**<https://www.bilibili.com/video/BV1pJ411U7G8>**

**信息论与编码**

1. 绪论
2. **正式内容前--**信息论的概念、形成和发展
   1. 信息系统模型
   2. 通信系统模型
   3. 信息论的研究内容
   4. 本书章节结构

**第一章 绪论**

学一个东西，我们脑子里出现的第一个问题就是：这是个什么东西？

这门课叫“信息论与编码”。那么问题来了：**“信息论”是个什么东西？“编码”是用来干什么的？**

为便于理解，在正式开始本书内容之前，先看一点“不正式”的东西，就叫做“正式内容前”吧。

**0 正式内容前--**信息论的概念、形成和发展

**（1）信息是什么？**

信息论是研究信息的理论，但：“信息是什么”？

奥古斯丁曾对“时间是什么”有过思考：“时间是什么？如果无人问我则我知道，如果我欲对发问者说明则我不知道”。

“信息”大概也类似。没有人问的时候，我们都以为我们当然知道什么是信息；但当人问起的时候，事实上我们不怎么知道。很多事情皆如此：我们以为非常熟悉的，我们甚至并不真正知道。

**麦克斯韦妖：**一个[绝热](https://baike.baidu.com/item/%E7%BB%9D%E7%83%AD" \t "_blank)容器被分成相等的两格，中间是由“妖”控制的一扇小“门”，容器中的空气分子作无规则热运动时会向“门”上撞击，“门”可以选择性的将速度较快的分子放入一格，而较慢的分子放入另一格，这样，其中的一格就会比另外一格温度高，可以利用此温差，驱动[热机](https://baike.baidu.com/item/%E7%83%AD%E6%9C%BA" \t "_blank)做功。这是[第二类永动机](https://baike.baidu.com/item/%E7%AC%AC%E4%BA%8C%E7%B1%BB%E6%B0%B8%E5%8A%A8%E6%9C%BA" \t "_blank)的一个范例。

永动机当然不可能实现。但麦克斯韦妖的假说也告诉我们：如果我们知道某些信息（比如分子的运动速度），我们确实可以把无序的系统（比如绝热容器内做无规则分子运动的气体）变得更有序一些（按分子运动速度快慢分成两个部分）。

信息论的创始人香农给信息下了一个定义：**信息是能使不确定性减少或者消除的东西**。

绝热容器内无规则运动的气体分子，其状态是不确定的。麦克斯韦妖通过获知分子运动速度的快慢（信息），可以将其分成温度不同的两个部分，虽然分子的状态还不是完全确定的，但至少已经部分确定了（一部分运动的快，一部分运动的慢），或者说：分子状态的不确定性减少了。

掷硬币的时候，告诉你掷出的是哪一面（获得信息），则关于掷硬币的结果，从完全未知变成了完全已知，不确定性消除了。

掷骰子的时候，告诉你掷出的是几（获得信息），则关于骰子上面的点数，从完全未知变成了完全已知，不确定性消除了。

所以，**信息是什么？信息是能使不确定性减少或者消除的东西**。

当然，上述有关“信息”含义的解释，只是众多解释中的一种（香农的解释）。实际上，有关信息的定义，不下几十种。或者也可以说，没有一个统一的有关信息的定义和解释。

这有点像著名的古代故事瞎子摸象：

从前有个王公，对他的仆人说，“去找十个天生的瞎子到我的宫殿来……并展示给他们一头大象” 。仆人按照王公的命令完成了差遣。当盲人到来的时候，王公对他们说：“这里是一头大象，检查它，然后告诉我大象是种什么东西”。

第一个个子高的盲人发现了大象的头，他说“大象像个大罐子”；

第二个个子矮的盲人触摸到大象的腿，他声称“大象像个柱子”；

第三个盲人是个办事有条不紊的人，他仔细听并感觉大象摆动耳朵扇动的空气，然后抓住大象的耳朵，感觉到它稀疏蓬乱的毛发，他高兴地笑起来，说道：“这头大象像把扇子”；

第四个非常谦逊的盲人触摸着大象的尾巴，他说“大象像一根有些磨损的绳子”；

第五个是胆大的盲人，他走近了这头大象的长牙，感觉长牙坚硬光滑乳白色的表面和长牙的尖头，他得出结论，“这大象像长矛一样坚硬和锋利”；

第六个个子矮小的盲人触摸到尾巴末端的一绺毛发，他说“大象像把刷子”；

第七个盲人触到大象的鼻子，他坚持大象像根树枝；

第八个总是慌慌张张的盲人撞到了大象的背部，他认为大象像个研钵；

第九个盲人个子很高，匆忙之中，他直接跑到了大象的侧边，他伸开他的手臂感觉这个动物宽大光滑的侧面，他说“这是个像一堵墙的动物”；

然后，这九个盲人开始争吵。

第十个盲人非常聪明，他静静地等着，不说话，直到所有其他人观察并说出他们的发现。他听了一会儿他们的争吵，然后在这头大象的周围转了转说：“我不知道一头大象像什么，这就是我打算写一部大象文章的原因，以证明不可能说出大象是种什么东西”。

这也是我写这本关于信息论的书的原因，用于证明不可能说出信息是种什么东西。

**注**：部分内容引自（美）布尔金著，王恒君等译，“信息论 本质·多样性·统一”，知识产权出版社，2015年。

**（2）信息的度量**

毕达哥拉斯说：“万物皆数”。

那么，信息的多少，也得有个“数”，叫做信息的量，简称信息量。

所以我们一定会问：“知道了掷硬币的结果”，相比“知道了掷骰子的结果”，哪个获得更多的信息量？各获得多少信息量？

既然“**信息是能使不确定性减少或者消除的东西”**，获得多少信息，就可以使不确定性减少多少。或者说，通过某种途径使不确定性减少了多少，你就通过这种途径获得了多少信息。

因此，（从某个途径获得的）**信息的多少（信息量）=事件不确定性的减少量**。

也可以说成：这个**不确定事件本身包含的信息量，等于完全消除其不确定性所需的信息量。**

盘古开天辟地，可以假定盘古知道天地间所有的事。当我们遇到不能确定答案的事儿时，我们可以去问他。不过盘古有点故弄玄虚，所有的问题只回答“是”或者“否”。因此，我们不能问盘古“刚才掉到大海里的硬币是正面朝上还是反面朝上”，而可以问“刚才掉到大海里的硬币是正面朝上吗（或者：“是反面朝上吗？”，效果是一样的）”？当盘古回答“是”（或者“否”，效果是一样的）时，你关于硬币面的朝向问题的不确定性就完全消除了。你得到了一定量的信息。

有多少信息呢？

所有的度量，都需要定义一个基本单位。1米被定义为光在真空中行进1/299792458秒的距离；1秒被定义为[铯](https://baike.baidu.com/item/%E9%93%AF" \t "_blank)-133原子[基态](https://baike.baidu.com/item/%E5%9F%BA%E6%80%81" \t "_blank)的两个超精细[能阶](https://baike.baidu.com/item/%E8%83%BD%E9%98%B6" \t "_blank)之间[跃迁](https://baike.baidu.com/item/%E8%B7%83%E8%BF%81" \t "_blank)时所辐射的电磁波的周期的9,192,631,770倍的时间。

于是，我们定义信息量的基本单位为比特（bit），1比特定义为你从盘古回答你一个“是”（或者“否”，效果一样）所获得的信息量。

所以，2选1的不确定事件，知道答案后获得的信息量是1比特（因为只要盘古回答一次“是”或者“否”即可完全知道答案），换句话说：2选1的不确定事件，只要获得1比特的信息量，就可以完全消除其不确定性（此处假定两个选项等概率出现，不等概率的情况后面再讨论，下同）；或者说：2选1的不确定性里面，包含了1比特的信息。

4选1的不确定事件呢？你可以通过问盘古两次得到准确的答案：有A、B、C、D四个备选答案，你可以通过两两分组，通过两次提问得到明确的答案（比如：第1次你可以问盘古，答案是A、B之一吗？如回答“是”，则第2次你可以问，是Ａ吗？），所以，4选1的不确定性事件，知道答案后获得的信息量是2比特；换句话说：4选1的不确定事件，只要获得2比特的信息量，就可以完全消除其不确定性；或者说：4选1的不确定性里面，包含了2比特的信息。

8选1的不确定事件，可以通过3次二分分组，得到答案，所以包含3比特的信息。

……

以此类推，*n*选1的不确定事件，包含有比特的信息。为简单起见，我们经常把上述以2为底的对数中的“2”省略，简写为。

可选项越多，不确定越大，包含的信息量就越多。

换个角度来看：可选项越多，不可预测性越大（不确定性越大），每个选项是正确答案的概率越小（2选1的时候每个选项是正确答案的概率是1/2，4选1的时候是1/4，8选1的时候是1/8，…，*n*选1的时候是1/*n*），其包含的信息量越多。

一条信息，其信息量的大小与不可预测有着关联！

新闻之所以是新闻在于其出乎意外（难以预测到），越是意外越是有新闻价值。

所以，今天的新闻如果和昨天的一样，也就是没有新闻，信息量也就是零。

归纳一下，我们知道：

1. 一个选项，其出现的概率越小，包含的信息量就越多；
2. 确定性的事件，其出现的概率为1（100%），包含的信息量为0；
3. 信息量具有可加性。从两个完全不同的信息来源得到两条消息，获得的信息量是这两条消息包含的信息量之和。

根据上面三条关于信息量的性质，我们可以得到信息量的表达式为（不是严格证明）：

 （1-1）

其中，*I*表示某个选项包含的信息量，*p*表示其出现的概率。或者说：某个选项包含的信息量等于其出现概率的倒数取对数。

现在我们以2选1为例，可以看看上述例子中不等概率时的情形。

如果两个选项中一个的概率为*p*，则另一个的概率为1-*p*，两个选项分别对应的信息量为：

 （1-2）

和

 （1-3）

于是，这条消息的平均信息量（答案可能为选项1，也可能为选项2，平均信息量为两者的概率平均）为：

 （1-4）

图1.4给出了平均信息量随*p*的变化关系曲线。

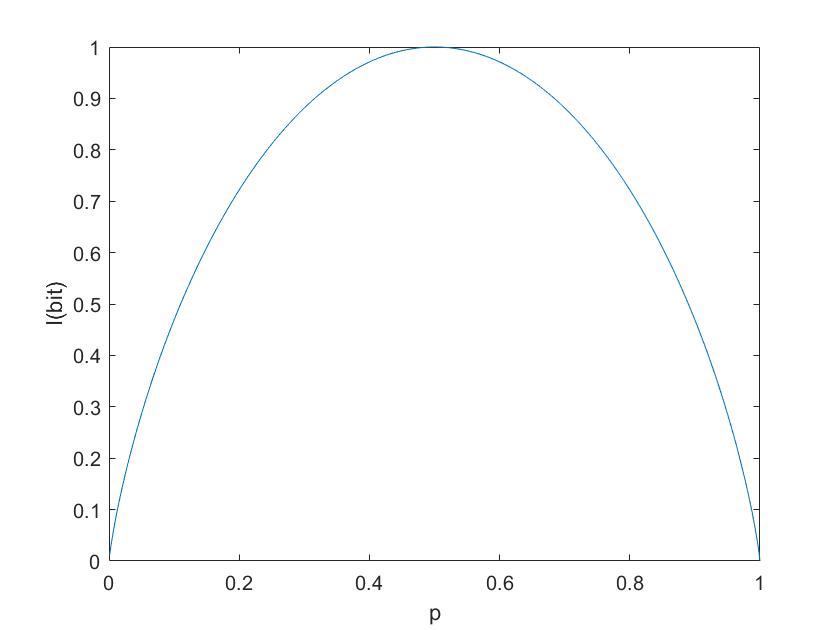


图1.4平均信息量随*p*的变化关系曲线

可以看出，当*p*=0.5时，*I*具有最大值1比特。这是容易理解的：当两个选项都是50%的概率出现时，具有最大的不确定性，所以包含的信息量也是最多的。

**（3）信息论的发展过程**

信息论，顾名思义，是研究信息的理论。

仅就术语而言，“信息”这个名词如今是如此普遍地充斥于我们生活的每一个角落，以至于我们都不会想到“信息到底是什么”这个问题。

科学家不一样。

数学家和控制论学家维纳（Norbert Wiener，1894-1964）大约是认真思考“信息是什么”的第一人。维纳关于“信息到底是什么”的思考导致他给出了关于信息的著名论述：“信息就是信息，不是物质，也不是能量”。

奈奎斯特（Harry Nyquist，1889-1976）和哈特莱（R. V. L. Hartley，1890~1970）在20世纪20年代致力于研究通信系统传输信息的能力。哈特莱第一次提出了信息量的概念，并企图用数学公式加以描述。

从维纳的开创性著作之后，信息科学诞生了许多种信息论，包括：统计信息论、语义信息论、算法信息论、语用信息论等等。

1948年，杰出工程师和数学家香农（Claude Elwood Shannon，1916-2001）在贝尔系统技术杂志上发表了他的一篇划时代论文--通信中的数学理论。香农本人称他创建的是一个通信理论，只是用到了信息量而已，所以使用的标题是“通信中的数学理论”。香农的追随者把这个理论改名为信息论。香农的理论影响之大，以至于一说到“信息论”，就让人想起香农的通信理论。所以，在以下的章节中，除非特别说明，我们提及的“信息论”，指的是香农信息论。

**（4）香农信息论的主要内容**

包括外在形式、内在含义和效用价值的认识论层次信息称为“全信息”。 仅计及其中形式因素的信息部分称为“语法信息”；同样的外在形式（语法），由于其语义不同，也是不同的信息，需考虑语义因素，称为“语义信息”；就算同样的语义，在不同的环境，对不同的对象，可能效用也是不一样的，计及效用的称为“语用信息”。

香农信息论只考虑信息的语法因素，不计及语义因素和语用因素。因此，香农信息论也被称为“狭义信息论”，只应用在它的严谨假设成立的情形下。换句话说，在现实中这样的假设从不成立（因为信息都是有语义和语用的），但这并不说明它毫无用处；有些东西只是“看起来很美”，比如“全信息”。丛林探索时，身着华服而寸步难行，倒不如衣衫褴褛而坚持前进。正因为香农抛弃了无法得出一致结论的语义和语用因素，才可以轻装上阵，使得关于信息的研究（特别是关于通信中的信息的研究），升华到一个新的层次。因此，香农被称为现代信息理论（特别是通信理论）的奠基人，一点也不为过。

香农信息论是在信息可以度量的基础上，研究**有效地和可靠地**传递信息的科学，它涉及信息度量、信息特性、信息传输速率、信道容量、干扰对信息传输的影响等方面。具体内容主要包括香农三定理：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。香农在 1948年论文中提出了无失真信源编码定理（香农第一定理），给出了简单的香农编码方法；在研究信源编码定理的同时，香农在1949年发表“噪声下的通信”，为信道编码奠定理论基础（信道编码定理，香农第二定理）。香农在1959年发表“保真度准则下的离散信源编码定理”，提出信息率失真理论，是频带压缩、数据压缩的理论基础（限失真信源编码定理，香农第三定理）。

