普通高等教育“十一五”电子信息类规划教材

信息论基础教程

焦瑞莉　李红莲　冷俊敏　编著

机械工业出版社
CHINA MACHINE PRESS

普通高等教育“十一五”电子信息类规划教材

信息论基础教程

焦瑞莉　李红莲　冷俊敏　编著
袁保宗　主审

机械工业出版社

本书系统地讲述了信息论的基本理论，共分8章，内容包括：信息的基本概念及信息度量、信源和信息熵、信道与信道容量、信源编码和信道编码定理与常用编码方法、网络信息论以及保密通信的信息理论。对于前6章内容在附录中提供相应内容的MATLAB仿真源程序供教学使用。

本书力求内容精炼、完备、准确，强调掌握信息论的基本理论以及在通信中的指导作用，在不影响内容完整性的前提下省略了部分繁琐的定理证明，可作为高等院校电气信息类专业本科生教材，也可供从事相关专业的科研和工程技术人员参考。

图书在版编目（CIP）数据

信息论基础教程/焦瑞莉，李红莲，冷俊敏编著. —北京：机械工业出版社，2008.7

普通高等教育“十一五”电子信息类规划教材

ISBN 978-7-111-24210-9

Ⅰ.信… Ⅱ.①焦…②李…③冷… Ⅲ.信息论-高等学校-教材 Ⅳ.G201

中国版本图书馆CIP数据核字（2008）第099007号

机械工业出版社（北京市百万庄大街22号 邮政编码100037）

策划编辑：闫晓宇 责任编辑：蔡家伦 责任校对：陈延翔

封面设计：张 静 责任印制：洪汉军

北京汇林印务有限公司印刷

2008年8月第1版第1次印刷

184mm×260mm · 15.25印张 · 374千字

标准书号：ISBN 978-7-111-24210-9

定价：28.00元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

销售服务热线电话：（010）68326294

购书热线电话：（010）88379639 88379641 88379643

编辑热线电话：（010）88379727

封面无防伪标均为盗版

前　言

信息论是长期的通信工程实践与概率论、随机过程和数理统计这些数学学科相结合而逐步发展起来的一门科学。随着信息概念的不断深化，它在科学技术上的重要性早已超越了狭义的通信工程的范畴，在许多领域中日益受到科学工作者的重视，他们竞相应用信息论的概念和方法理解和解决本领域的问题。随着信息技术的飞速发展，信息论课程被列为电气信息类本科生及研究生的必修课，其涉及内容可深可浅，可偏重理论推导也可偏重工程应用，有着不同的取舍和组织方式。

本书是编者总结多年为本科生讲授信息论课程的教学经验，在使用多年的教学讲义的基础上编写而成的。此次编写补充了马尔可夫信源和常用信道编码方法，增加了网络信息论和保密通信的信息理论，以使内容更加完整。本书突出信息论基础理论和信息与通信专业的结合，强调掌握信息论的基本理论及其对通信的指导作用。本书力求文字简洁，概念清晰，并选取了比较有代表性的习题供读者练习，以便对内容加强理解与掌握。因本书定位为本科生教材，因此在不影响内容完整性的前提下，省略了一部分繁琐的定理证明，以适合本科教学使用。在内容组织上有意提高了可读性，如定义、定理、例题等使用突出格式方便读者阅读。精炼、完备、准确是编者努力的目标，希望阅读本书能使读者在较短的时间内，对信息论的基本内容和方法有较为准确而深刻的理解与掌握。

全书共8章，系统地讲述了信息论的基本内容。第1章详细地描述了信息的概念和通信系统模型，为信息的度量和后续的信息论基本问题的讨论打下基础。第2、3章分别讨论信源的信息度量——熵和信道的信息传输特性——信道容量的特性与计算问题。第4、6章分别讨论信源的无失真编码和限失真编码问题。第5章论述信道编码定理和常用的几种信道编码方法。第1~6章（除5.4节讲述常用信道编码方法外）为香农信息论基础理论。第7章简要介绍了网络信息理论的一些基本内容。第8章主要介绍信息论在保密学中的应用，论述了数据加密标准（DES）和公开密钥密码的原理及其实现，并对信息安全与数字签名作了介绍。另外，在书后附录中提供了前6章相应内容的MATLAB仿真源程序，可依据教学需要选择使用。既可作为课后作业，又可作为单独的实验题目。

