信息论与编码

傅祖芸　赵建中　编著

http://www.phei.com.cn

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY

信息论与编码

傅祖芸　赵建中　编著

電子工業出版社
Publishing House of Electronics Industry
北京・BEIJING

内 容 简 介

本书系统地论述信息论与纠错编码的基本理论。共有 9 章，内容包括：信息的定义和度量；离散信源和连续信源的信息熵；信道和信道容量；平均失真度和信息率失真函数；三个香农信息论的基本定理：无失真信源编码定理、限失真信源编码定理和信道编码定理；若干种常见实用的无失真信源压缩编码的方法；以及信道纠错编码的基本内容和分析方法。

本书文字通顺、概念清晰、系统性和可读性强。可作为高等院校信息与计算科学、信息与通信工程等相关专业的本科生教材或教学参考书，也可供科研院所从事信息科学理论、技术、方法研究的科研和工程技术人员参考。

图书在版编目(CIP)数据

信息论与编码/傅祖芸，赵建中编著. —北京：电子工业出版社，2006.4
ISBN 7-121-02485-3

Ⅰ.信… Ⅱ.①傅…②赵… Ⅲ.①信息论—高等学校—教材②信源编码—编码理论—高等学校—教材③信道编码—编码理论—高等学校—教材 Ⅳ.TN911.2

中国版本图书馆 CIP 数据核字(2006)第 033346 号

责任编辑：陈晓莉
印　　刷：北京牛山世兴印刷厂
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编　100036
经　　销：各地新华书店
开　　本：787×960　1/16　印张：25.25　字数：565 千字
印　　次：2006 年 4 月第 1 次印刷
印　　数：5 000 册　定价：35.00 元

凡购买电子工业出版社的图书，如有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系。联系电话：(010)68279077。质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至dbqq@phei.com.cn。

前　言

信息论与编码是一门应用概率论、随机过程和数理统计等方法来研究信息的存储、传输、处理、控制和利用一般规律的科学。它主要研究如何提高信息系统的可靠性、有效性、保密性和认证性，以使信息系统最优化。自20世纪中叶香农信息论问世以来，信息理论本身得到不断地发展和深化，尤其是在这理论指导下信息技术也得到飞快发展。这又使对信息的研究冲破了香农狭义信息论的范畴，几乎渗透到自然科学与社会科学的所有领域，从而形成了一门具有划时代的新兴学科——信息科学。所以信息论是信息科学发展的源泉，也是信息科学的基础理论。

在近七八年中，当人类迈入21世纪——高度信息化时代以来，移动通信、互联网通信、多媒体技术、计算机技术、空间技术等信息技术出现了超出人们想像的、前所未有的发展速度。在这些领域中，只要涉及信息的存储、传输和处理的就要用到香农信息论的理论——无失真通信的传输速率极限（即香农极限）、无失真和限失真信源编码理论（即数据压缩原理）和信道编码理论（即纠错码理论）等。甚至日常生活娱乐中如数字激光影碟机、数字家庭音像系统等都普遍采用了纠错码技术和数据压缩技术。所以，现在人们对于信息论的基本理论已不再感到陌生、抽象和难以理解和掌握，也越来越感到学习和掌握信息理论的必要和重视。

在这种形势下，各高校的热门专业“信息工程技术专业”也得到快速发展，专业的知识结构也作了相应调整，都先后开设了“信息论与编码”的有关课程，作为本科生、研究生必修的专业基础课。与此同时，于1998年以来，全国几百所高校先后在理学院（或数学系）内新增设了“信息与计算科学专业”，报考这一专业的学生也与日俱增。根据2005年教育部下发的“信息与计算科学专业”的教学规范，“信息与计算科学专业”就是要培养具有良好的数学基础和数学思维能力，掌握信息或计算科学的基本理论、方法与技能，能解决信息科学技术和信息工程计算中实际问题的高级理论研究型人才和技术型人才。无可置疑，信息论与编码理论必然是此专业的必修基础课之一。

