程，本书第2章总结了全书用到的大部分数学知识。

(2) 注重基本概念的理解。例如第3章中提出了信源熵、联合熵、条件熵、互信息等概念，读者应深刻掌握其物理意义和计算方法，基于这些概念的理解，才能学好后面章节中的信源压缩、信道容量等概念。

(3) 注意理论与实际相结合。某种意义上说信息论是一门较为抽象的科学，其定理的证明较为艰深，如果能结合例题对信息论的概念、定理进行理解，将获得事半功倍的效果。

(4) 注意把握相关概念定理的逻辑关系。例如第3章的主线是：事件—概率—信息量—熵—联合熵、条件熵、互信息；在第4章中，这条线还将引申为：互信息—信道容量—信道容量的计算方法。读者在学习中应该注意把相关概念串起来，用图表的方法总结所学知识是学习信息论的一个好方法，切忌孤立地去理解某个概念。

第2章 数学基础回顾

本章要点

概率论和随机过程是信息论的重要数学基础，为方便读者更好地掌握信息论这门课程，本章将信息论中经常用到的概率论和随机过程的相关知识做简单的回顾。本章中条件概率、全概率公式和贝叶斯公式是第一个重点，二维随机变量的联合分布和边缘分布是第二个重点，这两部分知识是学习第3章的重要基础；马尔可夫链是第三个重点，它是学习第4章的重要基础。读者应牢固掌握这三方面内容的基本概念和相关运算，为信息论的学习打好扎实基础。

2.1 离散概率论

2.1.1 随机事件的概率

定义 2.1 一般的，随机试验 E 的每个可能结果被称为**样本点**；所有可能结果组成的集合称为 E 的**样本空间**，记为 S；S 的子集被称为 E 的**随机事件**；由一个样本点组成的单点集称为**基本事件**，由所有样本点组成的集合 S 称为**必然事件**，空集 $\varnothing$ 称为**不可能事件**。事件 $A\cup B$ 称为事件 A 和事件 B 的**和事件**，事件 $A\cap B$ 称为事件 A 和事件 B 的**积事件**，简记为 AB。若 $A\cap B=\varnothing$，则称 A,B 为不相容事件。

定义 2.2 如果对随机试验 E 的每个事件 A 赋予一个实数，记为 $p(A)$，且映射 $p()$ 满足下列条件，则称 $p(A)$ 为事件 A 的**概率**：

Ⅰ. 对于每一个事件 A，恒有 $p(A)\geqslant 0$；

Ⅱ. $p(S)=1$；

Ⅲ. 设 $A_1, A_2, \cdots$ 是两两互不相容的事件，恒有

$$p(A_1 \cup A_2 \cup \cdots) = p(A_1) + p(A_2) + \cdots \tag{2.1}$$

通常，概率 $p(A)$ 用来表示事件 A 发生的可能性的大小。第3章我们将知道，事件的概率决定了其信息量的大小。

可以证明，事件的概率满足下列性质：

Ⅰ. $p(\varnothing)=0$；

Ⅱ. 设 A, B 是两个事件，若 $A \subset B$，则有 $p(B-A)=p(B)-p(A)$ 且 $p(B) \geqslant p(A)$；

Ⅲ. 对于任一事件 A，恒有 $p(A) \leqslant 1$；

Ⅳ. 对于任意两事件 A, B，恒有 $p(A \cup B)=p(A)+p(B)-p(AB)$，特别的，当 A, B 不相容时，$p(A \cup B)=p(A)+p(B)$。

2.1.2 条件概率、全概率公式与贝叶斯公式

1. 条件概率

定义 2.3　设 A, B 是两个事件，且 $p(A)>0$，称

$$p(B|A)=\frac{p(AB)}{p(A)} \tag{2.2}$$

为在事件 A 发生的条件下事件 B 发生的**条件概率**。不难验证，条件概率 $p(\cdot|A)$ 符合概率定义中的3个条件，即

Ⅰ. 对于每一事件 B，有 $p(B|A) \geqslant 0$；

Ⅱ. $p(S|A)=0$；

Ⅲ. 设 $B_1, B_2, \cdots$ 是两两互不相容的事件，恒有

$$p\left(\bigcup_{i=1}^{\infty} B_i \mid A\right) = \sum_{i=1}^{\infty} p(B_i \mid A) \tag{2.3}$$

既然条件概率符合上述3个条件，故2.1.1节中的一些重要结果都适用于条件概率，例如，对于任意事件 B_1, B_2，恒有

$$p(B_1 \cup B_2|A)=p(B_1|A)+p(B_2|A)-p(B_1B_2|A) \tag{2.4}$$

2. 乘法定理

由条件概率的定义变形可得

$$p(AB)=p(B|A)p(A) \tag{2.5}$$

该式称为**乘法定理**，它也可写为

$$p(AB)=p(A|B)p(B) \tag{2.6}$$

式(2.6)容易推广到多个事件的积事件的情况。例如，设 A, B, C 为事件，且 $p(AB)>0$，则有

$$p(ABC)=p(C|AB)p(B|A)p(A) \tag{2.7}$$

更一般的，设 $A_1, A_2, \cdots, A_n$ 为 n 个事件，$n \geqslant 2$，且 $p(A_1A_2\cdots A_{n-1})>0$，则有

$$p(A_1A_2\cdots A_n)=p(A_n|A_1A_2\cdots A_{n-1})p(A_{n-1}|A_1A_2\cdots A_{n-2})\cdots p(A_2|A_1)p(A_1) \tag{2.8}$$