本书可作为高等院校电气信息类专业本科生教材，也可供从事相关专业的科研和工程技术人员参考。使用本书作为教材时，授课教师可依据实际情况对书中内容进行取舍。

本书第1~6章由焦瑞莉编写，其中2.7节和5.4节内容由李红莲编写，第7章由李红莲编写，第8章及附录中MATLAB仿真源程序由冷俊敏编写，全书由焦瑞莉统稿。

感谢北京交通大学袁保宗教授在百忙之中对本书进行了认真的审阅，并提出了十分宝贵的修改意见和建议。还要感谢北京信息科技大学光电信息与通信工程学院对本书编写的支持。

由于编者水平所限，书中难免有漏误不当之处，敬请广大读者批评指正。

编　者

2008年5月于北京信息科技大学

目　录

前言

第 1 章　绪论 …… 1

1.1　信息的概念 …… 1

1.2　信息论研究的内容 …… 2

1.3　通信系统模型 …… 4

1.4　信息论发展简史和现状 …… 5

第 2 章　信源和熵 …… 7

2.1　信源特性和分类 …… 7

2.2　离散信源的熵 …… 10

2.2.1　信息量的定义 …… 11

2.2.2　熵 …… 14

2.2.3　条件熵和联合熵 …… 16

2.3　熵函数的数学特性 …… 18

2.4　离散随机变量之间的互信息 …… 24

2.4.1　互信息量 …… 24

2.4.2　条件互信息量和联合互信息量 …… 26

2.4.3　平均互信息量 …… 29

2.5　信息不增性原理 …… 37

2.6　平稳离散信源 …… 39

2.6.1　平稳信源的定义 …… 39

2.6.2　平稳信源的熵 …… 41

2.6.3　信源的冗余度 …… 44

2.7　马尔可夫信源 …… 46

2.7.1　马尔可夫信源的定义 …… 46

2.7.2　马尔可夫信源的熵 …… 47

2.8　连续随机变量的熵和互信息 …… 49

2.8.1　连续随机变量的相对熵和绝对熵 …… 50

2.8.2　最大相对熵 …… 53

2.8.3　熵功率 …… 57

习题 …… 59

第 3 章　信道与信道容量 …… 63

3.1　信道的数学模型和分类 …… 63

3.2　离散无记忆信道的信道容量 …… 67

3.2.1　信道容量的定义 …… 67

3.2.2　离散无噪声信道 …… 71

3.2.3　准对称与对称离散无记忆信道容量 …… 73

3.2.4　可逆矩阵信道的信道容量 …… 76

3.3　信源与信道的匹配 …… 78

3.4　信道的组合 …… 78

3.4.1　积信道（独立并行信道） …… 79

3.4.2　和信道（并信道） …… 81

3.4.3　输入并接信道 …… 82

3.4.4　级联信道 …… 83

3.5　时间离散的无记忆连续信道 …… 83

3.5.1　可加噪声信道 …… 84

3.5.2　平均功率受限可加噪声信道 …… 85

3.5.3　香农公式 …… 87

3.5.4　平行可加高斯信道的容量 …… 89

习题 …… 91

第 4 章　离散信源的无失真编码 …… 94

4.1　编码器 …… 94

4.2　等长码和等长信源编码定理 …… 95

4.3　变长码 …… 98

4.4　变长信源编码定理 …… 103

4.5　变长码的编码方法 …… 109

4.5.1　费诺（Fano）码 …… 109

4.5.2　霍夫曼（Huffman）码 …… 110

4.5.3　r 进制霍夫曼码 …… 114

习题 …… 116

第 5 章　信道编码 …… 118

5.1　错误概率和译码规则 …… 118

5.2　错误概率与编码方法 …… 121

5.3　信道编码定理与逆定理 …… 127

5.3.1　信道编码定理 …… 127

5.3.2　信道编码定理的逆定理 …… 127

5.4　常用信道编码方法 …… 128

5.4.1　检错和纠错的基本原理 …… 129

5.4.2　奇偶校验码 …… 129

5.4.3　线性分组码 …… 130

5.4.4　汉明码 …… 134

5.4.5　循环码 …… 135

5.4.6　卷积码 …… 138

习题 …… 141
第 6 章 限失真信源编码 …… 145
6.1 引言 …… 145
6.2 率失真函数的定义 …… 146
6.2.1 失真函数 …… 146
6.2.2 率失真函数的定义 …… 148
6.3 率失真函数的性质 …… 150
6.4 率失真函数的计算 …… 153
6.5 连续信源的率失真函数 …… 156
6.5.1 连续信源的率失真函数及其计算 …… 156
6.5.2 高斯信源的率失真函数及其计算 …… 157
6.6 限失真信源编码定理 …… 160
6.6.1 信源编码定理及其逆定理 …… 160
6.6.2 编码定理的意义 …… 160
习题 …… 161
第 7 章 网络信息论 …… 164
7.1 网络信道分类 …… 164
7.1.1 多源接入信道 …… 164
7.1.2 广播信道 …… 164
7.1.3 中继信道 …… 165
7.1.4 串扰信道 …… 166
7.1.5 双向信道 …… 167
7.1.6 反馈信道 …… 167
7.1.7 多用户通信网信道 …… 167
7.2 相关信源编码 …… 168
7.2.1 基本概念 …… 168
7.2.2 相关信源独立编码 …… 170
7.2.3 相关信源协同编码 …… 171
7.3 典型网络信道 …… 172
7.3.1 多源接入信道 …… 172
7.3.2 高斯多源接入信道 …… 174
7.3.3 中继信道 …… 176
7.3.4 广播信道 …… 177
7.3.5 反馈信道 …… 178
习题 …… 180
第 8 章 保密通信的信息理论 …… 181
8.1 保密通信基础知识 …… 181
8.1.1 保密学的发展史 …… 181
8.1.2 基本概念 …… 182
8.2 保密系统的数学模型 …… 184
8.2.1 保密通信系统 …… 184
8.2.2 密码学中熵的概念 …… 186
8.2.3 理想保密性 …… 187
8.3 数据加密标准（DES） …… 188
8.3.1 替代密码与置换密码 …… 188
8.3.2 DES 密码算法 …… 190
8.3.3 DES 密码的安全性 …… 195
8.4 国际数据加密算法（IDEA） …… 195
8.4.1 算法原理 …… 195
8.4.2 加密解密过程 …… 196
8.4.3 算法的安全性 …… 197
8.5 公钥加密方法 …… 198
8.5.1 公钥密码体制的基本原理 …… 199
8.5.2 RSA 密码体制 …… 199
8.5.3 报文摘要 MD5 …… 202
8.6 信息安全与数字签名 …… 206
8.6.1 信息安全的基本概念 …… 206
8.6.2 数字签名 …… 208
习题 …… 211
附录 …… 212
附录 A 信道编码定理的证明 …… 212
附录 B MATLAB 源程序 …… 215
参考文献 …… 237

第1章 绪 论

信息论是长期的通信工程实践与概率论、随机过程和数理统计这些数学学科相结合而逐步发展起来的一门科学。随着信息概念的不断深化，它在科学技术上的重要性早已超越了狭义的通信工程的范畴，在许多领域中日益受到科学工作者们的重视。