“信息论与编码”是一门具有广泛的数学理论与知识，又有实际工程背景的课程，两者缺一不可。对信息工程技术类专业的学生来说，他们具有一定的通信、电子和计算机方面的实际工程知识，虽然已有一定的数学基础，但学生普遍对繁杂的数学公式感到困难而望而止步。对数学系信息与计算科学专业的学生来说，虽然他们具有较好的数学基础和数学思维能力，但更突出的是缺乏通信、电子等方面的实际工程知识。

针对上述情况，依据作者多年在信息论与编码方面的教学和科研经验及近年来对“信息与计算科学专业”的学生讲授《信息论与编码》课的教学实践总结，我们反复

商讨，决定以《信息论——基础理论与应用》[15]一书为蓝本，基本保留原有体系，删去一些对本科生来说较深的内容，以及在后续课程中将要讲述的内容，增加压缩编码和纠错码的内容，终而编写成本书，以期能适应不同专业的需求。

本书主要是系统地介绍香农(Shannon)信息论和编码理论。全书注重基本概念，基本定理和基本分析方法的论述，并列举许多例题，结合实例建立概念和数学模型，给出详细的、必要的数学推演过程和证明，力求物理概念清晰，逻辑性、系统性强，数学结构严谨完整又避免纯数学的枯燥乏味。在内容的编排上，力求由浅入深、循序渐进，合理而系统地安排章节。全书力求做到既有实际应用背景，又有清晰的数学概念和数学思想。

全书共分9章。第1、2、3、4章是全书的基础。首先阐述信息的概念，引出香农信息的定义，信息论研究的目的、内容。在这基础上讨论各类离散信源、连续和波形信源的信息测度——信息熵及离散信道、连续和波形信道的信息传输率和信道容量。第5、6、7章主要论述香农信息论的三个基本定理——离散信源的无失真编码定理、有噪信道编码定理及限失真信源编码定理。此部分内容是香农信息论的核心部分。第8章介绍若干种常见实用的无失真信源压缩编码方法。第9章在给出必要的数学知识基础上，论述信道纠错编码的基本内容和分析方法及一些主要的信道编码方法。为帮助读者掌握分析和解决问题的能力，各章配有大量的习题。书后的附录，为读者提供了所需的一些数学基础知识。为配合本书的学习和解题，作者编写出版了《信息论与编码学习辅导及习题详解》[27]一书，可供读者学习使用。

全书在三个定理证明上采用了统一的弱ε典型序列的分析方法，使定理证明简洁明了。但这些章节均标以“*”号出现。书中标有“*”的章节和小字体部分均属于严格的数学证明或加深、加宽的内容。各高校、各专业可根据学时的多少或学生的知识程度适当取舍，省略“*”的章节和小字体部分。省略后并不影响全书的系统性、逻辑性和可读性。所以本书可作为数学系信息与计算科学专业的教材，也可作为信息与通信工程技术相关专业的教材。

本书由傅祖芸主编，第8章字典码一节和第9章由赵建中编写，并由傅祖芸负责全书的修改和统编。孙建京、路而红、刘泉、彭一凡、陈立、赵黎明、施燕琼等同志参与了审稿、绘图、誊抄等大量工作，在此表示衷心的感谢。

在本书的编写过程中，参阅了国内外一些经典著作，均列于参考书目中，在此谨向作者表示深切谢意。

电子工业出版社陈晓莉编审对本书的出版做了大量的工作并提出了宝贵的意见，使本书得以顺利出版，也在此深表感谢。

有关书中的不妥和错误之处，殷切希望广大读者予以批评指正。

作　者

2005年12月

目　录

第1章 绪 论

信息论是人们在长期通信工程的实践中，由通信技术与概率论、随机过程和数理统计相结合而逐步发展起来的一门科学。通常人们公认信息论的奠基人是当代伟大的数学家和美国贝尔实验室杰出的科学家香农(C. E. Shannon)，他在1948年发表了著名的论文《通信的数学理论》，为信息论奠定了理论基础。近半个世纪以来，以通信理论为核心的经典信息论，正以信息技术为物化手段，向高精尖方向迅猛发展，并以神奇般的力量把人类社会推入了信息时代。随着信息理论的迅猛发展和信息概念的不断深化，信息论所涉及的内容早已超越了狭义的通信工程范畴，进入了信息科学这一更广阔、更新兴的领域。