香农信息论以[概率论](https://baike.baidu.com/item/%E6%A6%82%E7%8E%87%E8%AE%BA/829122" \t "_blank)、[随机过程](https://baike.baidu.com/item/%E9%9A%8F%E6%9C%BA%E8%BF%87%E7%A8%8B/368895" \t "_blank)为基本研究工具，研究广义通信系统的极限性能。包括香农、奈奎斯特在内的一批人，致力于研究接近极限性能的途径和手段，形成了较为完善的编码理论，提出了众多的编码方法。这些研究内容，和香农的信息理论一起，构建了香农信息论的主要结构。

* 1. **信息系统模型**

“正式内容前”看过以后，需要开始“正式内容”了。

“信息论”，顾名思义，是研究“信息”的理论。

所以，我们首先需要看一下，一个关于“信息”的系统大致是什么样的。

图1.1表示从信息角度看人类认识和改造外部世界的过程：

1. 人首先通过感觉器官（眼、耳、鼻、舌、皮肤）获取外部世界的信息（图像、颜色、声音、气味、味道、温度等）--称之为“信息获取”；
2. 通过神经系统把这些信息传递给大脑—称之为“信息传递”；
3. 大脑对感官获取的信息进行处理，并做出某种应对策略（如热了需要降温，或者冷了需要保暖等）--称之为“信息处理”；
4. 然后把这种决策信息通过神经系统告诉效应器官（手、脚等）--再次“信息传递”；
5. 效应器官做出某种行动，作用于外部世界—信息施效（或信息执行）。

人类认识和改造外部世界的过程，从“信息”的角度看，经历了“信息获取”—“信息传递”—“信息处理”—“信息施效”这样的步骤。

**神经系统**

**思维器官**

**神经系统**

**信息**

**策略**

**感觉器官**

**信息源**

**效应器官**

**信息宿**

**外部世界**

**信息**

**行为**

**信息**

**策略**

图1.1 人类认识和改造外部世界的过程框图

因此，我们可以从这个例子抽象出来一个信息系统的通用框图，如图1.2所示。

图1.2 信息系统框图

**信息传递**

**神经系统**

**信息处理 → 信息再生**

**思维器官**

**信息传递**

**神经系统**

**信息**

**策略**

**信息获取**

**感觉器官**

**信息源**

**信息施效**

**效应器官**

**信息宿**

**对象系统**

**-----------------------**

**外部世界**

**信息**

**行为**

**信息**

**策略**

* 1. **通信系统模型**

从图1.2中可以看出，为了使获取的信息能够进行信息处理，从而形成决策指挥执行机构（信息施效），信息的传递是必不可少的环节。我们把“信息的传递”叫做“通信”。

因此，我们有必要进一步了解通信系统。图1.3是通信系统简化框图。

**信源**

**信道**

**信宿**

噪声源

**信息**

**信息**

**干扰**

**图1.3 通信系统模型**

通信，就是把信息从一个地方（信息的出发点，信源）通过信息的传输通道（信道）传送到另一个地方（目的地，信宿）。在信息的传输过程中，信道不可能是理想的，存在噪声和干扰（噪声源）。

在通信的过程中，我们最关注的是以下三个方面：

* + - 1. 信息传递的**有效性**：用尽可能少的符号，把尽可能多的信息传递到目的地（或者说，在尽可能短的时间内，把尽可能多的信息传送到目的地）；
      2. 信息传递的**可靠性**：信息传递过程中，要尽可能减少差错；
      3. 信息传递的**安全性**：信息传递过程中，要尽可能保密。

**本书只考虑有效性和可靠性问题。**安全性问题可参见有关计算机加密算法等课程。

* 1. **信息论与编码的研究内容**

衡量通信系统质量好坏，主要有**有效性、可靠性和安全性**等指标，这些指标，当然越高越好。问题是，它们能高到多少（极限值是什么）？用什么样的手段能够达到或者接近这些极限值？有关这些问题的解答，就是信息论与编码这门课程的主要内容：**信息论解答性能极限值的问题，而编码解答达到或接近这些性能极限值的手段和途径问题。**

* 1. **本书章节结构**

既然信息论与编码是紧紧围绕通信（信息传递）过程来进行研究和讨论的，那么其内容也就是紧紧围绕通信系统来展开的。从通信系统模型方框图1.3可以看出，通信系统主要包括信源、信道和信宿三部分（干扰可看成信道的一部分），也可以看成发送端（信源端）、信道和接收端（信宿端）三部分。接收端所做的工作，基本上是发送端工作的逆过程（发送端对信息做某种包装以方便传输，而接收端则要拆除包装以获取信息）。

所以，要研究的主要部分，就是：**信源**和**信道**。

因此，本书以下的主要章节安排包括：

第二章：信源及信源熵

首先，对信源进行分类；然后，按照从简单到复杂的顺序，用数学模型表示各种信源，并研究信源的一个重要参量—信源熵；以及围绕信源熵的若干相关内容（互信息量等）。

第三章：信道及信道容量

首先，对信道进行分类；然后，按照从简单到复杂的顺序，用数学模型表示各种信道，并研究信道的一个重要参量—信道容量；以及围绕信道容量的若干相关内容。

可以看出：第二章和第三章是完全平行的，**信源对应信道；信源熵对应信道容量**。

实际上，**信源熵是对应有效性**的重要参数，而**信道容量是对应可靠性**的重要参数。在对信源和信道有了一定的认识（第二章和第三章）以后，接下来的任务，就是用什么手段和途径（编码），来实现高有效性和高可靠性。

因此，就有了：

第五章：信源编码

信源编码用于提高通信系统的有效性。本章在介绍一些基本概念的基础上，引出信源编码定理（有效性的极限值），并介绍一些常用的信源编码方法。

第六章：信道编码

信道编码用于提高通信系统的可靠性。本章在介绍一些基本概念的基础上，引出信道编码定理（可靠性的极限值），并介绍一些常用的信道编码方法。

可以看出：第五章和第六章是完全平行的，**信源编码对应信道编码；有效性对应可靠性**。

除此之外，还要讨论允许一定失真的信源编码（限失真信源编码，压缩编码）问题，讨论失真允许限度和信息率之间的关系（信息率失真函数），这些问题在第四章讨论。

最后，还要对信息论的前沿问题做简单介绍，主要介绍网络信息论的有关概念和知识，是为本书的第七章。

**归纳起来**：本书围绕信源和信道两个内容：第二章（信源及通信的有效性）和第三章（信道及通信的可靠性）研究信源和信道的基本属性，两章是平行的（包括第四章）；而第五章和第六章则在第二章和第三章的基础上，研究实现有效性和可靠性的手段和途径（信源编码和信道编码），两章是平行的。最后在第七章介绍了网络信息论。

另外，在每一章的后面，都附有若干附录。这些附录，或为该章正文中某些结论的严格数学推导（为聚焦基本概念和物理意义，正文中涉及到较繁琐数学推导的地方，大多放在附录中，供有需要的读者参考），或为扩展阅读部分。附录为选读内容，不影响正文的理解。

信息论与编码

有效性

可靠性

信 源

（第二章）

信 源 编 码

（第四、五章）

信 道

（第三章）

信 道 编 码

（第六章）

**本章思路：**信息论（关于信息的理论）🡪信息🡪信息系统组成🡪通信系统组成🡪通信系统评价指标🡪指标的最优值（极限值）（信息论）🡪达到最优值（极限值）的路径和手段（编码）🡪信息论与编码

第二章 信源与信息熵

2.1 信源的分类与数学模型

2.1.1 信源的分类

2.1.2 信源数学模型

2.1.3 马尔科夫信源的稳态分布

2.2离散单符号信源信息熵

2. 2.1自信息

2.2.2离散信源熵

2.2.3 熵的性质

2.2.4 互信息

2.2.5 数据处理定理

2.3离散序列信源信息熵

2.3.1 离散无记忆信源及其信息熵

2.3.2 离散有记忆信源及其信息熵

2.3.2 离散平稳信源的极限熵

2.4 马尔可夫信源的极限熵

2.5 连续单符号信源信息熵

2.5.1 连续单符号信源的差熵

2.5.2 连续单符号信源的最大熵定理

2.6 连续序列信源信息熵

2.7 波形信源信息熵

2.8 信源冗余度

2.1 信源的分类与数学模型

本章我们讨论信源。

2.1.1 信源的分类

信息是无形的。但为使用起见，信息需要表现为某种形式。比如天气预报说明天要下雨，在没有人告诉我们之前，我们不能确定明天是否下雨，是有不确定性的，天气预报可以消除或者减少这种不确定性，因此天气预报包含有信息。但天气预报必须以某种形式呈现，或者是文字，或者是声音，或者是一幅图画。这种携带有信息的文字（或声音、图像等）叫做**信号**。因此，**信息是信号里的内容，信号是信息的载体。**

信源是发出信息的，具体体现为从信源输出某种信号。我们可以根据这些信号的特征来对信源进行分类。

在信号与系统这门课里，我们按信号在时间轴和幅度轴是离散的还是连续的，把信号分成离散信号（时间和幅度均离散）、连续信号（时间和幅度均连续）、时间离散幅度连续信号（抽样信号）和幅度离散时间连续信号四种，通常，我们关注前三种。相应地，信源也可以分为三类：

1. 离散信源：时间离散、取值也离散的信源。比如：一次次地掷骰子，掷出的点数（取值）是离散的（六种里面的某一种），时间轴上也是离散的（一次一次的，每次之间有间隔）。离散信源按时间轴上只有一点还是很多点，又分成两种：
2. 离散单符号信源：在时间轴上只有一点（比如：掷骰子，只掷了一次）；
3. 离散序列信源：在时间轴上有一系列的点（比如：掷骰子掷了很多次）。
4. 时间离散幅度连续信源：时间轴上是离散的、取值是连续的信源。比如：测室内温度，温度的取值是连续的。也可以按时间轴上只有一点还是很多点，分成两种：
5. 连续单符号信源：在时间轴上只有一点（比如：测室温，只测了一次）；
6. 连续序列信源：在时间轴上有一系列的点（比如：测室温测了很多次）。
7. 连续信源：时间轴是连续的，取值也是连续的。比如：从早上八点到晚上八点室温变化情况。为避免跟连续单符号信源或连续序列信源混淆，我们把这种信源称为波形信源。

对于离散序列信源，我们又可根据前后符号取值之间是否有关联，分成：

1. 无记忆离散序列信源：前后符号之间的取值没有关联性。比如：口袋里有100个乒乓球，白色的和黑色的各50个，随机取出一个，记下颜色，再取第2次、第3次…，但每次记下颜色后都把球放回口袋。
2. 有记忆离散序列信源：前后符号之间的取值有关联性。比如：口袋里有100个乒乓球，白色的和黑色的各50个，随机取出一个，记下颜色，再取第2次、第3次…，但每次都不把球放回口袋。

如果记忆长度有限，比如某个符号的取值只和之前的*m*个符号有关，而与更前面的符号无关，我们把这样的信源叫做*m*阶马尔科夫信源。

对连续序列信源，也分为有记忆的和无记忆的两类。

由于马尔科夫信源是一类非常重要的信源，因此我们单独把它作为一类。因此，归纳起来，（基本）按照由简单到复杂的顺序，信源分为以下六类：

1. 离散单符号信源；
2. 离散序列信源；
3. 马尔科夫信源；
4. 连续单符号信源；
5. 连续序列信源；
6. 波形信源。

2.1.2 信源的数学模型

科学研究，不用文学语言，而用数学描述。比如一幅图片，在科学文献中，我们不会说这幅图片美得“此景只应天上有”，不会说“美不胜收”；而可能会说“在被调查的人群众，认为美到100分的占50%，60分以下的占20%”，等等。科学文献不可能出现“洛神赋”。这一方面是因为：科学研究需要精确的描述，而文学语言是模糊的；更重要的是，科学研究需要深入和严谨推理。因此，科学研究需要用数学的方法和语言。只有把一个模糊的社会应用问题变成一个严谨的数学问题，才可以做到这一点。

因此，我们需要把研究对象进行数学描述，这就是所谓的数学模型。

关于信源的研究也需要先建立其数学模型。

信息是用来消除不确定性的（参见第一章），信息量的多少，取决于消除不确定性的多少。不确定性可以用概率来描述。因此，信源的数学模型可以建立为概率空间。

1. 离散单符号信源的数学模型

由于离散单符号信源的取值有不随机性（如掷骰子，骰子的点数可以是1到6之间的任意整数），所以我们可以用离散随机变量X来表示。其可能的取值为*xi*（*i*=1，2，…，*n*），对应的概率分别为*p*(*x1*)，*p*(*x2*)，…，*p*(*xn*)，简记为*p1*，*p2*，…，*pn*.

则其数学模型为如下的概率空间：

 （2-1）

注：以下一般用大写字母X（或大写字母Y、Z等）表示随机变量，用小写字母*xi*表示随机变量X的某一特定取值。

1. 离散序列信源的数学模型

离散序列信源可以看成是一系列离散单符号信源按顺序组合而成。因此，可表示为离散随机矢量**X**，其可能的取值为***xi***（*i*=1，2，…，*t*），对应的概率分别为*p*(***x1***)，*p*(***x2***)，…，*p*(***xt***)，简记为*p1*，*p2*，…，*pt*.

注：以下一般用黑体大写字母**X**（或大写字母**Y**、**Z**等）表示随机矢量，用黑体小写字母***xi***表示随机矢量**X**的某一特定取值。

若序列的长度为L，则。其任一元素X*j*（*j*=1，2，…，L）均为随机变量，该随机变量的可能取值为*xi*. 因此可知，随机矢量**X**可能取值（*i*=1，2，…，*t*）的个数*t*=*n*L.

离散序列信源的数学模型为如下的概率空间：

 （2-2）

其中，

 （2-3）

若为无记忆信源序列，则

 （2-4）

1. 马尔科夫信源的数学模型

马尔科夫信源是记忆长度有限的有记忆信源序列，*m*阶马尔科夫信源，一个符号出现的概率，只和之前的*m*个符号有关，和更前面的符号无关。于是，马尔科夫信源序列可以建模成以下概率空间：

 （2-5）

其中，

 （2-6）

1. 连续单符号信源的数学模型

连续信源由于其取值是连续的，所以有无穷多可能的取值，此时不能用概率来描述其不确定性（因为任一取值的概率都趋于0），而应该用概率密度来表示。

设取值范围在（*a ,b*）的连续随机变量X的概率密度函数为，且满足，则连续单符号信源的概率空间为：

 （2-7）

1. 连续序列信源的数学模型

连续序列信源可以看成是一系列连续单符号信源按顺序组合而成。因此，可表示为连续随机矢量**X**，取值***xi***对应的概率密度为*p*(***xi***)，简记为*pi*.

若序列的长度为L，则，其任一元素X*j*（*j*=1，2，…，L）均为连续随机变量，.