本章首先引出信息的概念，进而讨论信息论研究的内容、历史发展与现状，最后介绍了本教材后文的内容结构。

1.1 信息的概念

在信息论中，信息是最基本、最重要的概念。那什么是信息呢？

信息概念的定义很多，但直到现在，还没形成完整、明确、为世人所普遍公认的定义。信息是客观事物状态和特征的反映，具有形式和内容之分。不同的定义从不同的侧面、不同的层次上揭示了信息的特性。通常有三类定义：广义信息、技术术语信息和统计信息。

广义信息是将信息的形式和内容等全部包含在内的最广泛意义上的信息。信息是在人类社会互通情报的实践过程中产生的，人们总能感受到信息的存在，常把“消息”、“情况”、“情报”等认为就是信息，是形式和内容的统一。当人们接到一个电话，或从收音机里听到气象预报，或看了电视里的新闻之后，就从“消息”中获得了信息。这种对信息的理解，我们可以把它称为广义信息。

技术术语信息是计算机所处理的海量对象，如音频和视频数据、文档资料等，在技术层面上统称为信息。信息作为技术术语被广泛使用是在计算机，特别是微处理器得到广泛应用以后，其特点是把信息的形式或载体与具体包含的内容分离开来。虽然信息形式和内容之间存在着联系，但计算机处理的技术角度只关心信息的形式或载体，不考虑信息的内容。

统计信息是统计意义上的信息，可用数学公式严格定义，反映了信息表达形式中统计方面的性质，是统计学上的抽象概念。统计信息是有明确定义的科学名词，它与内容无关，而且不随信息的具体表达式的变化而变化，也独立于形式。**香农[㊀]信息论中定义的信息是“事物运动状态或存在方式的不确定性的描述”，即遵循了统计信息的定义。**信息论中讨论的信息就是统计信息，因此我们在此后提及的、讨论的信息都是统计信息。

信息本身看不见、摸不着，因此在信息获取、传输、存储和处理时，总要借助一定的载体或形式——不同形式的消息，如文字、图像、语言、数据等，而消息最终又是由某种物理信号形式表达、传输和存储的，如声、光、电、磁等信号。因而，信息、消息和信号三者之间存在着必然的联系，但又相互区别，如图 1-1 所示。

㊀ 克劳德·香农（Claude Elwood Shannon，1916—2001），信息论与数字通信时代的奠基人。

信息这一抽象概念是包含在消息之中的，是通信系统传递的对象，是通信的价值所在，没有信息就没有信息传递的意义。**信息是无形的、可共享的**。同时，信息是**无限的**，这是由于信息是事物运动状态或存在方式的不确定性描述，因此和事物及其运动一样是永恒和无限的。信息的无限性还表现在时空扩展性上，如今天的气象数据所包含的信息，对明天来讲由于不确定性的消除而失去价值，但将它存档积累起来就变成重要的气象数据资料，又成为研究气候演变的有用信息。信息因其统计意义上的科学定义又是**可度量的**。显然，信息的大小与不确定性的程度有关。用数学的语言来讲，不确定性就是随机性。因此，可通过数学工具——概率论与随机过程来度量信息的大小。具体如何度量信息的大小将在后面章节详细讨论。

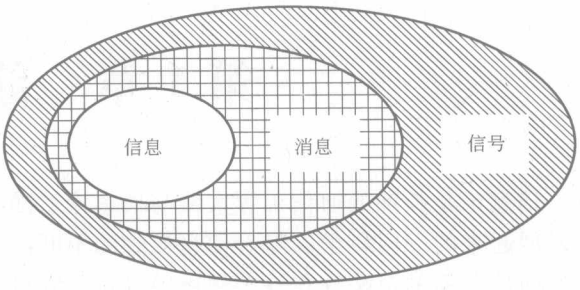

图 1-1 信息、消息和信号三者的关系

消息是信息的载体。消息具有不同的形式：语言、文字、符号、数据、图片等。构成消息的两个条件是能被通信双方所理解和可以在通信中传递和交换。从电报、电话或电视中得到的是一些具体的消息，是描述主、客观各种事物运动状态的消息。而电报、电话、电视等则都是这些消息的传递系统。需要指出的是，在消息传递系统中实质上传输的是信息，而消息只是表达信息的工具。同一消息对不同的客体来说可以含有不同的信息量，同一信息可以用不同形式的消息来承载。

信号是表示消息的物理量，包括声波、光波、电信号、机械信号等。信号是消息的表现形式，消息则是信号的具体内容。信号相对于信息和消息是最具体的，它是可测量、可显示、可描述的物理量，是承载信息的最终实体。

目前，信息已被人们看做是物质世界的基本属性之一。现代哲学家和科学家认为：物质、能量、信息是物质世界的三大要素，是科学历史上三个最重要的基本概念。而这三者之间有着密切的内在联系。物质运动的动力是能量，而信息是关于物质运动的状态的知识。只要有运动的事物，就需要有能量，也就会存在信息，可见信息是普遍存在的。因此，可以说人类在改造世界和认识世界的一切活动中无不牵涉到信息的交换和利用。信息的概念已远远超出原来通信的范畴。可以预言，随着人们对信息这一概念的不断深化，信息论和信息科学将会揭示出客观世界和人类主观世界的更多的内在规律，从而使人们有可能创造出更多更好的信息系统——信息获取系统、信息传输系统、信息控制系统以及信息处理系统。