本章首先引出信息的概念，进而讨论信息论这一学科的研究对象、目的和内容，并简述本学科的发展历史、现状和动向。

1.1 信息的概念

人类从产生那天起，就生活在信息的海洋之中。

人类社会的生存和发展，一时一刻都离不开接收信息、传递信息、处理信息和利用信息。

自古以来，人们就对信息的表达、存储、传送和处理等问题进行了许多研究。原始人的“结绳记事”也许是最初期的表达、存储和传送信息的方法。我国古代的“烽火告警”是一种最早的快速、远距离传递信息的方式。语言和文字则是人类社会用来表达和传递信息的最根本的工具。自从造纸术和印刷术的发明，使信息表示和存储方式产生了一次重大的变化，使文字成为信息记录、存储和传递的有效手段。特别是电报、电话和电视的发明，使信息传送快速、便利、远距离，再次出现了信息加工和传输的变革。近百年来，随着生产和科学技术的发展，使信息的处理、传输、存储、提取和利用的方式及手段达到了更新、更高的水平。

近代，电子计算机的迅速发展和广泛应用，尤其个人微型计算机得以普及，大大提高了人们处理加工信息、存储信息及控制和管理信息的能力。

20世纪后半世纪，计算机技术、微电子技术、传感技术，激光技术、卫星通信和移动通信技术、航空航天技术、广播电视技术、多媒体技术、新能源技术和新材料技术等新技术的发展和应用，尤其近年来以计算机为主体的互联网技术的兴起和发展，它们相互结合、相互促进，以空前未有的威力推动着人类经济和社会高速发展。正是这些现代新科学新技术汇成了一股强大的时代潮流，将人类社会推入到高度化的信息时代。

在当今“信息社会”中，人们在各种生产、科学研究和社会活动中，无处不涉及信息的交换和利用。迅速获取信息，正确处理信息，充分利用信息，就能促进科学技术和国民经济的飞跃发展。可见，信息的重要性是不言而喻的。

那么，什么是信息呢？

信息是信息论中最基本、最重要的概念，它是一个既抽象又复杂的概念。这一概念像在实践中提出来的其他科学概念一样，是在人类社会互通情报的实践过程中产生的。在现代信息理论形成之前的漫长时期中，信息一直被看做是通信的消息的同义词，没有赋予它严格的科学定义。到了 20 世纪 40 年代末，随着信息论这一学科的诞生，信息的含义才有了新的拓展。

在日常生活中，信息常常被认为就是“消息”、“情报”、“知识”、“情况”等。的确，信息与它们之间是有着密切联系的。但是，信息的含义是要更深刻、更广泛，它是不能等同于消息、情报、知识和情况的。

信息不能等同于情报。情报往往是军事学、文献学方面的习惯用词。如“对敌方情况的报告”，“文献资料中对于最新情况的报道或者进行资料整理的成果”等称为情报。在“情报学”，这一新学科中，它们对于“情报”是这样定义的，“**情报**是人们对于某个特定对象所见、所闻、所理解而产生的知识”。可见，情报的含义要比“信息”窄得多。它只是一类特定的信息，不是信息的全体。

信息也不能等同于知识。**知识**是人们根据某种目的，从自然界收集得来的数据中，整理、概括、提取得到有价值的、人们所需的信息。知识是一种具有普遍和概括性质的高层次的信息。例如，如图 1.1 所示。有一堆 A、B 两所大学学生的考试成绩数据。为了了解 A、B 两所大学学生的学习成绩水平的差别，而进行统计处理，得到一张曲线图。从中获得了有关 A、B 两所大学学生学习水平的知识。当然，还可以从这堆数据中获得其他有关知识（两所大学男、女生成绩差别等）。又例如，获得大量的遥感图片数据，根据不同目的，处理后可以得到不同的知识（地质知识、地形知识、水源知识等）。由此可知，知识是以实践为基础，通过抽象思维，对客观事物规律性的概括。知识信息只是人类社会中客观存在的部分信息。所以知识是信息，但不等于信息的全体。