连续序列信源的数学模型为如下的概率空间：

 （2-8）

其中，

 （2-9）

若为无记忆信源序列，则

 （2-10）

1. 波形信源的数学模型

对于限频波形信源，设其最高频率为*fm*，根据奈奎斯特抽样定理，只要抽样频率，则可用其抽样信号无失真恢复原信号。设该波形信源持续时间为*tB*，如果我们取最低抽样频率，，则只需要个采样点即可。

因此，一个代表波形信源的平稳随机过程{*x(t)*}，可以用序列长度的随机矢量来表示即可，其概率空间可用序列长度的连续序列信源的概率空间来描述。

注：根据时频测不准原理，一个信号不可能时间和频率都是有限的，限时信号不可能是限频的，限频信号一定不是限时的。上述限时（持续时间为*tB*）限频（最高频率为*fm*）波形信号理论上是不存在的。但在一定条件下，现实中的信号可以近似为限时限频信号。有关内容本书不再深入讨论。

2.1.3 马尔科夫信源的稳态分布

由于马尔可夫信源是一类非常重要的有记忆序列信源，除了上面讲到的信源概率空间外，我们还需要知道得更多一些。

1. 马尔科夫信源的状态转移概率矩阵

对于*m*阶马尔可夫信源序列，其某一时刻符号出现的概率，与之前的*m*个符号有关，于是需要把当前随机变量和之前的*m*个随机变量一起考虑，需要引入矢量进行分析运算，处理较复杂，有必要寻找简便的处理办法。

当考虑信源序列第*i*时刻的符号X*i*时，其之前的*m*个符号已经出现过，是确定的。如果我们把当前时刻之前的*m*个符号的组合叫做当前时刻的状态，当前时刻符号X*i*的概率，就只和当前时刻的状态有关（相当于一阶马尔可夫信源）。于是，就可以把*m*阶马尔可夫信源的复杂问题，转化成较简单的一阶马尔可夫信源问题。

当下一时刻（第*i*+1时刻）成为当前时刻时，当前时刻的状态，由于第*i*时刻符号的输出而发生了变化。因此，考虑序列信源符号的依次输出，就只需考虑随着符号的依次输出，而导致状态的转移情况。

定义第*i*时刻的状态（由*i*时刻之前的*m*个符号决定）

 （2-11）

共有*Q*=*nm*种不同的状态。

如果已知第*i*时刻的状态，随着第*i*时刻的符号*xi*输出，第*i*+1时刻的状态为：

 （2-12）

于是，描述马尔可夫信源概率空间的条件概率，即为以当前状态为条件的符号条件概率。对于齐次马尔可夫信源（所谓齐次，指该条件概率与时间起点无关。平稳信源的概率分布特性与时间起点无关，而齐次马尔可夫信源只要求条件概率与时间起点无关。所以一般情况下，平稳包含齐次，齐次不一定平稳），该条件概率就等于状态由转移为的概率（一步状态转移概率）：

 （2-13）

由于共有*Q*=*nm*种不同的状态，因此，可用的矩阵来描述一步状态转移概率：

 （2-14）

一步状态转移概率矩阵（简称状态转移概率矩阵）**P**的每一行代表一种原状态，每一列代表一种转移到的新状态。系统从一个状态出发，一定会转移到一个新状态，所以矩阵的每一行元素之和一定等于1. 但有可能有某些状态，无论从什么状态（或者是从某些状态）出发，都不可能转移到这些状态（类似于孤岛），所以每一列元素之和不一定等于1（有可能小于1）。

“系统从状态经过**两步**转移到状态”这个事件，等同于“系统从状态经过**一步**转移到任一状态，再由任一状态**一步**转移到状态”的事件之和，因此，两步转状态移概率可以由一步状态转移概率决定。以此类推，多步状态转移概率均可由一步状态转移概率决定。容易推导得出，*k*步状态转移概率矩阵**P**(*k*)等于一步状态转移概率矩阵**P**的*k*次方，即：

 （2-15）

**例2-1** 如图2-1所示的二进制相对码编码器，初始状态Y1=X1；其余时刻，如输入X=0，则当前时刻输出等于上时刻输出，Y*i*+1=Y*i*；如输入X=1，则当前时刻输出异于上时刻输出，；即。若已知P(X=0)=*p*，P(X=1)=1-*p*=*q*，试写出其转移概率矩阵，并画出状态转移图。

X

**+**

T

Y

图2-1 相对码编码器

**解**：由于当前时刻的输出，取决于当前时刻的输入，以及前一个时刻的输出，所以输出序列是一个一阶马尔可夫链。对于二进制信源，前一个时刻的输出共有两种（0或者1），所以状态共有两种：“0”状态和“1”状态。



即转移概率矩阵为：



其状态转移图为：

*q*

*p*

*q*

*p*

1

0

图2-2 状态转移图

1. 马尔可夫信源的稳态概率分布

如果马尔科夫链满足一定的条件（不可约性和非周期性，参见？？），则马尔科夫链具有遍历性，即不论系统从哪个状态出发，当转移步数足够大时，转移到状态的概率都近似等于某个常数。也就是说，无论系统的初始状态是什么样的，经过足够多步的状态转移后，系统处于某一状态（*j*=1，2，…，*Q*）的概率是一个常数。这种确定的状态分布概率称为稳态分布。

系统处于某一状态的概率，等于从各状态转移到该状态的概率平均，即

 （2-16）

其中，是状态转移前系统处于状态的概率。当系统已经处于稳态时，状态转移前后系统处于状态的概率是一样的，都是.

另外，稳态概率还必须满足：

 （2-17）

式（2-16）和（2-17）一起，决定了马尔可夫信源的稳态概率分布。其中共有*Q*个未知数（*j*=1，2，…，*Q*），但有*Q*+1个方程，因此，式（2-16）中的*Q*个方程不是完全独立的，其中只有*Q*-1个独立方程，与式（2-17）的一个方程一起，使稳态概率分布有唯一解。

**例2-2** 有一个二阶马尔可夫信源，输入符号为二进制，设其符号条件概率如表2-1所示，试写出其状态转移概率矩阵，画出状态转移图，并求出稳态概率分布。

表2-1 符号条件概率*p*(*aj*/*si*)

|  |  |  |
| --- | --- | --- |
| 起始状态 | 符 号 | |
| 0 | 1 |
| *s*1 (00) | 1/2 | 1/2 |
| *s*2 (01) | 1/3 | 2/3 |
| *s*3 (10) | 1/4 | 3/4 |
| *s*4 (11) | 1/5 | 4/5 |

**解：**对二阶马尔可夫链，其状态共有四种，分别为：S=(00，01，10，11)，根据符号条件概率表2-1，可以写出符号条件概率矩阵为：



可得状态转移概率矩阵为



相应的状态转移概率如表2-2所示，状态转移图如图2-3所示。

表2-2 状态条件概率*p*(*sj*/*si*)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 起始状态  (*si*) | 终止状态 (*sj*) | | | |
| *s*1 (00) | *s*2 (01) | *s*3 (10) | *s*4 (11) |
| *s*1 (00) | 1/2 | 1/2 | 0 | 0 |
| *s*2 (01) | 0 | 0 | 1/3 | 2/3 |
| *s*3 (10) | 1/4 | 3/4 | 0 | 0 |
| *s*4 (11) | 0 | 0 | 1/5 | 4/5 |

00

10

01

11

(0)1/2

(0)1/5

(1)2/3

(0)1/4

(1)1/2

(1)4/5

(1)3/4

(0)1/3

图2-3 二阶马尔可夫链状态转移图

设四种状态的稳态概率分别为W1，W2，W3，W4，由式（2-16）和式（2-17）得：



解得稳态状态分布概率为：

.

达到稳态后的符号概率为：



所以



同理可得

.

以上我们按离散和连续两大类，从简到繁，分析了离散单符号信源、离散序列信源、马尔可夫信源、连续单符号信源、连续序列信源、波形信源的数学模型，也对马尔可夫信源的状态描述和稳态分布作了进一步的分析，这是我们后续内容的数学基础。

2.2 离散单符号信源信息熵

上一节我们对信源进行了分类，并分别建立了其数学模型。

从本节开始，我们将以数学模型为基础，以信源的一个重要参数—熵—为中心，对信源进行深入分析。本节讨论最简单的信源—离散单符号信源。

2.2.1 自信息量

公式（1-1）给出了一个具有不确定性的信源*xi*所包含信息量多少的度量，为方便使用起见，我们重写如下：

 （2-18）

我们把*I*(*xi*)叫做信源*xi*的自信息量。

**自信息量的物理含义是**：由于离散单符号信源表示为其概率空间，也就是说，该信源X在某一时刻的取值*xi*可能是*x*1，可能是*x*2，…，也可能是*xn*，分别以概率*p*1、*p*2、…、*pn*取得；或者说，在通信过程发生之前，接收者对该信源符号是有不确定性的，只有当该信源符号通过信道传输到接收端以后，收信者才能够知晓该符号，消除了关于该符号的不确定性；根据“**信息是能使不确定性减少或者消除的东西**”的信息定义，收信者得到了信息；得到信息的多少，等于该符号所包含的全部信息量（假定接收到该符号后完全知晓了该符号，也就完全消除了其不确定性）。我们把这个信息量叫做该符号*xi*的自信息量，其大小等于该符号*xi*出现的概率的倒数再取对数，当取以2为底的对数时，得到的数值单位比特，用*bit*（或*b*）表示。

例如：二进制单符号离散信源，其取值可以为0或者1，设其概率空间为，则符号取值0和符号取值1各包含的自信息量为：





不难看出，自信息量具有以下性质：

（1），确定性事件不包含任何信息量；

（2），出现的概率越小，包含的信息量越大；

（3）非负性：，所以，故；

（4）可加性：如果符号*xi*中包含的信息量，符号*yj*中包含的信息量，则符号*xi*和*yj*同时出现包含的信息量，为两者的自信息量之和（假定两者互相独立）：



这是因为，两者同时出现的概率（联合概率），故有。

2.2.2 离散单符号信源熵

简单来说，离散单符号信源的熵就是其平均不确定性。

离散单符号信源的不同取值，由于其出现的概率不同，故所包含的自信息量也不同。大多数时候，我们更关心该离散单符号信源的整体性质，即各种可能取值的平均自信息量。

若离散单符号信源的概率空间为，其某一取值*xi*的自信息量为，则其平均信息量为：

 （2-19）

注意：上述的平均值是指概率平均，不是算术平均。

由于信息量在数值上等于不确定性（多少信息量就可以消除或者减少多少不确定性），所以，离散单符号信源的平均信息量，在数值上就等于其平均的不确定性。在热物理学中，用“**熵**”来表示分子热运动的混乱程度（或者说不确定程度），借用热物理学熵的概念，我们把信源的平均不确定性叫做该信源的熵，记为H(X)，在数值上等于该信源的平均信息量。因此，离散单符号信源的熵为：

 (2-20)

**例2-3**：设某二进制离散单符号信源X，其概率空间为：



求该信源的熵。

**解：**由信源熵的定义式（2-20）得：



是*p*的函数，用H(*p*)表示。

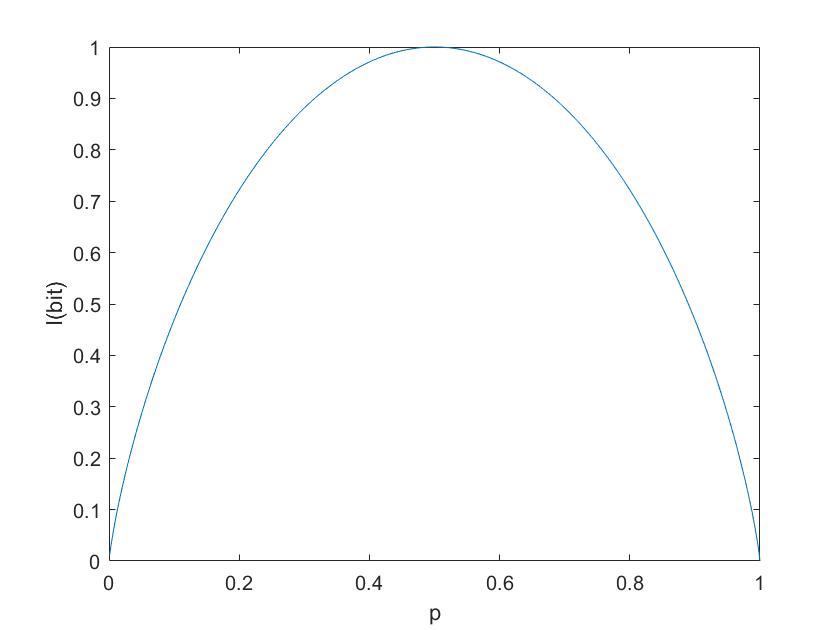


图2-4 熵函数曲线

图2-4绘出了H(X）和*p*的关系曲线。从图中可以看出：

1. 当*p*=0.5时，具有最大熵，此时H(X)=1*bit*；
2. 当*p*=1或*p*=0时，信源熵H(X)=0；
3. 信源熵是*p*的上凸函数。

解释如下：①当*p*=0.5时，两个可能的取值“0”和“1”出现的概率相等，此时信源具有最大的不确定性，因此具有最大熵；②当*p*=1（或*p*=0）时，信源取值为“0”（或为“1”）的概率为100%，取值为“1”（或为“0”）的概率为0，换句话说，信源的取值是确定的（“0”或者“1”，分别对应于*p*=1和*p*=0），没有不确定性，因此信源熵H(X)=0；③上凸性可以证明，本书从略。

容易看出，当*m*进制离散单符号信源的各种可能取值出现的概率相等（为1/*m*）时，信源具有最大的不确定性，此时信源的熵H(X)=*logm*。

2.2.3 离散单符号信源条件熵

如果给定某个条件*yj*，则信源X的概率空间有可能发生变化，相应地，其信息量和熵也可能发生变化。我们把这种情况下的信息量称为其条件信息量、信源熵称为其条件熵，分别记为*I*(*xi/y*j)和H(X/*yj*)，其中*yj*表示条件。则由熵的定义式（2-20），可得条件熵为

 （2-21）

若Y的取值空间为{*yj*，j=1，2，…,*m*}，即Y∈{*yj*，j=1，2，…,*m*}，则当Y取不同的值作条件时，根据式（2-21），就会有不同的条件熵。

我们把以Y为条件X的条件熵定义为式（2-21）的条件熵H（X/*yj*）对各*yj*的概率平均，即：

 （2-22）

条件熵H(X/Y)表示当已知Y后，X仍然具有的不确定度。由于计算X的不确定度时，需要对X的各可能取值*xi*的自信息量做概率平均，需要一次加权求和；Y的各种可能取值*yj*做条件，也要做概率平均，所以又有一次加权求和。因此在式（2-22）中，是二重求和。

同样可得：

 （2-23）

**例2-4：**某二进制数字通信系统如图2-5所示。发送端信源X为二进制信源，其概率空间为。由于通信信道中有干扰和噪声，导致接收端判决结果除了“0”和“1”以外，还有一种未知的状态（既不是“0”也不是“1”），我们表示为“？”状态。设信道的符号转移概率为：

*p*(*y=*0/*x=*0)=3/4， *p*(*y*=0/*x*=1)=0；

*p*(*y=*1/*x=*0)=0， *p*(*y*=1/*x*=1)=1/2；

*p*(*y=*?/*x=*0)=1/4， *p*(*y*=?/*x*=1)=1/2；

求信源熵H(X)、条件熵H(X/Y)和H(Y/X)， 以及信源熵H(Y)。

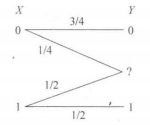


图2-5

**解：**（1）信源熵H(X)可直接由公式求得



（2）根据条件熵公式，求条件熵H(Y/X)，，需要知道条件概率*p*(*y =*0/ *x =*0)、*p*(*y =*1/*x =*0)、*p*(*y =*？/*x =*0)和*p*(*y =*0/*x =*1)、*p*(*y =*1/*x =*1)、*p*(*y =*？/*x =*1)；以及联合概率*p*(*x =*0， *y =*0)、*p*(*x =*0， *y =*1)、*p*(*x =*0， *y =*？)和*p*(*x =*1， *y =*0)、*p*(*x =*1， *y =*1)、*p*(*x =*1 *，y =*？)等：