1.2 信息论研究的内容

目前，对信息论研究的内容一般有以下三种理解。

（1）狭义信息论　狭义信息论是信息论基础理论，主要研究信息的测度、信道容量以及信源和信道编码理论等问题，也就是香农信息论。

（2）一般信息论　一般信息论主要是研究信息的传输和处理问题。但除了香农理论以外，还包括噪声理论、信号滤波和预测、统计检测与估计理论、调制理论以及信息处理理

论等。

后一部分内容以美国科学家维纳（N. Wiener）为代表，其中最有贡献的是维纳和前苏联科学家柯尔莫哥洛夫。

虽然维纳和香农等人都是运用概率和统计数学的方法来研究准确地或近似地再现消息的问题，都是为了使消息传输和接收最优化，但他们之间却有一个重要的区别。维纳研究的重点是在接收端。研究信号（消息）如果在传输过程中被某些因素（如噪声、非线性失真等）所干扰时，接收端如何把它恢复、再现，从干扰中提取出来。在此基础上，创立了最佳线性过滤理论（维纳滤波器）、统计检测与估计理论、噪声理论等。而香农研究的对象则是从信源到信宿之间的全过程，是收、发端联合最优化问题，其重点是编码。他指出，只要在传输前后对消息进行适当的编码和译码，就能保证在干扰存在的情况下，最佳地传送和准确或近似地再现消息。为此发展了信息测度理论、信道容量理论和编码理论等。

（3）广义信息论　广义信息论不仅包括上述两方面的内容，而且包括所有与信息有关的领域，如心理学、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。

本教材遵循第一种理解，把其他所有与信息有关的领域都看成是信息论的应用。

综上所述，信息论是一门应用概率论、随机过程、数理统计和近代代数等方法研究信息提取、传输和处理的科学。它的主要目的是提高信息系统的有效性和可靠性，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、调制理论和随机噪声理论等。

由于信息论研究的内容极为广泛，而各分支又有一定的相对独立性，因此本教材仅论述信息论的基础理论即香农信息论。

香农信息论又称为“通信的数学理论”。所谓通信是指消息的传递。通信的目的在于使接收端能够准确地或者在允许失真限度内重现发送的消息。香农信息论正是研究通信系统（即信息传输系统）中普遍存在的信息传输的共同规律性问题，研究信源、信宿和信道的统计特性以及有关编码问题，为设计有效而可靠的通信系统提供理论依据。这些信息论的基本问题，在香农的早期著作中都已系统地提出并给出了启发式证明。香农信息论的最大特点是将概率统计的观点和方法引入通信理论的研究之中，揭示了通信系统中传输的对象是信息，并对信息给出了科学的、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信源编码和信道编码，并且从理论上证明了可以逼近的最佳性能的极限。

信源编码的作用是根据失真度准则对信源的输出消息进行编码，用码字表示消息。信源译码的任务是根据收到的消息恢复出信源的原始输出。显然，失真度要求越小，对信源编码的码长也就越大。在给定信源和失真度的条件下，要多大信息速率才行呢？或者，对给定信源和保留一定信息速率的条件下，可以达到的最小失真是多少呢？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。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，也是一个十分重要的问题。

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，但讨论它不是为了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信息传输能力。当送入信道的信息速率为 R，信道容量为 C 时，则根据信道编码基本定理得知，若 $R<C$，则可以将速率为 R 的信息以任意高的可靠性送至接

收端；若 $R > C$，则不可能。这是信道编码和“可行性”问题。信道编码也存在寻找实际可行的编码和译码方法问题。

总之，香农信息论的中心问题是如何提高信息传输系统的有效性和可靠性。香农信息论已建立了自己完整的理论体系，它是一般信息论和广义信息论的理论基础。对香农信息论进行讨论，无疑是有助于将信息论渗透到更广泛的各种科学技术领域之中，有助于进一步发展和深化信息概念和信息理论的。

1.3 通信系统模型

信息论研究的主要问题是如何提高信息传输系统的有效性和可靠性。为了便于描述、分析和研究信息传输系统，我们首先把信息传输系统抽象成如图 1-2 所示的通信系统模型。这个模型主要分成 5 个部分。

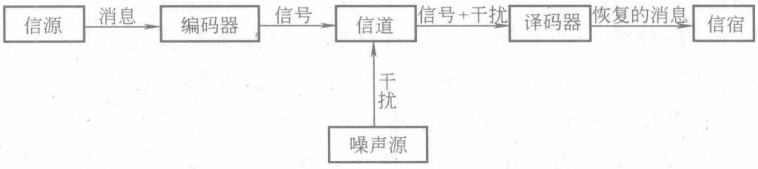

图 1-2 通信系统模型

（1）信源 顾名思义，信源是产生消息和消息序列的源，是同一类消息的集合。“母亲的身体情况”、“各种气象状态”等客观存在是信源，人的大脑思维活动也是一种信源。信源的输出是消息，消息是具体的，但它不是信息本身。消息携带着信息，消息是信息的载体。

（2）编码器 在许多场合，为了克服时间或空间的限制而进行通信，必须对消息进行加工处理，把消息变换成适合信道传输的物理现象。这种物理现象称为信号（如电信号、光信号、声信号等）。所以编码是把消息变换成信号的措施，而译码就是编码的反变换。编码器输出的是适合信道传输的信号，信号携带着消息，它是消息的负荷者。

编码器可分为两种，即信源编码器和信道编码器。为了提高信息传输的有效性和可靠性，它们分别对原始消息进行适当的处理。信源编码器的目的在于提高信息传输的有效性，是在一定的保真度准则下对信源输出进行变换，而信道编码是对信源编码器的输出进行变换，用以提高信息传输的抗干扰能力。当然，对于各种实际的通信系统，编码器还应包括换能、调制、发射等各种变换处理。

（3）信道 信道是指通信系统把承载消息的信号从甲地传输到乙地的媒质。在狭义的通信系统中实际信道有电缆、光纤、波导、无线电波传播空间等，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质。