3. 独立性

定义 2.4 设 A,B 是两事件，如果满足等式

$$p(AB)=p(A)p(B) \tag{2.9}$$

则称 A,B 为**相互独立的事件**。

当 A,B 为相互独立的事件时，显然有 $p(B)=p(B|A)$ 且 $p(A)=p(A|B)$。

定义 2.5 设 A,B,C 是三事件，如果满足等式

$$p(AB)=p(A)p(B) \tag{2.10}$$

$$p(BC)=p(B)p(C) \tag{2.11}$$

$$p(AC)=p(A)p(C) \tag{2.12}$$

$$p(ABC)=p(A)p(B)p(C) \tag{2.13}$$

则称 A,B,C 为**相互独立的事件**。

类似的可以得到 n 个事件相互独立的条件。

4. 全概率公式

定义 2.6 设 S 为试验 E 的样本空间，$B_1,B_2,\cdots,B_n$ 为 E 的一组事件，若

Ⅰ. $B_iB_j=\varnothing, i\neq j, i,j=1,2,\cdots,n$；

Ⅱ. $B_1\cup B_2\cup\cdots\cup B_n=S$；

则称 $B_1,B_2,\cdots,B_n$ 为样本空间 S 的一个**划分**。

若 $B_1,B_2,\cdots,B_n$ 是样本空间的一个划分，其直观理解是样本空间 S 被既不遗漏又不重复的分成了 n 份。对每次试验，事件 $B_1,B_2,\cdots,B_n$ 中必有一个且仅有一个发生。

定理 2.1 设试验 E 的样本空间为 S，A 为 E 的事件，$B_1,B_2,\cdots,B_n$ 为 S 的一个划分，且 $p(B_i)>0(i=1,2,\cdots,n)$，则

$$p(A)=p(A|B_1)p(B_1)+p(A|B_2)p(B_2)+\cdots+p(A|B_n)p(B_n) \tag{2.14}$$

式(2.14)称为**全概率公式**。

5. 贝叶斯公式

定理 2.2 设试验 E 的样本空间为 S，A 为 E 的事件，$B_1,B_2,\cdots,B_n$ 为 S 的一个划分，且 $p(A)>0, p(B_i)>0(i=1,2,\cdots,n)$，则

$$p(B_i\mid A)=\frac{p(A\mid B_i)p(B_i)}{\sum_{j=1}^{n}p(A\mid B_j)p(B_j)},\ i=1,2,\cdots,n \tag{2.15}$$

式(2.15)称为**贝叶斯(Bayes)公式**。其中 $p(B_i)$ 称为**先验概率**，$p(B_i|A)$ 称为**后验概率**。

【小结】 条件概率、全概率公式和贝叶斯公式是概率论中十分重要的基础知识，对于信息论也十分重要，第 3 章中，它们将被广泛运用于求解条件熵、联合熵和互信息。读者应注意理解其概念，掌握它们之间的运算关系。

2.1.3 离散型随机变量及其分布

定义 2.7　设 E 是随机试验，它的样本空间是 $S=\{e\}$，如果对于每一个 $e\in S$，都有一个实数 $X(e)$ 与之对应，这样就得到一个定义在 S 上的单值实值函数 $X=X(e)$，称为**随机变量**。有些随机变量，它全部可能取到的值是有限个或可列无限多个，这种随机变量称为**离散型随机变量**。

定义 2.8　设离散型随机变量 X 的所有可能取值为 $x_k(k=1,2,\cdots)$，事件 $\{X=x_k\}$ 的概率为

$$p\{X=x_k\}=p_k,k=1,2,\cdots \tag{2.16}$$

称式(2.16)为离散型随机变量 X 的**概率分布**或**分布律**，分布律也可以用表格的形式来表示，即

X	x_1	x_2	$\cdots$	x_n	$\cdots$
$p(X)$	p_1	p_2	$\cdots$	p_n	$\cdots$

定义 X 的**概率空间**为

$$\begin{bmatrix} X \\ p(X) \end{bmatrix}=\begin{bmatrix} X=x_1 & \cdots & X=x_k & \cdots \\ p(x_1) & \cdots & p(x_k) & \cdots \end{bmatrix} \tag{2.17}$$

由概率的定义，p_k 满足如下两个条件：

Ⅰ. $0\leqslant p_k\leqslant 1,k=1,2,\cdots$；

Ⅱ. $\sum\limits_{k=1}^{\infty} p_k=1$。

下面给出两种典型离散型随机变量的分布律。

1. (0-1)分布

设随机变量 X 只可能取 0 与 1 两个值，它的分布律是

$$p(X=k)=p^k(1-p)^{1-k},k=0,1 \quad (0<p<1) \tag{2.18}$$

则称 X 服从**(0-1)分布**，其分布律也可写成

X	0	1
$p(X)$	$1-p$	p

2. 二项分布与贝努利试验

将试验 E 重复进行 n 次，若各次试验的结果互不影响，即每次试验结果出现的概率都不依赖于其他各次试验的结果，则称这 n 次试验是相互独立的。设试验 E 只有两个可能结果：A 及 $\overline{A}$，$p(A)=p,p(\overline{A})=1-p=q(0<p<1)$。将 E 独立地重复进行 n 次，则称这一串重复的独立试验为 n 重贝努利(Bernoulli)试验，简称**贝努利试验**。用 X 表示这 n