*p*(*x =*0， *y =*0)= *p*(*x =*0) *p*(*y =*0/ *x =*0)=1/2

*p*(*x =*0， *y =*1)= *p*(*x =*0) *p*(*y =*1/ *x =*0)=0

*p*(*x =*0， *y =*?)= *p*(*x =*0) *p*(*y =*?/ *x =*0)=1/6

*p*(*x =*1， *y =*0)= *p*(*x =*1) *p*(*y =*0/ *x =*1)=0

*p*(*x =*1， *y =*1)= *p*(*x =*1) *p*(*y =*1/ *x =*1)=1/6

*p*(*x =*1， *y =*?)= *p*(*x =*1) *p*(*y =*?/ *x =*1)=1/6

则由条件熵公式可得：



（3）根据条件熵公式，求条件熵H(X/Y)，，需要知道条件概率*p*(*x =*0/*y =*0)、*p*(*x =*0/*y =*1)、*p* ( *x =*0/*y =*？)和*p*(*x =*1/*y =*0)、*p*(*x =*1/*y =*1)、*p*(*x =*1/*y =*？)；以及联合概率*p*(*x =*0， *y =*0)、*p*(*x =*0， *y =*1)、*p*(*x =*0， *y =*？)和*p*(*x =*1， *y =*0)、*p*(*x =*1， *y =*1)、*p*(*x =*1 *，y =*？)等：







故













因此



1. 由







得：



2.2.4 离散单符号信源联合熵

当两个离散单符号X和Y联合出现时，若，，则其概率空间为

 （2-24）

联合自信息量为

 （2-25）

联合熵为

 （2-26）

**例2-5**：例2-4中的信源X和信源Y，求联合熵H(X,Y)。

**解：**根据例2-4的求解过程，已知

*p*(*x =*0， *y =*0)= *p*(*x =*0) *p*(*y =*0/ *x =*0)=1/2

*p*(*x =*0， *y =*1)= *p*(*x =*0) *p*(*y =*1/ *x =*0)=0

*p*(*x =*0， *y =*?)= *p*(*x =*0) *p*(*y =*?/ *x =*0)=1/6

*p*(*x =*1， *y =*0)= *p*(*x =*1) *p*(*y =*0/ *x =*1)=0

*p*(*x =*1， *y =*1)= *p*(*x =*1) *p*(*y =*1/ *x =*1)=1/6

*p*(*x =*1， *y =*?)= *p*(*x =*1) *p*(*y =*?/ *x =*1)=1/6

所以

=1.8*bit*

2.2.5 互信息

从**例2-4**可以看出，H(X)=0.92*bit*，H(X/Y)=0.33*bit*。也就是说，给定Y（或者说如果知道了Y），使得X的平均不确定性由0.92比特减少到0.33比特，减少了0.59比特。为什么？是什么原因使得X的不确定性减少了？

实际上，图2-5所示的是一个数字通信系统。在通信之前（发送端发送信源X之前），接收端对发送端要发送的符号X具有不确定性，平均不确定性就是信源熵H(X)。当在信道的发送端发送信源符号X后，在信道的接收端收到了符号Y。由于接收端符号Y和发送端的符号X是有某种相关性的（X和Y之间的相关性由信道转移概率*p*(*yj/xi*）确定，若*p*(*yj/xi*）=1，则说明*xi*和 *yj*完全相关)，因此，由于Y的已知，使得X的不确定性减少到H(X/Y)，减少了H(X)-H(X/Y)。换句话说：**Y里面包含有部分（或全部）X的信息，或者说，从Y里面可以得到部分X的信息**。同样地，已知X，也可以使Y的不确定性由H(Y)减少到H(Y/X)，减少了H(Y)-H(Y/X)。可以证明，H(X)-H(X/Y)= H(Y)-H(Y/X)，也就是说，Y里面包含的有关X的信息，等于X里面包含的有关Y的信息。我们把这种X和Y互相包含有的对方的信息，叫做X和Y之间的互信息，记为I（X；Y）。

显然，

 （2-27）

式（2-27）也可以从另一个角度得到：

设发送端发送的符号为*xi*，接收端收到的符号为*yj*，则*xi*和*yj*之间的互信息量为：



即：*xi*和*yj*之间的互信息量（*yj*里所包含的有关*xi*的信息量），等于已知*yj*这个事件所导致的有关*xi*不确定性的减少量，也就是*xi*原来具有的不确定性（数值上等于其自信息量I(*xi*)），减去当*yj*已知后，*xi*仍然具有的不确定性（数值上等于其条件信息量I(*xi*、*yj*)）。

如果在X取值集合上做概率统计平均，即得：



如果进一步在Y取值集合上做概率统计平均，即得：



式（2-28）可以看作是互信息量的定义式。

而且：

 （2-29）

上式推导中，用到了。

**例2-6：**求**例2-4**中，符号X和符号Y之间的互信息。

**解：**由互信息定义式（2-27）得：

符号X和符号Y之间的互信息

=0.59*bit*

2.2.6 信源熵、条件熵、联合熵、互信息等之间的相互关系

很容易理解，关于信源的熵、条件熵、联合熵、以及互信息量之间，有以下关系：

1. 信源熵、条件熵、联合熵之间的关系：

 （2-30）

因为不确定性具有可加性，两个符号X和Y的联合熵（联合不确定性），是两个符号不确定性之和，只是，在考虑第二个符号的不确定性时，第一个符号已经考虑过了，也就是说，第二个符号的不确定性，是以第一个符号已经确定为前提的，因此加上的是条件熵。另外，由于加法满足交换律，就有了式（2-28）的两个等式。

一种特殊情况是：如果信源X和信源Y是相互独立的，则

 （2-30.1）

上述关系式也可以由各量的定义式出发得到证明。

1. 信源熵、条件熵、联合熵、互信息之间的关系

 （2-31）

式（2-31）可由式（2-30）代入式（2-29）得到。

式（2-31）中各量之间的关系可由图2-6形象地表示出来。

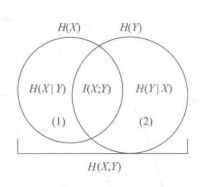


图2-6

图2-6中，左边的圆代表信源X的不确定性H（X），右边的圆代表信源Y的不确定性H（Y），中间的重叠部分，表示已知X后（已知左边圆的部分），能知道的有关Y的信息（右边圆的一部分），即为X和Y之间的互信息量。

将上面一段话中的“X”和“Y”互换，也一样成立。所以X和Y之间的互信息量，是X中包含的有关Y的信息量，也是Y中包含的有关X的信息量。表示的是X和Y之间的相关程度，在图中即为两个圆重叠的程度。

两个圆的外边界（即形如的边界）所包括的区域，表示X和Y合在一起的不确定性，即为H（X，Y）。

也可以用图2-7描述各量之间的关系。

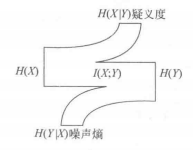


图2-7

在图2-7中，左边为发送端，带状区域的宽度表示发送信源熵H（X）；右边为接收端，带状区域的宽度表示接收信号信源熵H（Y）；从左往右的中间部分表示信息在信道中的传输过程。

如果信道是理想的，完全没有干扰和噪声，则发送和接收信号取值是一一对应的，表现在图中，就是H（X）和H（Y）是对齐的。接收端知道了接收信号Y，就可以完全知道发送信号X，因此，接收信号Y里包含有发送信号X的全部信息，即互信息量I（X；Y）=H（X）=H（Y），如图2-7.1.

I（X；Y）

H（X）

H（Y）

图2-7.1

如果信道中有干扰和噪声，由于干扰或噪声的影响，使得信道的输出和输入（Y和X）不是一一对应的，而是按一定的概率关联，如**例2-4**中的图2-5. 在图2-7中，就表现为H（X）和H（Y）不是对齐的，其公共部分，表示已知Y（或已知X）以后，能够知道的关于X（或者Y）的信息，即为互信息量I（X；Y）。Y的不确定性（知道Y后能够得到的信息量，H（Y）），除了由X传送过来的一部分不确定性（I（X；Y））引起，还有一部分是由于干扰和噪声的不确定性（H（Y/X））引起，因此，H（Y/X）称为噪声熵。

相似地，X的不确定性（X里面所包含的信息量，H（X）），除了一部分传送到接收端（互信息量，I（X；Y）），还有一部分因为信道的非理想性而丢失了，因此，在接收端，即使已经知道了Y，对X仍然有一定的不确定性，这部分不确定性叫做疑义度，是已知Y以后X仍然具有的不确定性，为H（X/Y）。

所以，与图2-6相比，图2-7不但表达了信源熵、条件熵、联合熵、互信息之间的关系，也清楚表明了有噪信道下通信系统的信息传输情况，对理解通信系统的信息传递，具有十分重要的作用。

2.2.6 信源熵的性质

这里仅给出结论，有关证明请参见附录。

（1）非负性

对信源X的所有可能概率分布P={*p1*，*p2*，…，*pn*}，都有

 (2-32)

信源熵描述信源的不确定性，不确定性不可能为负。

（2）对称性

当概率矢量P={*p1*，*p2*，…，*pn*}中的元素顺序任意互换时，熵函数H(P)的值不变，即

 (2-33)

信源熵是信源总体不确定性的描述，仅与信源的总体统计特性(含有的消息数和概率分布)有关，不同信源，只要其总体统计特性相同，信源熵就相同。

（3）确定性

概率矢量P={*p1*，*p2*，…，*pn*}中的任一概率分量*pi*=1时，其余概率分量必全部为零，此时：

 （2-34）

即：如果信源发出某个消息或符号的概率为1，而发出其它消息或符号的概率为0，则信源熵为零。这样的信源被称为确定信源，其熵等于零。

信源熵的确定性很容易理解：对确定性信源，其发出某一个特定取值的概率为100%，所以不存在任何不确定性，因此熵为零。

（4）可扩展性

 （2-35）

该性质说明，增加信源的取值个数，若这些增加的取值出现的概率很小，则信源熵不变。这是因为，虽然这些小概率取值的自信息量很大，但从总体来考虑时，它在熵的计算中起的作用很小。这也进一步说明熵是信源总体平均不确定性的体现，它由信源中所有符号的概率和不确定性共同决定。

（5）递增性

设信源X的概率分布为P={*p1*，*p2*，…，*pn*}，若将其某一取值（比如*xn*）拆分为*m*个不同的取值{ *xn1*，*xn2*，…，*xnm* }，其对应的概率分别为{ *q1*，*q2*，…，*qm* }，它们共享*xn*的概率*pn*，即，则其熵会增加，满足：

 （2-36）

熵增加是因为：信源X中某一元素被分割成*m*个元素后，X的取值可能性变多，增加了信源的不确定性。

（6）上凸性

熵函数H(P)是概率矢量P={*p1*，*p2*，…，*pn*}的严格∩型凸函数（上凸函数）。因为熵函数具有上凸性，所以熵函数具有极大值。

（7）极值性

 （2-37）

式（2-37）表明，离散信源*n*个可能的取值等概率分布时熵最大。

**最大离散熵定理：**对于具有*n*个可能取值的离散无记忆单符号信源，只有在*n*个可能取值等概率出现的情况下，信源熵才能达到最大值。即

 （2-38）

极值性是上凸性的自然结果。

2.3 离散序列信源信息熵

除极少数情况外，实际信源很少会是单符号，而会是符号序列。如离散单符号信源用随机变量X表示，则离散序列信源要用随机矢量**X**表示：**X**={X1X2…XL}，其中，L表示序列的长度，X*i*为一个离散单符号随机变量。（黑体符号表示矢量，请注意其与非黑体符号所表示变量之不同）

根据前述描述离散序列信源的数学模型式（2-2）、式（2-3）和式（2-4），可得离散序列信源的信息熵等。

（2-39）

2.3.1 离散无记忆信源及其信息熵

若随机矢量**X**中各时刻的随机变量X*i*统计独立，则由式（2-4）和式（2-39）得：

即： （2-40）

式（2-40）看起来好像推导很复杂，实际上其结论和物理意义非常简洁：**离散无记忆序列信源的熵，等于组成该序列的各离散单符号信源熵之和。**这是很容易理解的，因为根据信源联合熵的性质式（2-30.1），可以将离散序列信源看成是各离散单符号信源的集合，若这些符号之间统计独立，则其联合熵为各符号信源熵之和。

若信源序列同时满足平稳性，即信源的特性在各时刻是相同的，与序号*l*无关，则各时刻的单符号信源熵相同，，故

 （2-41）

平均符号熵（即平均到每个单符号上的熵值）为

 （2-42）

2.3.2 离散有记忆信源及其信息熵

若随机矢量**X**中各时刻的随机变量X*i*不是统计独立的，则由式（2-3）和式（2-39），与式（2-40）的推导过程相似，可以得到：

 （2-43）

2.3.3 离散平稳信源的极限熵

在式（2-42）中，我们想知道，如果，有极值吗？如果有，极值等于什么？

**结论1**：离散序列信源的极限熵为（证明见附录？？）

 （2-44）

**结论2**：H（XL/XL-1）是L的单调非增函数。

其中，XL-1表示X1X2…XL-1的序列。

**结论3**：。

**结论4**：是L的单调非增函数。

以上结论，其物理意义相当明显，容易理解。

对结论2：随着L的增加，H（XL/XL-1）中的条件序列XL-1中的条件增多，而XL的特性并不随L改变（对平稳序列）。增加了条件，会使不确定性减小（如果增加的条件与XL统计独立，则增加条件对XL的熵没有影响），故随着L增加，H（XL/XL-1）会变小，或者保持不变。因此，H（XL/XL-1）是L的单调非增函数。

对结论3：观察式（2-43）的右边，根据结论2，连加式的每一项是单调非增的，由的定义式（2-42）可知，是各项的平均值，介于最大值（第一项）和最小值（最后一项）之间，故有本结论。

对结论4：根据对结论3的分析，式（2-43）右边的项数越多，越往后的项其值就会变得越小，导致均值越小。故有本结论。

最后，对结论1：式（2-43）右边各项，是一个非增序列。随着L的增加，XL序列越来越长， XL与XL序列中前面的一些符号（如X1X2等）之间的相关性越来越小，所以，非增序列后面的一些项数值的变化越来越小，几乎趋近于一个极限值。当，在求其均值时，前面有限项较大的数值，对均值的影响可以忽略，因此，即为那个序列的极限值，即：。

**例2-7**：某离散有记忆信源，其每一单符号的概率空间为



设信源发出L=2的符号序列，记为；记忆特性可用条件概率*p*（*xj*/*xi*）表/示，如表2-3所示。

表2-3 条件概率

|  |  |  |  |
| --- | --- | --- | --- |
| *xj* | *xi* | | |
| *x1* | *x2* | *x3* |
| *x1* | 9/11 | 1/8 | 0 |
| *x2* | 2/11 | 3/4 | 2/9 |
| *x3* | 0 | 1/8 | 7/9 |