（4）译码器 译码就是把信道输出的信号（已迭加了干扰）进行编码的反变换。一般认为这种变换是可逆的。译码器也可分成信源译码器和信道译码器。

（5）信宿 信宿是消息的接收者，可以是接收消息的人或机器。

此外，在通信系统模型中我们引入了噪声和干扰。只在信道中引入噪声和干扰，这是一种简化的表达方式，也就是把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，它将作用于信道中传输的信号上。

研究这样一个通信系统的目的就是要使设计出来的各种通信系统具有高有效性和高可靠性。

所谓有效性，即用尽可能少的代价（时间、带宽）来传送一定数量的信息。而可靠性，就是要使信源发出的消息经过信道传输以后，尽可能准确而不失真地在接收端再现，或者在不超过限定失真的前提下再现信源发出的消息。两者的结合就能使系统达到最优化。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。

香农在1948年发表的文章的序言中有这样一句话："通信的基本问题是要在某一端准确地或近似地再现从另一端选择出来的消息"。这句话恰如其分地表达了信息论研究的目的——提高通信系统的有效性和可靠性。

1.4 信息论发展简史和现状

信息论从诞生到今天，已有50多年历史，现已成为一门独立的科学理论。

信息论是在长期的通信工程实践和理论研究的基础上发展起来的。电的通信系统（电信系统）已有170年的历史了，从1832年的莫尔斯电报到20世纪60年代的光纤通信。随着通信工程技术的发展，有关理论问题的研究也逐步深入，这对信息论的形成与发展起到了积极的推动作用。

1832年莫尔斯电报系统中高效率编码方法对后来香农的编码理论是有启发的。

1885年凯尔文（L. Kelvin）曾经研究过一条电缆的极限传信率问题。

1922年卡逊（J. R. Carson）对调幅信号的频谱结构进行了研究，并建立了信号频谱概念。

1924年奈奎斯特（H. Nyquist）指出，如果以一个确定的速度来传输电报信号，就需要一定的带宽。他把信息率与带宽联系起来了。

1928年哈特莱（R. V. Hartley）发展了奈奎斯特的工作，并提出把消息考虑为代码或单语的序列。在 s 个代码中选 N 个码即构成 s^N 个可能的消息。他提出"定义信息量 $H = N\log s$"，即定义信息量等于可能消息数的对数。其缺点是没有统计特性的概念。他的工作对后来香农的思想是有影响的。

1936年阿姆斯特朗（E. H. Armstrong）认识到在传输过程中采用增加带宽的办法对抑制噪声干扰有明显的效果。根据这一思想他提出了宽偏移的频率调制方法，该方法具有划时代的意义。

1939年达德利（H. Dudley）发明了声码器。当时他提出的概念是：通信所需要的带宽至少要同所要传送消息的带宽一样。达德利和莫尔斯都是信源编码的先驱者。

20世纪40年代初期，由于军事上的需要，维纳在研究防空火炮的控制问题时，发表了《平稳时间序列的外推、内插与平滑及其工程应用》的论文。他把随机过程和数理统计的观点引入通信和控制系统中来，揭示了信息传输和处理过程的统计本质。他还利用他本人早在20世纪30年代初提出的"广义谐波分析理论"对信息系统中的随机过程进行谱分析。这就

使通信系统的理论研究面貌焕然一新，引起了质的飞跃。

1948 年香农在《贝尔系统技术》杂志上发表了两篇有关“通信的数学理论”的文章。在这两篇论文中，他用概率测度和数理统计的方法系统地讨论了通信的基本问题，得出了几个重要而带有普遍意义的结论，并由此奠定了现代信息论的基础。

香农理论的核心是：在有噪信道中只要信息传输速率低于某个值——信道容量，则就存在着一种编码方法，它可以使消息传输过程的差错概率任意小。这个结论完全出乎人们的意料，引起了科学家们的巨大兴趣。当然，从数学观点看，香农的结论是一个最优编码的存在定理。从工程观点看，尽管香农没有提出实现最优编码的具体途径，但存在性的严格证明促进了人们去寻找通往有效通信系统的途径。

从 20 世纪 50 年代开始，一部分科学家（包括香农本人）又把已经得到的数学结论作了进一步的严格论证和推广。在这方面，1954 年范恩斯坦（A. Foinstein）的论著是有很大贡献的。香农本人在 50 年代发表了许多重要文章，1959 年他发表了“保真度准则下的离散信源编码定理”，以后发展成为“信息率失真理论”。这一理论是频带压缩、数据压缩的理论基础，它一直到今天还保持着继续发展的势头。

与此同时，另外一部分科学家在从事寻找最佳信道编码（纠错码）的研究工作。这一工作取得了很大的进展，并已经形成一门独立的分支——纠错码理论。

1961 年香农发表论文《双路通信信道》，开拓了网络信息论的研究。1970 年以来，随着卫星通信、计算机网络通信的快速发展，网络信息论的研究十分活跃，发表了大量的论文，网络信息论日趋完善，已成为信息论的重要研究课题之一。

关于保密理论问题，香农在 1949 年发表的论文《保密通信的信息理论》中首次用信息论的观点对信息保密问题进行论述。到 1976 年狄非（Diffie）和赫尔曼（Hellman）提出公开密钥（简称“公钥”）密码体制后，保密通信问题才得到广泛研究。现已经形成独树一帜的分支——密码学理论。

在香农信息论方面，目前值得注意的研究动向是：信息概念的深化、网络信息理论和多重相关信源编码理论的发展与应用、通信网的一般信息理论研究、磁记录信道的研究、信息率失真理论的发展及其在数据压缩和图像处理中的应用、信息论在大规模集成电路中的应用等。这些研究方向都是与当前信息工程的前景——光通信、空间通信、计算机网络、语音和图像的信息处理等紧密相关的。