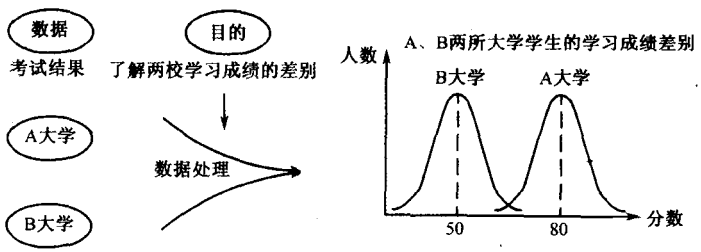

图 1.1 统计处理后的分布曲线

信息也不能等于消息。人们也常常错误地把信息等同于消息，认为得到了消息，就是得到了信息。例如，当人们收到一封电报，接到一个电话，收听了广播或看了电

视等以后，就说得到了“信息”。的确，人们从接收到的电报、电话、广播和电视的消息中能获得各种信息，信息与消息有着密切的联系。但是，信息与消息并不是一件事，不能等同。

我们知道，在电报、电话、广播、电视（也包括雷达、导航、遥测）等通信系统中传输的是各种各样的消息。这些被传送的消息有着各种不同的形式，例如：文字、符号、数据、语言、音符、图片、活动图像等。所有这些不同形式的消息都是能被人们感觉器官所感知的，人们通过通信，接收到消息后，得到的是关于描述某事物状态的具体内容。例如，听气象广播，气象预报为“晴间多云”，这就告诉了我们某地的气象状态，而“晴间多云”这广播语言则是对气象状态的具体表述。又如，我们收到一份电报为“母病愈”，则得知了母亲的身体健康状况，报文“母病愈”是对母亲身体健康状况的一种描述。再例如，电视中转播球赛，人们从电视图像中看到了球赛进展情况，而电视的活动图像则是对球赛运动状态的描述。可见，语言、报文、图像等消息都是对客观物质世界的各种不同运动状态或存在状态的表述。当然，消息也可用来表述人们头脑里的思维活动。例如，朋友给你打电话，电话中说：“我想去上海”，你就得知了你的朋友的想法。这时，此语言消息则反映了人的主观世界——大脑物质的思维运动所表现出来的思维状态。

因此，用文字、符号、数据、语言、音符、图片、图像等能够被人们感觉器官所感知的形式，把客观物质运动和主观思维活动的状态表达出来就成为**消息**。

可见，消息中包含信息，是信息的载体。得到消息，从而获得信息。同一则信息可用不同的消息形式来载荷。如前例中，球赛进展情况可用电视图像、广播语言、报纸文字等不同消息来表述。而一则消息也可载荷不同的信息，它可能包含非常丰富的信息，也可能只包含很少的信息。因此，信息与消息是既有区别又有联系的。

既然信息不同于消息，当然也不同于信号。

在各种实际通信系统中，往往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，它是消息的运载工具。如前例中，“母病愈”这种关于母亲身体健康状况的信息，用汉文“母病愈”的消息来表述，然后通过电报系统传送到另一地的收信者。由于这个电报系统的传递信道是无线电电波信道，所以汉文消息不能直接在信道中传输。一般，需先将汉文（例如“母病愈”）变换成四位码，然后变换成由点、划和空隔三种符号组成的莫尔斯码，再转换成脉冲电信号，然后经过调制变成高频调制电信号，才能在信道中传输。此时，脉冲电信号或高频调制电信号都载荷着汉文消息，表述了母亲身体健康的一种状态。在通信系统的接收端，通过解调，反变换，若无干扰的话就可恢复成原汉文消息——“母病愈”。收信者收到报文后，就得知了母亲病愈，身体健康，从而获得了信息。可见，信号携带信息，但不是信息本身。同样，同一信息可用不同的信号来表示。同一信号也可表示不同的信息。例如，红、绿灯信号。若在十字路口，红、绿灯信号表示能否通行的信息。若在电子仪器面板上，红、绿灯信号却表示仪器是否正常工作或者表示高低电压等信息。所以，信息、消息和信号是既有区别又有联系的三个不同的概念。

关于信息的科学定义，到目前为止，国内外已有不下百余种流行的说法。它们都是从不同的侧面和不同的层次来揭示信息的本质的。

最早对信息进行科学定义的，是哈特莱(R. V. L. Hartley)。他在1928年发表的《信息传输》一文中，首先提出“信息”这一概念。他认为，发信者所发出的信息，就是他在通信符号表中选择符号的具体方式，并主张用所选择的自由度来度量信息。