**注**：由表2-3可以看出：每一列元素之和为1；但每一行元素之和不一定等于1。其物理意义是：对任一前一符号*xi*，其后面发生的符号一定是*x1*、*x2*或*x3*中的一个；但即使前一符号遍历所有可能（*x1*、*x2*或*x3*），其后续符号也有可能不会出现某一特性符号*xj*（其概率不一定等于1）。

**试求：**信源序列熵H（**X**）和平均符号熵H2（**X**）。

**解：**

（1）信源序列熵

由式（2-43）可知：

H（**X**）=H（X1）+H（X2/X1）

① 计算单符号熵H（X1）

由单符号概率空间可得：



② 计算条件熵H（X2/X1）

由条件熵定义式（2-23），计算条件熵H（X2/X1）需要知道条件概率*p*（*xj*/*xi*）和联合概率*p*（*xi xj*）。

表2-3给出了条件概率*p*（*xj*/*xi*），根据条件概率与联合概率的关系式，可计算出联合概率，如表2-4所示。

表2-4 联合概率

|  |  |  |  |
| --- | --- | --- | --- |
| *xj* | *xi* | | |
| *x1* | *x2* | *x3* |
| *x1* | 1/4 | 1/18 | 0 |
| *x2* | 1/18 | 1/4 | 1/18 |
| *x3* | 0 | 1/18 | 7/36 |

条件熵



③信源序列熵为

H（**X**）=H（X1）+H（X2/X1）=1.5+0.87=2.37*bit*/序列

（2）平均符号熵



可见，H2（**X**）<H（X1），即L=2的有记忆信源序列的平均符号熵小于单符号无记忆信源的熵，这是由于符号之间存在相关性，使得符号的不确定性减小所致。

2.4 马尔可夫信源的极限熵

根据极限熵的公式（2-44）可得：

（2-45）

根据条件熵的定义式（2-22）可知：



对于具有稳态分布的马尔可夫链（见2.1.3）：

（2-46）

可见，马尔可夫信源的极限熵，是其状态条件熵H(X/*si*)对全部可能状态做统计平均得到的。

**例2-8**：图2-8所示的三状态马尔可夫信源，写出其状态转移矩阵，画出其状态转移图，求出稳态概率分布，并求其极限熵。



图2-8三状态马尔可夫信源状态转移图

解：由图2-8可以写出其状态转移矩阵为



设稳态概率分布，则由

**WP**=**W**和

可得：



在状态下每输出一个符号的平均信息量为







对3个状态取统计平均，即可得马尔可夫信源熵为：



总结一下：求解马尔可夫信源熵的步骤为：

1. 根据题意画出状态转移图，或写出状态转移矩阵；
2. 计算信源的平稳分布概率（假定信源具有平稳分布）；
3. 根据一步转移概率和稳态分布概率，计算信源熵（极限熵）。

2.5 连续单符号信源熵

幅度连续单符号，可以看作是离散单符号幅度取值无限多无限密集的极限情况。因此，我们可以从离散单符号信源熵出发，通过取极限来推导连续单符号信源熵。

设某信源变量X为在[a，b]区间连续取值的单符号连续信源，其概率空间为：



其中，*p*X(*x*)是随机变量X的概率密度分布函数。

将取值区间[a，b]分成*n*等份的小区间，每个区间长度，*xi*是第*i*个区间中的一点，即：。第*i*个区间的取值概率*p*(*xi*)为其概率密度在区间上的积分：



由积分中值定理可得：



经由以上处理，则连续单符号信源转化为离散单符号信源：



由离散单符号信源熵的定义可得信源Xn的熵为：



若，则，：



上式中的第二项为无穷大，这是可以理解的：因为连续信源有无穷多种可能的取值，其不确定性为无穷大。

但是，在很多情况下，我们并不在意熵的值有多大，而是更在意两个熵的差值。比如通信系统更关注互信息量，而互信息量是两个熵之差。

因此，我们可以不必在意上式中的第二项（因为熵差的时候，这一项被抵消掉了），而仅仅关注第一项。

我们定义**连续信源熵**为：

 （2-47）

可以看出，连续信源熵式（2-47）与离散信源熵非常相似，只不过把离散情况下的求和变成了积分。

与离散单符号信源情形类似，也可以定义相应的连续信源条件熵和联合熵：

 （2-48）

 （2-49）

2.6 连续序列信源熵

连续序列的熵和离散序列的熵性质完全相同，只要把离散信源熵用连续信源熵替代就可以了，式（2-40）--式（2-44）同样满足。

2.7 波形信源熵

波形信源可用随机过程*x*(*t*)表示。对限时*t*B和限频*fm*平稳随机过程，可以抽样成序列长度L=2*fmt*B的连续序列信源，所以其熵为：

 （2-50）

2.8 连续信源最大熵定理

对于离散单符号信源，等概率分布时，具有最大熵。

连续信源会如何？

可以证明，连续信源不存在如离散信源那样的无条件最大熵。如附加约束条件，则可以有最大熵，不同的约束条件，可能有不同的最大熵。

1. 限峰值功率最大熵定理：

对如下概率空间描述的连续信源：



若其定义域[*a*，*b*]为有限值，意味着其峰值功率是受限的（功率正比于幅度值的平方）。此时有如下最大熵定理：

随机矢量**X**的幅度取值限制在[*a*，*b*]时，当其概率密度分布符合均匀分布条件



时，信源达到最大熵，最大熵为：

 （2-51）

该最大熵定理与离散单符号信源最大熵定理类似。

1. 限平均功率最大熵定理：

对相关矩阵（可理解为平均功率）一定的随机矢量**X**，当其概率密度分布满足正态分布



时，具有最大熵。最大熵为：

 （2-52）

2.9 信源的冗余度

定义信息效率：

 （2-53）

则冗余度定义为：

 （2-54）

**本章小结：**

1. 信源分类：

（1）离散单符号信源

（2）离散序列信源

（3）马尔可夫信源

（4）连续单符号信源

（5）连续序列信源

（6）波形信源

2. 信源的数学模型

概率空间、概率密度空间

3. 离散单符号信源熵

（1）自信息量

（2）信源熵

（3）条件熵

（4）联合熵

（5）互信息量

（6）各量之间的关系

4. 离散序列信源的熵

（1）序列熵

（2）平均符号熵

（3）极限熵

5. 马尔可夫信源极限熵

6. 连续单符号信源熵

（1）连续熵

（2）连续条件熵

（3）连续联合熵

（4）连续单符号信源最大熵定理

7. 连续序列信源熵

8. 波形信源熵

9. 信源冗余度

第三章 信道与信道容量

3.1 信道的分类与数学模型

3.1.1 信道的分类

3.1.2 信道数学模型

3.1.3 信道容量的定义

3.2离散单符号信道容量

3. 2.1对称离散无记忆信道容量

3.2.2准对称离散无记忆信道容量

2.2.3 一般离散无记忆信道容量

3.3离散序列信道容量

3.3.1 离散无记忆序列信道情形

3.3.2 离散无记忆输入情形

3.3.2 扩展信道

3.4 连续单符号信道容量

3.4.1 连续单符号高斯加性噪声信道容量

3.4.2 连续单符号一般加性噪声信道容量

3.5 连续序列信道容量

3.5.1 多维无记忆加性连续信道

3.5.2 注水算法

3.6 波形信道容量

3.6.1 波形信道容量--香农公式

3.6.2 有关香农公式的讨论

3.7 信道冗余度

3.8 多输入多输出（MIMO）信道

附录1：信源熵性质证明

附录2：最大熵定理证明

3.1 信道的分类与数学模型

和第二章研究信源类似，对信道的研究，也按照分门别类、由简到繁的思路进行。所以要先对信道进行分类，并用数学模型进行描述。

3.1.1 信道的分类

角度不同，分类就不同，对信道的分类同样如此。比如可以按信道是否有线分为有线信道和无线信道；可以根据信道有无从输出到输入的反馈分为无反馈信道和有反馈信道等等。

为研究的方便，并与第二章信源分类相衔接，我们按照信道输入信号（信源）的不同，来对信道进行分类。因此，可以把信道分为：

1. 离散单符号信道；
2. 离散序列信道；
3. 连续单符号信道；
4. 连续序列信道；
5. 波形信道。

由于本教材并不研究有记忆信道，所以并不考虑输入是马尔可夫信源的情形。

3.1.2 信道的数学模型

可以用多种方法对信道进行建模。比如在无线信道情形，可以用射线追踪法按照电磁传播理论对信道进行建模，等等。

按照通信系统的模型方框图1.3，我们可以把信道看作是连接输入（信源）和输出的一个管道。管道的作用，就是要尽可能把特定输出符号和特定输入符号对应起来，以便在接收端可以根据特定输出符号判断出发送端通过信道传输的是哪个特定输入符号，从而达到通信的目的。

因此，信道把特定输入符号映射为特定输出符号的能力（正确概率）以及映射为其他输出符号的可能性（错误概率），就可以非常清楚地描述信道的特性。我们把这种输入符号和输出符号之间的依赖关系，叫做**转移概率**，如图3.1所示。描述这种依赖关系的数学方式，就是给出其转移概率P(Y/X)。

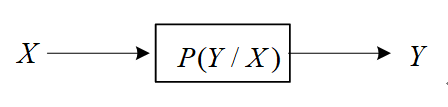


图3.1 信道模型图

1. 离散单符号信道的数学模型

离散单符号信道的输入信号为随机变量X，，用概率空间来描述。设输出随机变量为Y，。则信道模型可以描述为输入输出之间的转移概率矩阵：

 （3-1）

其中，表示在输入端符号为*ai*的条件下，接收端收到符号*bj*的概率。

转移概率矩阵P的每一行元素之和一定为1，因为当输入为某个符号*ai*时，总会在输出端出现某个符号*bj*（*j*=1,2,…,*m*）;而每列元素之和不一定为1，因为即使输入端遍历所有的输入符号*ai*（*i*=1,2,…,*n*），输出端也不能确定一定会出现某个特定输出符号*bj.*

**例3-1**：如图3.2所示的二进制对称信道（BSC），试写出其转移概率矩阵。



图3.2 二进制对称信道（BSC）

**解**：输入输出皆为二进制符号，故转移概率矩阵为2行2列。由图可知：

*p*11=1-*p*，*p*12=*p*，*p*21=*p*，*p*22=1-*p*

故转移概率矩阵为：



1. 离散序列信道的数学模型

当输入为离散序列时，用随机矢量**X**来表示，若输出随机矢量为**Y**，则信道可用转移概率P(**Y/X**）表示。

本章仅考虑无记忆信道。若信源也是无记忆的，则离散序列信道可看成是一系列离散单符号信道。

1. 连续单符号信道的数学模型

对连续单符号输入信源*x*，对应的输出为随机变量*y*，信道用转移概率密度PY(*y/x*)描述。若信道为加性高斯白噪声信道，即：

*y*=*x*+*n* （3-2）

且*n*为0均值、方差为的高斯随机变量，则当给定输入*x*为*x*0后，输出*y*是一个均值为*x*0、方差为的高斯随机变量：

 （3-3）

如给定输入X为离散单符号*a*0，由于噪声*n*是连续高斯随机变量，输出*y*是一个均值为*a*0、方差为的连续高斯随机变量：

 （3-4）

1. 连续序列信道的数学模型

当输入为连续序列时，用随机矢量***x***来表示，若输出随机矢量为***y***，则信道可用转移概率密度PY(***y/x***）表示。若输入序列无记忆，无记忆连续序列信道可以看作是一系列连续单符号信道。

1. 波形信道的数学模型

若输入为波形信源，用随机过程{*x*(*t*)}表示，则输出也为随机过程，记为{*y*(*t*)}。若满足限时（*t*B）限频（*fm*）条件，则可以抽样成L=2*fmt*B的连续平稳随机序列，则信道用转移概率密度描述，为：

 （3-5）

若信源信道皆无记忆，式（3-5）可以写成：

 （3-6）

即此时波形信道转化为多维连续单符号信道。

对于多维连续加性噪声信道，输入随机矢量***x***、输出随机矢量***y***、噪声矢量***n***之间的关系是：***y***=***x***+***n***，此时：



考虑到噪声***n***和信号***x***是相互独立的，则

 （3-7）

3.1.3 信道容量的定义

信道的作用是传递信息。我们希望在信道中每传送一个发送端符号X，接收端符号Y里能包含尽可能多有关X的信息量，即互信息量I(X;Y)要尽可能大。那么我们要问：互信息量I(X;Y)和什么有关？最大可以达到多大？

1. 互信息量表达式

由互信息量的定义可知：



即：

 （3-8）

从式（3-8）可以看出，互信息量只是输入信源的概率分布和信道模型的函数。

**定理3.1：**在给定时，互信息量I(X;Y)是的上凸函数。

上述定理说明，当信道确定时（给定），互信息量仅是输入信源X的概率分布的函数，当取特定值时，互信息量有最大值，也就是说，这时候发送端每发送一个符号，接收端可以从接收符号Y中得到最多关于发送符号X的信息量。

通信的目的，当然希望每次信道使用（每发送一个符号），传送到接收端尽可能多的信息量（互信息量取最大值）。现在的问题是：

① 该最大值是多少？

② 是什么分布时，取得该最大值？

1. 信道容量的定义

信道容量用C表示（英文单词Capacity的首字母），定义为：

 （3-9）

其物理含义是指信道的最大传输能力，即信道给定时，信道中每传送一个符号接收端能够得到的最大信息量。C的单位是“比特/符号”。

从式（3-8）可知，互信息量是信源概率分布和信道特性的函数，当信道给定时，信道特性就已经确定，此时互信息量的大小仅仅取决于输入信源的概率分布。当信源满足某一特定概率分布时，互信息量达到最大值（信道容量）。这样的信源叫做该信道的匹配信源。

也可用单位时间信道传输的最大信息量来表示信道容量，记为C*t*：

C*t*=C/T (3-10)

T是每传送一个符号所需的时间（符号周期T，单位：秒）。C*t*的单位是“比特/秒（bit/s）”。

信道容量要研究的问题是：信道容量有多大？匹配信源的概率分布是什么样的？

3.2 离散单符号信道的信道容量

即使是离散单符号这样简单的信道，其信道容量也是不容易计算得到的。为此，我们进一步把离散单符号信道按照从简单到复杂的顺序，分成以下三类进行进一步的分析。

1. 对称离散无记忆信道（Discrete Memoryless Channel--DMC ）

为进一步简化分析，我们对离散单符号信道附加以下约束：

① 转移概率矩阵P的每一行元素都相同，但位置顺序可以不同（称为每一行都是第一行的置换）。该转移概率矩阵称为输入对称的；

② 转移概率矩阵P的每一列元素都相同，但位置顺序可以不同（称为每一列都是第一列的置换）。该转移概率矩阵称为输出对称的。

输入输出都对称的信道，称为对称信道。

如果离散信道是对称信道，且无记忆，则称为对称离散无记忆信道。

下面我们来推导对称离散无记忆单符号信道的信道容量表达式。

根据信道容量的定义，有：



而



其中，



对于对称离散无记忆信道，由于其每一行皆为第一行的置换，故与行号*i*无关，记为：



因此：



该条件熵与信源X的概率分布无关。

由此可以得到：

 （3-11）

即对称DMC信道容量等于输出随机变量Y（通过改变输入信源X的概率分布得到）的最大熵减去转移概率矩阵中任意一行元素的熵值。

输出随机变量Y为离散单符号信源，。

如果没有其他的约束，根据第二章的知识，其最大熵为*logm*，并且在等概率的情况下，即*p*(*bj*)*=1/m*的情况下取得最大熵。

现在的问题是，式（3-11）中Y的最大熵，可以改变的参数是信源X的概率分布，如果通过改变信源X的概率分布无法使随机变量Y达到等概率，则无法取得最大值*logm*.