当前，信息论及方法不仅在电子学领域应用，还广泛地应用到生物学、医学、生理学、语言学、社会学和经济学等领域，这得益于信息科学的兴起与发展。信息论是信息科学不可缺少的一块基石。信息科学是以信息为主要研究对象，以信息的运动规律和应用方法为主要研究内容，以计算机等技术为主要研究工具，以扩展人类的信息功能为主要目标的一门新兴的综合性学科。信息科学由信息论、控制论、计算机科学、仿生学、系统工程与人工智能等学科互相渗透、互相结合而形成。信息过程普遍存在于生物、社会、工业、农业、国防、科学实验、日常生活和人类思维等各种领域，因此信息科学将对工程技术、社会经济和人类生活等方面产生巨大的影响。

当今社会已步入高度信息化时代，信息、能源和材料是现代科学技术的三大支柱。毋庸置疑，信息论和信息科学的不断发展，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，将对现代科学技术的发展产生更为深远的影响。

第2章 信源和熵

从本章开始，将从如何有效而可靠地传输信息的视角来讨论信息传输系统（即通信系统）的各个组成部分。本章首先讨论信源。对于信源需要解决两个问题：一是如何描述信源，即如何计算各类信源信息量——熵；二是如何表示信源的输出，即信源编码问题。这两个问题都与信宿对通信质量要求有关。本章仅讨论信源的第一个问题，即各类信源的统计特性及其信息度量——熵，这些是香农信息论的基础。信源的第二个问题将在第4章和第6章讨论。

2.1 信源特性和分类

信源是信息的来源，可以是人、生物、机器或其他事物。信息是一个抽象的概念，而信源通常是以消息（或符号）的形式发出信息。作为信息论研究的问题而言，不是信源的内部结构及其特性，也不是消息产生的过程，而只是信源的输出。信源发出的消息（符号）对于其接收者来说存在不确定性（随机性），因此可以用随机变量或随机序列来描述信源输出的消息，用概率空间来描述信源。从随机变量出发来研究信息是香农信息论的基本出发点。

描述信源输出消息的随机变量，可以在某一离散集合内取值，也可以在某一连续区间内取值，相应的信源就分别称为离散信源和连续信源。

如果信源输出的消息是以一个个离散符号的形式出现，例如文字、字母、数字等，而且这些符号的取值都是有限的或可数的，则这种信源就属于**离散信源**。离散信源的数学模型是离散型概率空间，见式（2-1）。

$$\begin{pmatrix} X \\ p(x) \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{pmatrix} \tag{2-1}$$

式中，$x \in \{x_1, x_2, \cdots, x_n\}$为离散随机变量；$n \in I$，$I$为正整数集。

$p(x_i) = p(x = x_i)$ 为元素 $x_i (i = 1, 2, \cdots, n)$ 的概率，且应满足

$$\sum_{i=1}^{n} p(x_i) = 1 \tag{2-2}$$

式（2-2）表示信源可能取的消息（符号）只有 x_1、x_2、$\cdots$、x_n这 n 个，并且每次必定也只能取其中的一个。或者说，随机变量 x 的取值集合 X 是一个完备的、两两不相容的基本随机事件集。

如果信源输出消息的取值是连续的，即可能出现的消息数是不可数的无限值，例如语音、电视图像、遥感器测得的连续数据等，这样的信源就属于**连续信源**。连续信源的数学模型是连续型概率空间，见式（2-3）。

$$\begin{pmatrix} X \\ \omega(x) \end{pmatrix} = \begin{pmatrix} (a,b) \\ \omega(x) \end{pmatrix} \quad 或 \quad \begin{pmatrix} X \\ \omega(x) \end{pmatrix} = \begin{pmatrix} R \\ \omega(x) \end{pmatrix} \tag{2-3}$$

并应满足

$$\int_a^b \omega(x)\mathrm{d}x = 1 \quad 或 \quad \int_R \omega(x)\mathrm{d}x = 1 \tag{2-4}$$

式中，(a, b) 为连续随机变量 x 的取值区间；R 为整个实数集 $(-\infty, +\infty)$。$\omega(x)$ 是 x 的概率密度函数。

上面所说的离散信源和连续信源都只输出一个消息（符号），可以称为单符号信源。它是用一维离散或连续随机变量来描述的。这是信源最简单的情况。

实际信源输出的消息往往不只是一个符号，而是一系列符号，这时可以称为符号序列信源。例如，电报信号是在时间上离散的符号（脉冲）序列，而书信则是在空间上离散的符号（文字）序列。又如，语音是时间的连续函数 $X(t)$，静止的平面图像是空间的连续函数 $X(x, y)$，活动的立体图像则是空间和时间的连续函数 $X(x, y, z, t)$。这样，信源输出的消息在时间或空间上又可以分为离散的和连续的两种。这种多符号序列信源输出的消息在时间或空间任一点上每个符号的出现都是随机的，其取值也可以是离散或连续随机变量。

如果把信源输出的消息看做为时间或空间上离散的一系列随机变量，这样信源的输出可用如下 N 维随机变量来描述：

$$\boldsymbol{X} = (X_1, X_2, \cdots, X_l, \cdots, X_N)$$

式中，X_l为信源随机发出第 l 个消息符号；N 为随机消息（符号）序列 X 的长度，可为有限的正整数或可数的无限值。

在这 N 维随机变量 $\boldsymbol{X} = (X_1, X_2, \cdots, X_l, \cdots, X_N)$ 中，若每个随机变量 $X_l (l = 1, 2, \cdots, N)$ 都取值于同一离散集合 $X = \{x_1, x_2, \cdots, x_q\}$（其中 q 为离散集合 X 的元素个数），即

$$X_l \in X = \{x_1, x_2, \cdots, x_q\}, \quad l = 1, 2, \cdots, N$$

则 N 维随机变量

$$\boldsymbol{X} = (X_1, X_2, \cdots, X_N) \in X^N$$

共有 q^N种取值，这种信源输出符号序列的统计特性应该用 N 维随机变量 $\boldsymbol{X}$ 的联合概率分布来表示：

$$\begin{aligned} p(\boldsymbol{x}) &= p(\boldsymbol{X} = \boldsymbol{x}) \\ &= p(X_1 = x_1, \cdots, X_l = x_l, \cdots, X_N = x_N) \\ &= p(x_1, \cdots, x_l, \cdots, x_N) \end{aligned} \tag{2-5}$$