哈特莱的这种理解在一定程度上能够解释通信工程中的一些信息问题，但它存在着严重的局限性。首先，他所定义的信息不涉及信息的价值和具体内容，只考虑选择的方式。其次，即使考虑选择的方法，但没有考虑各种可能选择方法的统计特性。正是这些缺陷严重地限制了它的适用范围。

1948年，控制论的创始人之一，美国科学家维纳(N. Wiener)出版了《控制论——动物和机器中通信与控制问题》一书。维纳在该书中是这样来论述信息的，他指出：“信息是信息，不是物质，也不是能量”[①]。这就是说，信息就是信息自己，它不是其他什么东西的替代物，它是与“物质”、“能量”同等重要的基本概念。正是维纳，首先将“信息”上升到“最基本概念”的位置。

后来，维纳在《人有人的用处》[②]一书中，他提出：“信息是人们适应外部世界并且使这种适应反作用于外部世界的过程中，同外部世界进行互相交换的内容的名称。”又说：“接收信息和使用信息的过程，就是我们适应外部世界环境的偶然性变化的过程，也是我们在这个环境中有效地生活的过程。”“要有效地生活，就必须有足够的信息。”的确，信息对人类的生存是很重要的；但是，信息不仅仅与人类有关，不仅仅是人与外部世界交换的内容。在自然界中，一切生物体都在与外部世界进行着互相交换信息。一切生物体都有它们独自的接收信息和交换信息的方式。俗话说“禽有禽言，兽有兽语”，这是动物之间特别是群体动物之间传递信息的方式。人们发现动物之间可以利用气味、声音、不同的运动姿态、乃至超声波、电磁场等多种方式来传递信息。另外，信息确是人们与外部世界互相交换的内容，但是，人们在与外部世界相互作用过程中，还进行着物质与能量的交换。这样，就又把信息与物质、能量混同起来。所以，维纳关于信息的定义是不确切的。

关于信息的定义，有人提出用变异度、差异量来度量信息，认为“信息就是差异”。这种说法的典型代表是意大利学者朗格(G. Longe)。他在1975年出版的《信息论：新的趋势与未决问题》一书序言中，提出：“信息是反映事物的形式、关系和差别的东西。信息是包含于客体间的差别中，而不是在客体本身中。”，“在通信中仅仅差别关系是重要的”也就是说，他定义信息是客体之间的相互差异。的确，宇宙内到处存在着差异，差异的存在使人们存在着“疑问”和“不确定性”。从这个角度看，差异确是信息。但是，并不能说没有差异就没有信息。所以，这样定义的信息也是不全面的、不确切的。

而香农在1948年发表了一篇著名的论文，“通信的数学理论”。他从研究通信系

① N. Wiener，《控制论——动物和机器中的通信与控制问题》，科学出版社，1963年

② N. Wiener，《人有人的用处》，商务印书馆，1978年

统传输的实质出发，对信息作了科学的定义，并进行了定性和定量的描述。

如前所述，各类通信系统——电报、电话、广播、电视、雷达、遥测……等传送的是各种各样的消息。消息的形式可以不同，但它们都是能被传递的，能被人们感觉器官（眼、耳、触觉等）所感知的，而且消息表述的是客观物质和主观思维的运动状态或存在状态。

香农将各种通信系统概括成如图 1.2 所示的框图。在各种通信系统中，其传输的形式是消息。但消息传递过程的一个最基本、最普通却又不十分引人注意的特点是：收信者在收到消息以前是不知道消息的具体内容的。在收到消息以前，收信者无法判断发送者将会发来描述何种事物运动状态的具体消息；他也更无法判断是描述这种状态还是那种状态。再者，即使收到消息，由于干扰的存在，他也不能断定所得到的消息是否正确和可靠。总之，收信者存在着“不知”、“不确定”或“疑问”。通过消息的传递，收信者知道了消息的具体内容，原先的“不知”、“不确定”和“疑问”消除或部分消除了。因此，对收信者来说，消息的传递过程是一个从不知到知的过程，或是从知之甚少到知之甚多的过程，或是从不确定到部分确定或全部确定的过程。如果不具备这样一个特点，那就根本不需要通信系统了。试想，如果收信者在收到电报或电话之前就已经知道报文或电话的内容，那还要电报、电话系统干什么呢？