但对于对称DMC，如果输入信源X是等概率分布的，则输出随机变量Y一定也是等概率分布的：



由于对称DMC的列对称性，对任意列号*j*，其值都是相同的，故



即当输入随机变量X等概率分布时，输出随机变量Y也是等概率分布的，因此

 （3-12）

所以：

 （3-13）

**例3-2**：设某离散单符号信道的转移概率矩阵为：



试求其信道容量。

**解**：分析该转移概率矩阵可知，这是一个对称信道。按照信道容量公式（3-13），可得：



**例3-3**：某信道的转移概率矩阵为



求信道容量。

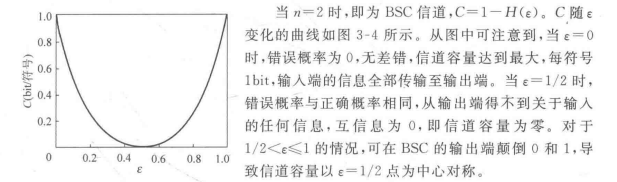
**解**：该信道输入输出符号取值个数都是*n*，正确（输入*ai*输出*bj，i=j*）的概率是1-ℇ，错误（输入*ai*输出*bj，i≠j*）概率ℇ被均匀地分给*n*-1个错误类型，该信道被称为强对称信道。

由式（3-13）得：

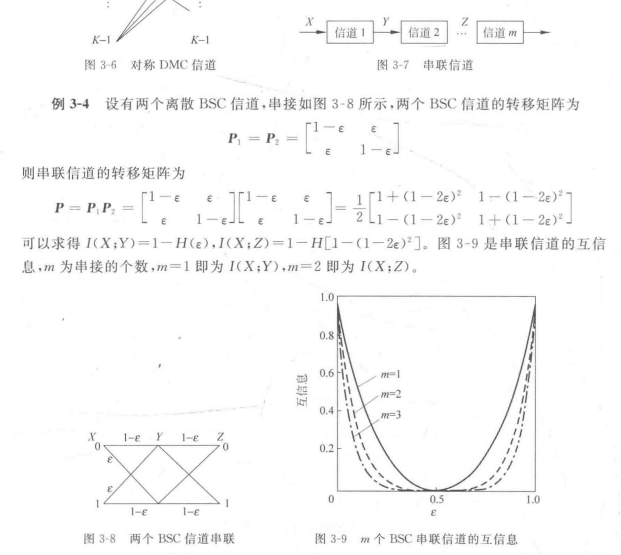


其中，*logn*为输入符号X可携带的最大信息量（最大熵），是由于信道非理想损失的信息量。

**例3-4**：BSC信道的转移概率矩阵为



**例3-5**：串联信道



1. 准对称离散无记忆信道

若信道的转移概率矩阵输入对称而输出不对称，即每一行皆为第一行的置换（包含相同的元素，但位置顺序可能不同），而每一列不是第一列的置换，这样的信道称为准对称离散无记忆信道。

对于准对称离散无记忆信道，即使输入信源X是等概率分布的，



由于不是列对称的，所以输出Y不是等概率的。因此，对于准对称离散无记忆信道，不一定能使得输出等概率， 。

可以有三种方法计算其信道容量。

① 求极值方法

由定理3.1可知，互信息量I(X;Y)是输入信源概率分布的上凸函数，因此，可以用求极值的方法，求得最大互信息量（信道容量）及其对应的概率分布。

**例3-6**：若一信道的转移概率矩阵为



求信道容量。

**解**：矩阵为输入对称、输出不对称的准对称矩阵，2行3列，说明输入符号取值有两种，输出符号取值有三种，记输入符号可能取值为*a1*和*a2*, 输出符号可能取值为*b1*、*b2*和*b3*,设，则，则有







故







因为，所以应有：



可得：

此时，I(X;Y)达到最大值，为C=0.036*bit*/符号，，

② 基于矩阵分解的公式法

将准对称DMC的转移概率矩阵分解为*r*个对称子矩阵，则准对称DMC的信道容量为：

 （3-14）

其中，为第*k*个对称子矩阵一行元素之和，为第*k*个对称子矩阵一列元素之和。

**例3-7**：用矩阵分解法求例3-6中信道的容量。

**解**：对转移概率矩阵分解，可得到两个对称矩阵，分别为：和，故：



**例3-8**：某信道的转移概率矩阵为



试求其信道容量。

**解**：矩阵行对称，列不对称，是准对称矩阵。

利用矩阵分解法，可把其分解为：、和三个子矩阵，于是



③ 观察法

**例3-9**：如例3-6，信道的转移概率矩阵为



求信道容量。

**解**：观察该矩阵可知，因为无论如何分布，为恒定值0.2，故要得到最大的H（Y），*b1*和*b2*应为均匀分布，即，此时：，



1. 一般离散无记忆信道

对于一般离散无记忆信道，很难计算其信道容量。但有以下充要条件：

 （3-15）

式（3-15）比较容易理解：如果，则可以小于C，因为该取值不会出现；对所有的，其不可能大于某一常数C，因为如果其>C，则可以通过增加其概率来增加，但增加会导致其自信息量减少，而其它取值提供的信息量增加。最终平衡的结果，是使得所有的，其都等于某一常数C，该常数C即为信道容量。

3.3 离散序列信道的信道容量

对于离散序列无记忆信道，输入和输出分别为长度为L的离散序列，，，如图3.2所示。

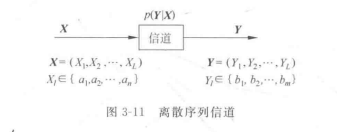
对离散无记忆序列信道：



若信道同时还是平稳的，则：







1. 信源信道皆无记忆情形

若信道无记忆，则：



若信源无记忆，则：



若信源信道皆无记忆，则：

 （3-16）

此时，



1. N次扩展信道

如果对离散单符号信道进行L次扩展，就形成了L次离散无记忆序列信道。

BSC的二次扩展信道：

00

00

10

01

11

11

10

01

***X***∈{00,01,10,11}，***Y***∈{00,01,10,11}，二次扩展无记忆信道的序列转移概率*p*(00/00)=*p*(0/0)*p*(0/0)=(1-*p*)2，*p*(01/00)=*p*(0/0)*p*(1/0)=*p*(1-*p*)，*p*(10/00)=*p*(1/0)*p*(0/0)=*p*(1-*p*)，*p*(11/00)=*p*(1/0)*p*(1/0)=*p*2

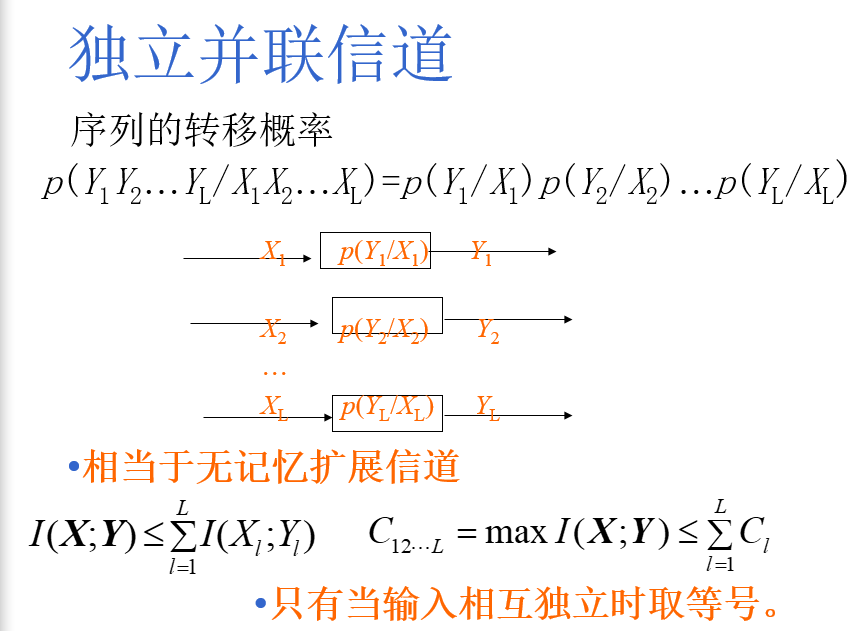




若*p*＝0.1，则C2＝2－0.938＝1.062比特/序列

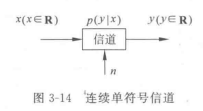
1. 独立并联信道

若信道无记忆，则：



3.4 连续单符号信道的信道容量

最常见的连续单符号信道如下图所示。输入信号x是一个幅度连续的单符号信源，n是加性噪声，y=x+n.



第二章我们曾经说过，一般情况下，我们假定噪声n是均值为零、方差为σ2的加性高斯噪声，即：



记为，其噪声熵为。在这种情况下，



 （3-17）

由限平均功率最大熵定理可知，如能使Y满足高斯分布，则其熵最大。

两个独立高斯分布和，若x=x1+x2，则x满足的高斯分布。

因此，对于y=x+n的单符号连续信道，因为噪声n是均值为零、方差为σ2的加性高斯噪声，如果输入信源x的概率密度函数也是高斯分布，则可使y为高斯分布，达到其最大熵。

设x概率密度为的高斯分布，S为输入信源的功率，则y=x+n是的高斯分布，其中P=S+σ2,是输出信源的功率。

因此，



即

 （3-18）

上述结论是在加性高斯噪声的情况下得到的。如果是非高斯噪声，则有：

 （3-19）

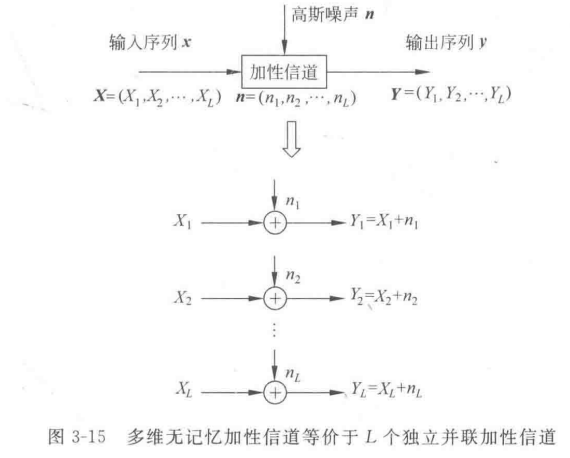
式（3-19）说明如下：

1. 式右边部分：由于噪声为非高斯噪声，其熵为Hc(n)，若输入符号x的概率密度分布pX(x)可以使接收符号y具有高斯分布的概率密度，则其最大熵可以达到，一般情况下，则不一定能达到此最大熵，所以；
2. 式左边部分：当噪声为高斯噪声时，具有最大噪声熵Hc(n)，此时信道容量最小。由于噪声为高斯噪声时， ，故对于非高斯噪声，。

3.5 连续序列信道的信道容量

若加性信道的输入为连续序列，输出为连续序列，**y**=**x**+**n**，其中是单元时刻1、2、…、L上的噪声，均值均为零。

若信道无记忆，则可将连续序列信道等价成并联连续单符号信道，如图所示。



若信源序列也无记忆，则。

一个问题是，若对输入序列有如下约束：



即各时刻输入符号平均功率之和为常数。此时，应该将总功率P如何分配给输入序列中的各个符号，才能得到最大的信道容量？

1. 当噪声序列矢量中各随机变量均为均值为零、方差为σ2的高斯噪声时，根据各独立并联信道的对称性，应将总功率P平均分配给各输入符号，此时：



1. 当噪声序列矢量中各随机变量为均值为零、但方差σl2各不相同的高斯噪声时，求最大容量问题，是如下的有约束极值问题：



且：

对有约束极值问题，可用拉格朗日乘子法构造无约束极值问题，即：





令

则

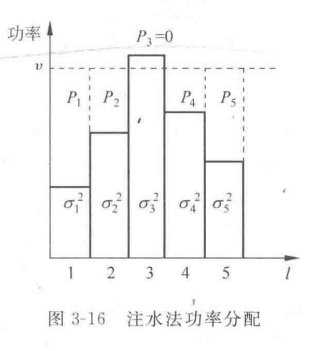
即：

 （3-20）

式（3-20）表明：任一时刻l（l=1,2,…，L）的信号功率Pl与噪声功率之和都相等，令该值为，则：

 （3-21）

该功率分配方法，可以形象地描述为图，如果把噪声功率看成是容器底部的凸起，把看成是水平面，则每个子信道上分配的功率，就如往底部有不同凸起的容器中注水，故形象地称为“注水算法”。



于是，信道容量为：

 （3-22）

**例3-10**：一独立并联高斯加性信道中，各独立并联信道均为零均值加性高斯噪声，其方差分别为：，，，，，，，，，；

求：（1）若输入信号总功率为P=5，各子信道上最佳功率分配；

（2）若输入信号总功率为P=3，各子信道上最佳功率分配；

（3）若输入信号总功率为P=1，各子信道上最佳功率分配；

**解**：（1）若P=5，则输出端总功率（信号功率加噪声功率）为，按照注水算法，平均输出功率（水平面）为10.5/10=1.05，故每个子信道上分配的功率分别为：0.95，0.85，0.75，0.65，0.55，0.45，0.35，0.25，0.15，0.05；

（2）若P=3，则输出端总功率（信号功率加噪声功率）为，按照注水算法，平均输出功率（水平面）为8.5/10=0.85，小于最后两个子信道的噪声功率，关闭此两个子信道，重新计算输出端总功率（此时不能把关闭的两个子信道噪声功率考虑在内），，平均输出功率（水平面）为6.6/8=0.825，故每个子信道上分配的功率分别为：0.725，0.625，0.525，0.425，0.325，0.225，0.125，0.025；

（3）若P=1，则输出端总功率（信号功率加噪声功率）为，按照注水算法，平均输出功率（水平面）为6.5/10=0.65，小于最后四个子信道的噪声功率，关闭此四个子信道，重新计算输出端总功率（此时不能把关闭的四个子信道噪声功率考虑在内），，平均输出功率（水平面）为3.1/6=0.517，仍小于第6个子信道上的噪声功率，关闭第6个子信道，重新计算水平面得：平均输出功率为0.5，故每个子信道上分配的功率分别为：0.4，0.3，0.2，0.1，0.0；由于第5个子信道上分配的功率为零，实际上只有四个子信道可用，总信道容量为。

3.6 波形信道的信道容量

若信道的输入信号是波形信号，即随机过程x(t)，对应的信道称为波形信道。

对于波形信号，若该波形信号是限时（时长tB）限频（最高频率fm）限功率（平均功率为P）的（注：严格的限时限频信号是不存在的，但大多数信号可近似为限时限频信号），可根据奈奎斯特抽样定理，将其变为L=2fmtB的序列信源（见第二章：波形信源），由上节内容可知，其信道容量为：



若每个抽样时刻的噪声均为零均值、方差为σ2的加性高斯噪声，则：

 bit/序列 （3-23）

其中：

L=2fmtB为抽样序列长度；

为第l时刻抽样信号的功率，由于信号功率为P（即单位时间内信号的能量为P），该功率要分配给单位时间内所有的抽样信号（共2fm个），因此，每个抽样信号的功率为：（由于噪声方差均为σ2,故信号功率平均分配给各采样信号）；

为噪声方差，即噪声功率（实际上应为噪声交流功率，但噪声均值为零，故直流功率为零），设单边噪声功率密度为N0，则噪声功率为N0W（W为信道带宽，可认为W=fm），因此，.