式中，$x_l \in X = \{x_1, x_2, \cdots, x_q\}$（$l = 1, 2, \cdots, N$）表示信源发出的第 l 个消息符号的取值，有 q 种取法。

应该注意，不同的消息 X_l，其取值 x_l可以相同，也可以不同。可见，这类多符号序列信源的数学模型是一个 N 维的离散概率空间，见式（2-6）。

$$\begin{pmatrix} \boldsymbol{X} \\ p(\boldsymbol{x}) \end{pmatrix} = \begin{pmatrix} (x_1, x_2, \cdots, x_q) & (x_2, x_1, \cdots, x_1) & \cdots & (x_q, x_q, \cdots, x_q) \\ p(x_1, x_2, \cdots, x_q) & p(x_2, x_1, \cdots, x_1) & \cdots & p(x_q, x_q, \cdots, x_q) \end{pmatrix} \tag{2-6}$$

该概率空间中共有 q^N个元素。

当信源发出的符号序列即 N 维随机变量 $\boldsymbol{X}$ 的各个分量 X_l相互之间是统计独立的，并具有同样的概率分布时，则 N 维随机变量 $\boldsymbol{X}$ 的概率分布即式（2-5）将满足

$$p(\boldsymbol{x}) = \prod_{l=1}^{N} p(X_l = x_l) = \prod_{l=1}^{N} p(x_l) \tag{2-7}$$

式中，$x_l \in X = \{x_1, x_2, \cdots, x_n\}$，$l = 1$、2、$\cdots$、$N$。

此时，N 维随机变量的联合概率分布等于它的 N 个分量概率分布的乘积。这种 N 维随机变量所表示的信源称为**无记忆离散信源**。

在 N 维随机变量 $\boldsymbol{X} = (X_1, X_2, \cdots, X_N)$ 中，若每个随机变量 $X_l (l = 1, 2, \cdots, N)$ 都是连续随机变量，则 $\boldsymbol{X}$ 是连续型 N 维随机变量。这种信源输出消息的统计特性，可用如下 N 维联合概率密度函数来表示：

$$\omega_N(\boldsymbol{x}) = \omega_N(X_1, X_2, \cdots, X_N) \tag{2-8}$$

式中，$X_l \in (a, b)$，$l = 1, 2, \cdots, N$，表示信源发出的第 l 个消息符号，并在实数区间内的取值。当其各个连续随机变量 X_l 彼此之间相互独立时，则 $\boldsymbol{X}$ 的联合概率密度函数等于各个随机变量概率密度函数的乘积，即式（2-8）将满足

$$\omega_N(\boldsymbol{x}) = \prod_{l=1}^{N} \omega_l(X_l) \tag{2-9}$$

式中，$\omega_l(X_l)$ 为信源发出的第 l 个随机变量的概率密度函数。这种 N 维连续型随机变量所表示的信源称为**无记忆连续信源**。

如果 N 维随机变量的各个随机变量不是相互独立的，则其联合概率分布或联合概率密度函数中就必然要引入条件概率或条件概率密度，用以说明各分量之间的关联性，与此相应的信源就称为**有记忆信源**。一般情况下，实际信源常常是有记忆的。对于有限记忆信源，不能用式（2-7）或式（2-9）来描述其输出的统计特性。当 N 很大（甚至无限大）时，表述有记忆信源要比表述无记忆信源困难得多。在实际问题中，信源发出的符号往往只与前面若干个符号的依赖性较强，而与更前面的符号依赖关系就很弱。因此往往限制记忆长度，并由此引出有限记忆信源和无限记忆信源的区分。

在 N 维随机变量 $\boldsymbol{X}$ 中的分量 X_l 只与前面有限个随机变量有关，则称为**有限记忆信源**，否则称为**无限记忆信源**。对于有限记忆信源可用条件概率分布

$$p(X_l \mid X_{l-1}, X_{l-2}, \cdots, X_{l-m})$$

或条件概率密度函数

$$\omega(X_l \mid X_{l-1}, X_{l-2}, \cdots, X_{l-m})$$

来描述信源的统计特性，其中 m 为正整数，称为记忆阶数或记忆长度。相应信源称作 ***m* 阶记忆信源**。

如果信源的输出是时间或空间上的连续函数，而且它的取值又是连续的和随机的，这种信源的输出要用随机过程来描述，可称为**波形信源**。一般分析随机过程比较困难。但根据抽样定理，只要是一个限频或限时过程，就可把随机过程用一系列取样值来表示，而每个取样值都是连续随机变量。这样，就可以把随机过程转换成 N 维随机变量来处理。对于一个限频 F、限时 T 的随机过程 $X(t)$，即 $X(t)$ 的上限频率是 F，而 $t \in (a, b)$ 或 $T = b - a$，可用 $N = 2FT$ 维取样来表示，或者说随机过程 $X(t)$ 有 $N = 2FT$ 个自由度。应该指出，任何一个函数是限频的时候，则这个函数的时间延伸必定是无限的；反之，真正的限时函数不可能是限频的，即

$$X(t) = 0 \quad t \notin (a, b)$$

上述的限频、限时过程是建立在一定条件之下的，限时是研究这一段时间之内的随机过程，在这段时间之外它并不恒等于零，而是令它与（a，b）区间内一样地周期重复着，这种周期性函数可能是限频的，但在（a，b）区间之外就不一定是实际的随机过程了。反之，在限频的情况下，只取（a，b）区间内的样本来表示，而在（a，b）区间之外就不确定了。