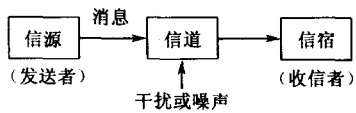

图 1.2 通信系统框图

由于主、客观事物运动状态或存在状态是千变万化的、不规则的、随机的。所以在通信以前，收信者存在“疑义”和“不知”。例如，在电报通信中，收报人在收到报文前，首先他不知何人会给他发电报，而且也不知将要告诉他什么事情。若当他收到报文是“母病愈”后，才能确定是他家人告诉他母亲的身体情况。其次，报文“母病愈”是母亲身体健康状态的一种描述，而母亲身体健康情况会表现出不同的状态，到底出现的是什么状态是随机的、变化的。收信者在看到报文以前，他不能确定母亲身体健康状态如何，也存在“不确定性”。只要报文是清楚的，在传递过程中没有差错，那么，他收到报文以后，他原来所有的“不确定性”都没有了，他就获得了所有的信息。如果在传递过程中存在着干扰，使报文完全模糊不清，收信者收到报文以后，原先所具有的不确定性一点也没有减少，他就没有获得任何信息。如果干扰使报文发生部分差错，使收信者原先的不确定性减少了一些，但没有全部消除，他就获得了一部分信息。所以，通信过程是一种消除不确定性的过程。不确定性的消除，就获得了信息。原先的不确定性消除得越多，获得的信息就越多。如果原先的不确定性全部消除了，就获得了全部的信息；若消除了部分不确定性，就获得了部分信息；若原先不确定性没有任何消除，就没有获得任何信息。由此可见，**信息是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**。

从以上分析可知，在通信系统中形式上传输的是消息，但实质上传输的是信息。

消息只是表达信息的工具，载荷信息的客体。显然，在通信中被利用的(亦即携带信息的)实际客体是不重要的，而重要的是信息。信息较抽象，而消息是较具体的，但还不一定是物理性的。通信的结果是消除或部分消除不确定性从而获得信息。

根据香农的有关信息的定义，信息如何测度呢？当人们收到一封电报，或听了广播，或看了电视，到底得到多少信息量呢？显然，信息量与不确定性消除的程度有关。消除多少不确定性，就获得多少信息量。那么，不确定性的大小能度量吗？

用数学的语言来讲，不确定就是随机性，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——概率论和随机过程来测度不确定性的大小。若从直观概念来讲，不确定性的大小可以直观地看成是事先猜测某随机事件是否发生的难易程度。

例如，假设有甲、乙两个布袋，各袋内装有大小均匀，对人手感觉完全一样的球100个。甲袋内红、白球各50个，乙袋内有红、白、蓝、黑四种球，各25个。现随意从甲袋或乙袋中取出一球，并猜测取出的是什么颜色的球，这事件当然具有不确定性。显然，从甲袋中摸出是红球要比从乙袋中摸出是红球容易得多。这是因为，在甲袋中只在“红”与“白”两种颜色中选择一种，而且“红”与“白”机会均等，即摸取的概率各为$\frac{1}{2}$。但在乙袋中，红球只占$\frac{1}{4}$，摸出是红球的可能性就小。自然，“从甲袋中摸出的是红球”比“从乙袋中摸出的是红球”的不确定性来得小。从这例子得出，不确定性的大小与可能发生的消息数目及各消息发生的概率有关。

再例如气象预报，我们知道可能出现的气象状态有许多种。以十月份北京地区天气为例，经常出现的天气是“晴间多云”、“晴”或“多云”，其次是“多云转阴”、“阴”、“阴有小雨”等，而“小雪”这种天气状态出现的概率是极小的，“大雪”的可能性则更小更小。因此，在听气象预报前，我们大体上能猜测出天气的状况。由于出现“晴间多云”、“晴”或“多云”的可能性大，我们就比较能确定这些天气状况的出现。因此，当预报明天白天“晴间多云”或“晴”，我们并不觉得稀奇，因为和我们猜测的是基本一致，所消除的不确定性要小，获得的信息量就不大。而出现“小雪”的概率很小，我们很难猜测它是否会出现，所以这事件的不确定性很大。如果预报是“阴有小雪”，我们就要大吃一惊，感到气候反常，这时就获得了很大的信息量。出现“大雪”的概率更小，几乎是不可能出现的现象，它的不确定性更大。如果一旦出现“大雪”的气象预报，我们将万分惊讶，这时将获得更大的信息量。由此可知，某一事物状态出现的概率越小，其不确定性越大；反之，某一事物状态出现的概率接近于1，即预料中肯定会出现的事件，那它的不确定性就接近于零。