将以上各量带入式（3-23），得：

 bit/序列 (3-24)

若考虑单位时间的信道容量，则有：



其中，P为输入波形信号的平均功率，为噪声功率，记为信噪比，则：

 （3-25）

式（3-25）即为著名的**香农公式**，是高斯噪声信道中可靠通信，信息传输速率的上限值。

**对香农公式的讨论：**

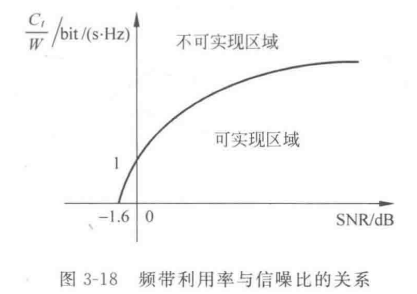
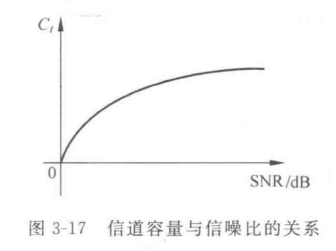
1. 由式（3-25）可知，带宽W和信噪比SNR可以互换：即在保持信道容量不变的前提下，可以通过增加信噪比减小所需带宽，也可以通过增加带宽减小所需信噪比。扩频通信技术即是通过扩展传送信号的带宽，在远高于信号原始带宽的信道上进行传输，从而减小所需信噪比，提高了抗干扰能力。
2. 带宽W一定时，信道容量Ct与信噪比SNR之间的关系：信道容量Ct与信噪比SNR的对数成正比，SNR增加，则Ct也增加，但增加的速率随着SNR的增大而减小；当时，.
3. 当输入信号功率PS一定时，信道容量Ct与信道带宽W之间的关系：由于噪声功率为N0W，当信道带宽W增加时，由于信号输入功率P一定，故信噪比SNR减小，会与式（3-25）中的W因子的作用相互抵消。因此，当时，Ct（此时记为）不一定趋于，利用，令：

 （3-26）

上式说明，即使带宽W趋于无穷大，信道容量仍是有限值。

香农探讨了在的情况下，所需的最小信噪比，即时，所需的，由式（3-26）可知，此时，即-1.6dB，该数值称为香农限，即在保证信道容量为1bit/s的前提下，所需的最小信噪比。

（4） 称为频带利用率，表示每单位频带信道的传输能力。当SNR=1（0dB）时，；当SNR=-1.6dB时，（因为此时需要无穷大的的带宽W）。其关系曲线如图所示。



3.7 信道冗余度

由信道容量的定义式可以看出，只有在信源概率满足某种特定分布时，互信息量I（X；Y）才可以达到其最大值，即信道容量，此时的信源称为该信道的匹配信源，或者说信源和信道是匹配的。

当信源不满足匹配信源概率分布时，互信息量I（X；Y）达不到信道容量，其与信道容量的差值，称为信道绝对冗余度，用表示，即：

 （3-27）

也可以用相对信道冗余度表示：

 （3-28）

信道冗余度反映了信源与信道的匹配程度。信道冗余度大，说明信源与信道的匹配程度低，信道传输能力未得到充分利用；信道冗余度为零，说明信源与信道完全匹配，信道传输能力得到完全利用。

本章小结：

1. **信道的分类：**按照信道输入信号的分类，把信道分成离散单符号信道、离散序列信道、连续单符号信道、连续序列信道、波形信道；、
2. **信道的数学模型：**信道是连接发送端发送信号X和接收端接收信号Y的通道，描述信道用转移概率P（Y/X）；
3. **信道容量的定义：**信道容量是互信息量I（X；Y）的最大值，描述信道的最大传输能力，可改变的参数是信源的概率分布p（x）。能够达到最大互信息量的信源称为匹配信源；
4. **各种信道的信道容量：**
5. 离散单符号信道：
6. 对称DMC：
7. 准对称DMC：
8. 离散序列信道：

，信源信道皆无记忆；

1. 连续单符号信道：

，针对y=x+n，n为加性高斯噪声；

1. 连续序列信道：
2. ，信源信道皆无记忆；
3. 波形信道：
4. ，Ct是**单位时间**的互信息量最大值，**香农公式**；
5. W一定时，Ct随SNR增大而增大，当时，；
6. 信号功率PS一定，时，；
7. **信道冗余度：**；
8. **MIMO信道：**可实现空间复用或空间分集。

第四章 失真与信息率失真函数

4.1 失真的概念和性质

4.1.1 单符号信源失真的定义和度量

4.1.2 序列信源失真的定义和度量

4.2 信息率失真函数

4.2.1 信息率失真函数的定义

4.2.2 信息率失真函数的性质

4.2.3 信息率失真函数与信道容量的比较

4.3 信息率失真函数的计算

本章和第二章与第三章各有若干相似之处。第二章研究信源，本章也是研究信源的；但本章的研究手段，却类似于第三章，因此，会把信息率失真函数与信道容量进行比较。

第二章着重研究信源的熵。在第五章我们将会看到，为提高通信系统的“有效性”，需要进行“信源编码”，即编码后用尽可能少的符号表示原信源信息。如果要求编码是无失真的，则编码后的编码序列所能包含的信息量（是其有能力包含的最大信息量，并不是指其实际包含的信息量）必须大于（至少等于）编码前信源所含有的信息量，在这个过程中，一个核心的要素即是“信源熵”（因为它是编码前每个信源符号所含有的平均信息量），它决定了无失真信源编码能够达到的极限。

但是，有很多场合，或者做不到无失真信源编码，或者没有必要进行无失真信源编码。

比如：对于连续信源，因为其熵为无穷大，若要求对其进行无失真编码，需要用无穷多个比特才能完全无失真地来描述，这显然是不可能的。

比如：由于人耳能够接收的带宽和分辨率是有限的，因此对数字音频，就允许有一定的失真，并且对音乐欣赏没有影响。我们可把频谱范围从20Hz~8000Hz的语音信号去掉低频和高频，保留带宽范围300Hz~3400Hz的信号。这种失真是允许的。

又如：对于数字电视，由于人的视觉系统对低频比较敏感，对高频不太敏感，且人眼分辨率有限，因此可以在一定限度内损失部分高频分量。

等等这些，都决定了限失真信源编码的重要性。

4.1 失真的概念和性质

1. 单符号失真的定义和度量

所谓失真，就是编码前后符号不相同（或者不是一一对应）了。设编码前符号为，编码后的符号为。失真的大小，用失真函数表示。有多种失真函数的定义方法：

1. 均方失真：
2. 绝对失真：
3. 相对失真：
4. 误码失真：

前三种失真度量适用于模拟信源，后一种适用于离散信源。

将失真度量写成矩阵形式，可得失真矩阵：

 （4-1）

平均失真定义为：

 （4-2）

2. 序列失真的定义和度量

设编码前符号序列为，编码后的符号序列为。失真的大小，用失真函数表示：

 （4-3）

式中，是编码前序列中的第*l*个符号，是编码后序列中的第*l*个符号。

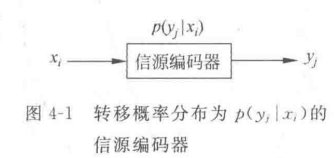
平均失真定义为：

 （4-4）

4.2 信息率失真函数

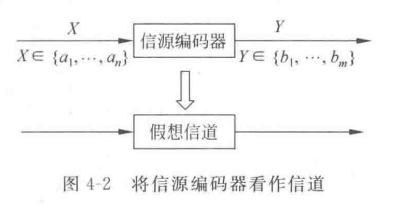
1. 信息率失真函数的定义

信源编码过程方框图如图所示：



从图中可以看出，所谓信源编码，就是将编码前的信源符号按照某种转移概率映射为新的符号，不同的转移概率，意味着不同的信源编码方法。映射的目的，是要使编码后得到的新符号在满足失真要求的前提下，信息率要尽可能低，从而达到信息压缩，提高系统有效性。

可以把上图的信源编码，和第三章分析信道传输进行类比，如图所示：



根据第三章，信道的特性，用信道的转移概率来描述，这和信源编码时的映射概率类似，因此，可以把信源编码器看成是一个信道，叫做假想信道，不同的编码方法，就意味着不同的假想信道。

若允许失真的最大限度是D，则所有满足的映射（不同的假想信道），就构成了一个信道集合，称为D允许实验信道PD：

 （4-5）

信源编码的假想信道，与第三章真正信息传输通道目的不同。

信息传输通道是要尽可能增加信道传输能力，也就是需要了解信道的最大传输能力，即信道容量，它是互信息量I（X；Y）的最大值；

信源编码的假想信道，是要尽可能降低输出码率，即要了解假想信道的最小输出速率，即互信息量I（X；Y）的最小值。

因此，我们定义信息率失真函数为：

 （4-6）

式（4-6）的意义是：在所有的允许实验信道PD（即满足失真要求的所有假想信道）上，最低的输出信息速率（最有效的信源编码），就叫做信息率失真函数。或者说，信息率失真函数是在满足给定最大允许失真的情况下，能够达到的最低编码输出信息率。该信息率是最大允许失真D的函数，所以称为信息率失真函数。

对离散无记忆信源，根据其互信息量表达式，有：

 （4-7）

另一方面：



从信源编码（假想信道）的角度来看，为了使输出信息率尽可能小，在信源给定的情况（一定）下，要在满足失真要求的前提下，尽量增大。是在Y给定的条件下X仍然具有的不确定性，即信源有失真编码造成的信息丢失。

**例4-1**：已知编码器输入的概率分布为，两种信源编码器转移矩阵分别为：



定义单符号失真度



计算两种信源编码方法带来的平均失真。

**解**：

可得





由转移概率矩阵，可以计算出信源编码后的信息传输率为：

比特/符号

比特/符号

两种信源编码器转移概率不同，代表了不同的编码方法，例中经过两种不同的编码方法编码后每个符号传送的信息量，即编码后的信息率不同，，说明编码方法一对信源数据的压缩率高，但是编码方法一带来的失真，说明编码方法一对信源数据带来的失真要大一些。

2. 信息率失真函数的性质

1） 信息率失真函数的定义域和值域

i）Dmin和R（Dmin）

Dmin=0

R（Dmin）=R（0）=H（X）

ii）Dmax和R（Dmax）

R（Dmax）=0

Dmax=min{D:R(D)=0}

设当平均失真时，R(D)达到其下界0。当允许更大失真时，即时，R(D)仍只能继续是0。因为当X和Y统计独立时，平均互信息I(X;Y)=0，可见当时，信源X和接收符号Y已经统计独立了，此时，，与无关。因此，就是在R(D)=0的条件下，在某种分布下，能够得到的平均失真D的最小值，即

 (4.1.21)

也可以改写成

 (4.1.22)

可以让这样分布以达到上式的最小值：当某一个使得为最小时，就取，而其余的，此时求得的一定是最小的。此时，有

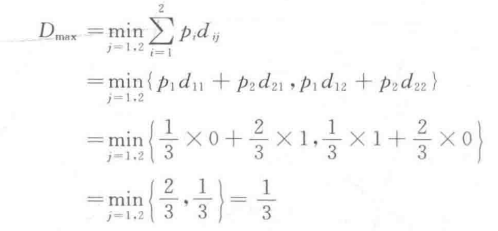
 (4.1.23)

**例4-2**：设输入输出符号表为X=Y={0,1}，输入概率分布为，失真矩阵为



求

**解**：



故输出符号概率为

2）信息率失真函数的下凸性

第二章我们讨论过，互信息量I（X；Y）=f[p(x),p(y/x)]，即互信息量是信源概率分布p（x）和信道（或者假想信道）的转移概率p(y/x)的函数。

当给定信道的转移概率p(y/x)时，互信息量I（X；Y）是信源概率分布p(x)的上凸函数，有极大值，该极大值即为信道容量：.

当给定信源概率分布p(x)时，互信息量I（X；Y）是信道（假想信道）的转移概率p(y/x)的下凸函数，有极小值，该极小值即为信息率失真函数：.

3）信息率失真函数的单调递减性

可以证明，信息率失真函数R（D）随着容许失真D的增大单调递减，即：

若D>D’，则R（D）<R(D’).

3. 信息率失真函数与信道容量的比较

信息率失真函数与信道容量的比较见下表：



可以看出，信息率失真函数与信道容量主要有以下几点相同之处：

1. 两者都可以看作信息传输率；
2. 两者都有信息损失，即H（X/Y）>0,一个是由于信道噪声引起的，一个是由于有失真编码引起的。

但是，信息率失真函数与信道容量是根本不同的量，主要表现在：

1. 研究对象不同：信道容量是关于信道的，信息率是真函数是关于信源的；
2. 给定条件不同：研究信道容量，要给定信道，即给定转移概率。研究信源编码，要给定信源，即给定信源概率分布；
3. 可改变的参数不同：求信道容量时，可改变的是信道上传输的信源，即改变信源概率分布；求信息率失真函数时，可改变的是信源编码方法，即假想信道转移概率；
4. 定义式不同：信道容量是通过改变信源概率分布求得互信息量的最大值；信息率失真函数是通过改变编码方法（即改变假想信道转移概率）求得满足失真要求的互信息量最小值。

4.3 信息率失真函数的计算

除了少数几种特殊情况外，对大多数信源，通常很难计算出信息率失真函数。本书对信息率失真函数的计算不做过多介绍。

第五章 信源编码

5.1 信源编码的概念

5.2 无失真信源编码定理

5.3 限失真信源编码定理

5.4 常用信源编码方法

信源编码的目的，是要提高通信系统的有效性，即通过编码，用尽可能短的编码序列符号来表示原信源序列。其基本思想是：

1. 尽可能去除原消息序列中符号之间的相关性（因为相关性意味着一定的确定性或可预测性，而确定性会减少序列的熵，从而减少序列携带的信息量）；
2. 尽可能使编码后各符号出现的概率相等（因为等概率比不等概率所包含的自信息量大）。

信源编码需要解决两个方面的问题：

1. “信源编码是用尽可能短的编码序列符号来表示原信源序列”，这个“短”，有没有极限？如果有，该极限是多少？无失真信源编码定理和限失真信源编码定理解决该问题，是本章第二节和第三节的内容；
2. 用什么方法达到或者接近该极限？这是信源编码方法需要解决的问题，一些常用的信源编码方法见本章第四节。