一般情况下，限频 F、限时 T 的随机过程，取样得到的 $2FT$ 个随机变量之间是线性相关的，而不是相互独立的。就是说，这个 $N=2FT$ 维连续型随机序列是有记忆的，在分析上是比较困难的。因此，希望能找到一组完备的正交函数集，按这组正交函数集展开后，能使所得到的随机序列的各随机变量之间是相互独立的或至少是相互线性无关的。满足这一要求的是卡休宁-勒维（Karhunen-Laeve）展开。假设$\{\varphi_l(r)\}$（$l=1$、2、…、n，其 n 可以是有限的，也可以是无限可数的）是在（a，b）区间内对于随机过程 $X(t)$ 满足卡休宁-勒维展开条件的一组归一化的完备正交函数集，如果随机过程 $X(t)$ 是平方可积的，即是能量有限的，则 $X(t)$ 可展开为

$$X(t) = \sum_{l=1}^{n} X_l \varphi_l(t) \tag{2-10}$$

其中展开式的系数

$$X_l = \int_a^b X(t)\varphi_l(t)\,\mathrm{d}t \tag{2-11}$$

$$l=1、2、\cdots、n$$

是按卡休宁-勒维展开得到的随机变量。这样，随机过程 $X(t)$ 就完全可以用相互间线性无关的随机变量 X_l（$l=1$、2、…、n，其中 n 可以是有限的，也可以是无限可数的）构成的随机序列来表示。

综上所述，可以用一维随机变量、N 维随机变量或者随机过程来描述不同统计特性的信源所输出的消息。关于信源的分类还有多种方法，讨论具体信源时再进一步介绍。下面各节主要讨论具体信源输出消息的信息度量及其有关性质。

2.2 离散信源的熵

信息论是建立在信息可以度量的基础上，研究如何有效、可靠地传输和处理信息的科学。也就是说，要从理论上研究有关信息传输和处理问题，必须对信息给予定量描述。信息度量是建立信息论的基础，是一个十分重要的概念。

本节首先从单符号离散信源的信息度量开始讨论，给出了信息量和熵的定义，并给出了条件熵和联合熵的定义。

单符号离散信源的输出是信源符号（消息）集合中的单个符号，是一种最简单的信源。表 2-1 列出了 27 个字符（其中包括 26 个英文字母和 1 个间隔符号）的出现概率，这是对英文普通文章大量统计的结果。如果一个离散信源是由 27 个字符所构成的符号（消息）集合，则表 2-1 就是描述该信源的概率空间。如果每次依据表 2-1 给出的各字符的概率随机地选取出一个字符，且每次必定选取出其中的一个字符，这样的信源就是单符号离散信源。

表 2-1 英文普通文章中字符的概率统计表

序 号	字 符	概 率	序 号	字 符	概 率	序 号	字 符	概 率
1	间隔	0.1817	10	H	0.04305	19	P	0.01623
2	E	0.1073	11	D	0.03100	20	W	0.01260
3	T	0.0856	12	L	0.02775	21	B	0.01179
4	A	0.0668	13	F	0.02395	22	V	0.00752
5	O	0.0654	14	C	0.02260	23	K	0.00344
6	N	0.0581	15	M	0.02075	24	X	0.00136
7	R	0.0559	16	U	0.02010	25	J	0.00108
8	I	0.0519	17	G	0.01633	26	Q	0.00099
9	S	0.0499	18	Y	0.01623	27	Z	0.00063

实际中也存在着许多这样的单符号离散信源。例如研究掷一骰子下落后朝上一面的点数，每次试验结果必然是1点、2点、3点、4点、5点、6点中的一个面朝上。其输出消息是“朝上面是1点”、“朝上面是2点”、……、“朝上面是6点”等6个不同消息。每次试验出现哪一个消息是随机的，但必定是出现其中一个消息。如果骰子是均匀的，各个点数都是以等概率出现，即都是1/6。因此，根据式（2-1）可把这个信源抽象后得到数学模型

$$\begin{pmatrix} X \\ p(x) \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

并满足

$$\sum_{i=1}^{6} p(x_i) = 1$$

这种离散信源能输出多少信息呢？信源输出的每个消息又能够为其接收者提供多少信息呢？下面就来研究这些有关的信息度量问题。

2.2.1 信息量的定义

通信系统传送的对象是消息，信息包含其中。信源发出的消息是随机的，接收者在收到消息之前对信源发出的消息是不预知的。消息对接收者来说的这种不确定性是客观存在的。消息的传递过程就是一种使接收者从不知到有所知或完全知的过程。通信的过程就是一种消除不确定性的过程，从不确定到比较确定或完全确定的过程。接收者获得信息的多少是与不确定性的减少直接有关的。这样可以直观地把信息量定义为：

信息量 = 消息不确定度的减小量

= （收到消息之前关于某事件发生的不确定度）-（收到消息之后关于该事件发生的不确定度）

信息量的这个定义，应该理解为消息接收者收到消息后获得关于某事件发生的信息量，即收到某消息获得的信息量。

如果在消息传递过程中没有产生任何差错，即在没有干扰的情况下，可以完全不失真地收到信源所发生的消息。因此，收到消息之后，关于某事件发生的不确定性就完全消除了，即收到消息后对该事件发生的不确度变为零。那么，在没有干扰的情况下，信息量的定义就变为：

信息量 = 收到消息之前关于某事件发生的不确定度

= 消息的不确定度