这两例子告诉我们：某一事物状态的不确定性的大小，与该事物可能出现的不同状态数目及各状态出现的概率大小有关。既然不确定性的大小能够度量，可见，信息是可以测度的。

我们把某事物各种可能出现的不同状态，即所有可能选择的消息的集合，称为**样本空间**。每个可能选择的消息是这个样本空间的一个元素。对于离散消息的集合，概率测度就是对每一个可能选择的消息指定一个概率(非负的，且总和为1)。一个样

本空间和它的概率测度称为一个**概率空间**。

一般概率空间用$[X,P]$来表示。在离散情况下，X的样本空间可写成$\{a_1,a_2,\cdots,a_q\}$。样本空间中选择任一元素a_i的概率表示为$P_X(a_i)$，其脚标X表示所考虑的概率空间是X。如果不会引起混淆，脚标可以略去，写成$P(a_i)$。所以在离散情况下，概率空间为

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} a_1, & a_2,\cdots, & a_q \\ P(a_1), & P(a_2),\cdots, & P(a_q) \end{bmatrix}$$

其中$P(a_i)$就是选择符号a_i作为消息的概率，称为**先验概率**。在接收端，对是否选择这个消息(符号)a_i的不确定性是与a_i的先验概率成反比的，即对a_i的不确定性可表示为先验概率$P(a_i)$的倒数的某一函数。我们取该函数为对数函数，并把这样定义的不确定性称为该消息(符号)a_i的**自信息**：

$$I(a_i) = \log \frac{1}{P(a_i)} \tag{1.1}$$

由于信道中存在干扰，假设接收端收到的消息(符号)为b_j，这个b_j可能与a_i相同，也可能与a_i有差异。我们把条件概率$P(a_i|b_j)$称为**后验概率**，它是接收端收到消息(符号)b_j后而发送端发的是a_i的概率。那么，接收端收到b_j后，发送端发送的符号是否是a_i尚存在的不确定性应是后验概率的函数，即是$\log \frac{1}{P(a_i|b_j)}$。于是，收信者在收到消息(符号)$b_j$后，已经消除的不确定性为：先验的不确定性减去尚存在的不确定性。这就是收信者获得的信息量，定义为互信息：

$$I(a_i;b_j) = \log \frac{1}{P(a_i)} - \log \frac{1}{P(a_i \mid b_j)} \tag{1.2}$$

如果信道没有干扰，信道的统计特性使a_i以概率1传送到接收端。这时，收信者接到消息后，尚存在的不确定性就等于零，即$P(a_i|b_j)=1$，$\log \frac{1}{P(a_i|b_j)}=0$，不确定性全部消除。由此得互信息

$$I(a_i;b_j) = I(a_i) \tag{1.3}$$

以上就是香农关于信息的定义和度量。通常也称为**概率信息**。

香农定义的信息概念在现有的各种理解中，是比较深刻的，它有许多优点。

首先，它是一个科学的定义，有明确的数学模型和定量计算。

其次，它与日常用语中的信息的含意是一致的。例如，设某一事件a_i发生的概率等于1，即a_i是预料中一定会发生的必然事件，如果事件a_i果然发生了，收信者将不会得到任何信息(日常含义)，因为他早知道a_i必定发生，不存在任何不确定性。

根据式(1.1)，因为$P(a_i)=1$，所以得

$$I(a_i)=\log \frac{1}{P(a_i)}=0$$

即自信息等于零。反之，如果a_i发生的概率很小，即猜测它是否发生的不确定性很大，一旦a_i果然发生了，收信者就会觉得很意外和惊讶，获得的信息量很大。根据式(1.1)，因为$P(a_i)\ll 1$，故得