5.1 信源编码的概念

一、几个概念

1. 非分组码和分组码

将长消息序列按顺序分成若干各符号一组，对每一组独立进行编码，称为分组码；否则称为非分组码。

普遍使用的是分组码，以下的概念都是针对分组码。

1. 定长码和变长码

若编码后的码字长度都相同，则称为定长码；否则称为变长码。

1. 奇异码和非奇异码

若编码前的信源组和编码后的码字是一一对应的，则称为非奇异码；否则称为奇异码。

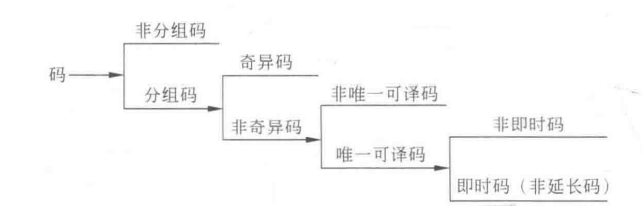
1. 非唯一可译码和唯一可译码

对各个信源组编码后的码字，会连接在一起成为一个码元序列。如果任意有限长的码元序列都只能被唯一地分割成一个个的码字（主要是对变长编码而言。对定长编码，分割是唯一的），则称为唯一可译码；否则称为非唯一可译码。

1. 非即时码和即时码

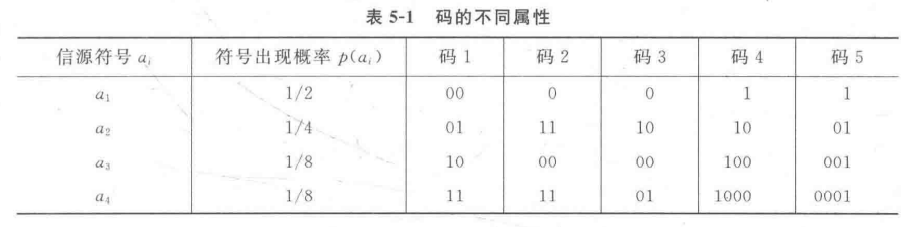
在接收端接收码元序列的过程中，如果接收到的码元序列已经组成了一个码字，但接收端并不能立即判断出，还要等下一个码字开始接收的时候才能判断出前者已经是一个完整码字，从而开始译码，则称为非即时码；反之则称为即时码。

综上所述，编码可分成如图所示的各种类型：



一般来说，我们经常使用的是分组码类型的非奇异码中的唯一可译码中的即时码。

**例5-1**：表5-1中的分组码1、码2、码3、码4和码5，从上述的4个方面分别判断属于什么类型的码。



**解**：码1：非奇异码、定长码、唯一可译码（非奇异定长编码都是唯一可译码）、即时码（定长码都是即时码）；

码2：奇异码、变长码、非唯一可译码（奇异码都是非唯一可译码）、非唯一可译码无所谓即时码或非即时码；

码3：非奇异码、变长码、非唯一可译码（如“00”，既可以理解为*a3*,也可以理解为“0”和“0”，即*a1 a1*）、非唯一可译码无所谓即时码或非即时码；

码4：非奇异码、变长码、唯一可译码、非即时码；

码5：非奇异码、变长码、唯一可译码、即时码（收到一个“1”，则可判断该码字一定结束）。

二、唯一可译码存在的充要条件-克劳夫特不等式

若要求某变长编码的各个码字长度分别为，则这样的唯一可译码编码存在的充要条件是：

 （5-1）

上式称为克劳夫特不等式，其中*m*是进制，*n*是信源符号数。

需要注意的是，上述定理是那种码长分布的唯一可译码**存在**的充要条件，而不是判断是否是唯一可译码的判据。也就是说，若满足该不等式，则这样码长分布的唯一可译码编码是存在的，但并不是说只要码长分布满足克劳夫特不等式，就一定是唯一可译码。

**例5-2**：设对符号集{进行二进制编码；

（1）对应的码长分别为，，,，试判断是否存在这样的唯一可译码；

（2）若K1=1,K2=2,K3=3,K4=3又如何？

**解**：（1）由克劳夫特定理，可得



因此不存在满足这种码长的唯一可译码。

（2）由克劳夫特定理，可知此时：



因此满足这种码长的唯一可译码是存在的，例如{0，10，110，111}。但并不是说，只要满足这种码长分布，就一定是唯一可译码，例如{0，10，010，111}虽然也是码长分别为K1=1,K2=2,K3=3,K4=3，但不是唯一可译码。

5.2 无失真信源编码定理（香农第一定理）

设编码前的符号序列为：

，

编码后的码序列（码字）为：



要求无失真信源编码，即可以从编码后的码字**Y**中完全无误地恢复**X**，同时为保证通信的有效性要求，希望传送Y的信息率最小，即Y的长度KL最短。

KL最短是多少？

这是无失真信源编码定理要解决的问题。下面分定长编码和变长编码两种情况分别加以讨论。

1. 无失真定长编码定理

由L个符号组成的、每个符号的熵为的无记忆平稳信源符号序列，用个符号进行定长变码。对任意，只要

 （5-2）

则当L足够大时，必可使译码差错小于；即可实现几乎无失真编码；

反之，当

 (5-3)

时，译码差错一定是有限值（即不可能实现无失真编码），当L足够大时，译码几乎必定出错（译码错误概率近似等于1）。

定理证明请参阅附件。

不等式两边同时乘以信源序列的长度L，则式(5-2)可改写为

 (5-4)

式（5-4）左边表示长为的码符号序列（码字）所能载荷的最大信息量，而右边代表长为L的信源序列携带的平均信息量。所以无失真定长信源编码定理告诉我们：只要码字传输的信息量大于信源序列携带的信息量，总可以实现几乎无失真的编码。

无失真定长信源编码定理说：“对于任意的，只要L足够大时，必可使译码差错小于”。

我们不禁要问：L、、三者之间有什么关系？对于给定的和，多大的L算足够大？

事实上，可以证明（证明见无失真定长信源编码定理的证明过程）：

 （5-5）

式中，Pe为差错概率，为信源序列的自信息方差。上式说明，对于给定的，只要L足够大，就可以使差错概率小于任意给定的。

从式（5-5）可以看出，如果给定差错概率上界，则越小，要求的编码长度L就越大。L越大，编码器越复杂，且时延越大，在有时延要求的场合，难以满足实时性要求。增加，可以减小对编码长度L的要求，但以牺牲编码效率为代价：

编码效率 （5-6）

越大，编码效率越低。

**例5-3**：设离散无记忆信源概率空间为



若要求编码效率为90%，译码差错概率，试求所需要的编码序列长度L。

**解**：信源熵为



自信息方差为



对信源符号采用定长二元编码，要求编码效率，无记忆信源有，因此



可以得到

如果要求译码错误概率，则



可见，在对编码效率和译码错误概率的要求不是十分苛刻的情况下，就需要个信源符号一起进行编码，这对存储和处理技术的要求太高，目前还无法实现。

**思考**：对例5-3中的信源，有8种不同的信源符号取值（*a*1~*a*8），如果用二进制序列来表示的话，每个符号需要3bit（3位二进制数可以表示8中不同的符号）。但由于不是等概率的，所以其熵H（X）=2.55bit，按照无失真定长信源编码定理，其极限编码长度是2.55bit，而，也就是说，只能表示5.856种不同的符号，其余的符号怎么办？

实际上，由于*a*1~*a*8中部分符号的概率较小，如果序列长度L足够大，则总有某种序列出现的概率足够小，对这些概率足够小的序列，如果不设计对应的编码码字，造成的错误概率也非常小。因此才有：“对于任意的，只要L足够大时，必可使译码差错小于”。

1. 无失真变长编码定理

定长编码中，由于所有的码字都使用相同的长度，限制了其灵活性，导致或者效率不高，或者复杂度太高（L太大）。

变长编码可以对出现概率大的信源尽量用短码，从而提高编码效率。

（1）单符号无失真变长编码定理：

若离散单符号信源的熵为，对每个单符号进行无失真变长编码，设。则

 （5-7）

其中，为平均码长。

定理证明请参阅附件。

（2）离散平稳无记忆序列无失真变长编码定理：

对平均符号熵为的无记忆平稳信源符号序列进行变长变码，必存在一种无失真编码方法，使平均信息率满足：

 (5-8)

需要注意，式（5-8）中是平均信息率，，是任意小正数。

式（5-8）可以由式（5-7）很容易推出：

由式（5-7），把长为L的序列看成一个单符号，则



式中，为平均码长，，故：



整理得：



即：

式中，，只要L足够大，总可以使足够小。

变长编码的效率为：

 （5-9）

**例5-4**：设离散无记忆信源的概率空间为



试讨论其编码效率。

**解**：其信源熵为



（1）若用二元定长编码（0，1）来构造一个即时码：，这时平均码长为



编码效率为



输出的信息率为



（2）若对长度为L=2的信源序列进行变长编码，其即时码如表5-2所示。

表5-2长度为2的信源序列对应的即时码



111

1/16

10

3/16

110

3/16

0

9/16

即时码

序列概率

序列

即时码

序列概率

序列

该码的平均长度为



单符号平均码长为



其编码效率为





这说明，采用序列变长编码，编码复杂了，但提高了信息传输率和效率。

同样可以求得信源序列长度增加到L=3和L=4时，变长编码效率和信息传输率分别为：





（3）若对该信源采用定长二元码编码，要求编码效率达到96%，允许译码错误概率，则：

自信息方差为



为



因此，需要的信源序列长度为



可以看出，使用定长编码时，为了使编码效率较高（96%），需要对非常长的信源序列进行编码，且总存在译码差错。而使用变长编码，只要使用L=2的序列编码，就能使编码效率达到96%，且可以实现无失真编码。

由于码长不是固定的，变长编码的译码相对来说要复杂一些。

5.3 限失真信源编码定理（香农第三定理）

前文说过，在一些场合，无失真信源编码或许是不可能的，也有可能是不必要的。为了更有效地使用通信资源，可以进行限失真信源编码。

**限失真信源编码定理：**

设R(D)为一离散无记忆平稳信源的信息率失真函数，并且有有限的失真测度。则对于任意的和，当信息率R>R(D)时，一定存在一种编码方法，其译码失真小于或等于，条件是编码的信源序列长度L足够长。反之，如果R<R(D)，则无论采用什么编码方法，其译码失真必大于D。

定理说明，在允许失真为D的条件下，信源最小可达的信息传输率是信源的R(D)。

限失真信源编码定理也叫做保真度准则下的信源编码定理，是有失真信源压缩的理论基础。定理说明了在允许失真D确定后，总存在一种编码方法，其编码的信息传输率大于且可以任意接近R(D)，且可以保证平均失真小于允许失真D。当信息传输率小于R(D)时，编码的平均失真将一定大于D。

可见，R(D)是允许失真度为D的情况下信源信息压缩的下限值。比较香农第一定理和香农第三定理可知，当信源给定后，无失真信源压缩的极限值是信源熵H(X)（或信源符号熵），而有失真信源压缩的极限值是信息率失真函数R(D)。在给定D后，一般R(D)<H(X)。R(D)可以作为衡量各种压缩编码方法性能优劣的一种尺度。

香农第三定理同样是一个存在性定理，它只说明一定存在一种满足要求的编码方法。至于如何寻找这种最佳压缩编码方法，定理中没有给出。

在实际应用中，该理论主要存在以下两类问题：

（1）符合实际信源的R(D)函数的计算相当困难；

（2）即使求得了符合实际的信息率失真函数，还需要研究采用何种编码方法，才能达到或接近极限值R(D)。

5.4 常用信源编码方法

香农第一定理和香农第三定理给出了无失真信源编码和限失真信源编码的有效性（编码码长）性能极限。对无失真信源编码来说，该极限就是信源符号熵；对限失真信源编码来说，该极限就是信息率失真函数R(D)。

但香农第一定理和香农第三定理只是存在性定理，它说明达到或接近这种性能限的编码方法是存在的，至于如何进行编码，以达到或接近编码性能极限，在香农定理中没有给出。

描述有效性的指标是编码效率。对定长编码，编码效率体现在编码后的码长；对变长编码，编码效率体现在编码后的平均码长。平均码长是各个码码长的概率平均，为尽可能减小平均码长，应使出现概率大的信源符号尽可能使用短码，出现概率小的信源符号则可以使用码长稍长一些的码。

本节给出几种常用信源编码方法。

1. 香农编码

香农编码是采用信源符号的累计概率分布函数来分配码字。

设信源符号集，并设所有的P(x)>0，香农编码方法如下：；

（1）将信源消息符号按其出现的概率大小依次排列：；

（2）确定满足下列不等式的整数码长：；

（3）为了编成唯一可译码，计算第个消息的累加概率；

（4）将累加概率变换成二进制数；

（5）取二进制数的小数点后位即为该消息符号的二进制码字。

香农编码的基本做法，是把长度为1的整个累积概率区间，按照信源符号集中q个信源符号的概率大小，分成q份不同长度的子区间（每份子区间的长度正比于对应符号的概率大小），将每中信源符号xi，映射到其对应子区间上的一个点。这样，每个信源符号xi映射区间都是不重叠的，从而可以保证唯一可译码，而且可以证明，也是即时码；另一方面，码字长度由其概率决定，概率大的用短码。

**例5-5**： 设信源共有7个符号组成，试求其香农编码。

**解**：香农编码过程如表5-3所示。

表5-3香农编码过程

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 信源符号  xi | 符号概率  p(xi) | 累加概率  Pi |  | 码字长度  Ki | 码字 |
| x1 | 0.20 | 0 | 2.34 | 3 | 000 |
| x2 | 0.19 | 0.2 | 2.41 | 3 | 001 |
| x3 | 0.18 | 0.39 | 2.48 | 3 | 011 |
| x4 | 0.17 | 0.57 | 2.56 | 3 | 100 |
| x5 | 0.15 | 0.74 | 2.74 | 3 | 101 |
| x6 | 0.10 | 0.89 | 3.34 | 4 | 1110 |
| x7 | 0.01 | 0.99 | 6.66 | 7 | 1111110 |

以为例，第4个信源符号所编码的码长为



即



取



第4个信源符号累加概率，变成二进制数，为0.1001…，取4位，得第4个信源符号所编码字为：1001。

小数变二进制数的方法是：用乘以2，如果整数部分有进位，则小数点后第一位为1，否则为0，将其小数部分再做同样的处理，得到小数点后的第二位，依此类推，直到得到了满足要求的位数，或者没有小数部分了为止。

例如现在，乘以2为1.14，整数部分有进位，所以小数点后第一位为1，将小数部分即0.14再乘以2，得0.28，没有整数进位，所以小数点后第二位为0，依此类推，可得到其对应的二进制数为0.1001…。

由表5.2.2可以看出，编码所得的码字，没有相同的，所以是非奇异码，也没有一个码字是其它码字的前缀，所以是即时码，唯一可译码。

例5-5中香农编码的性能可以用平均码长和平均信息传输率、编码效率评价。

平均码长为：



平均信息传输率为



编码效率为



压缩之前7个符号，平均每个符号需要3个比特表示，经香农编码压缩之后的平均码字长度为3.14，因此压缩比为



香农编码的效率不高，本例中压缩比为负值，并没有对信源进行压缩，实用意义不大，但对其它编码方法有很好的理论指导意义。

1. 哈夫曼编码

香农编码效率不高，不具有实际应用价值，但其概率匹配的思想很有指导意义。

哈夫曼编码是按照概率匹配思想构建的一种具有实际使用价值的编码方法。哈夫曼编码是唯一可译码、即时码。

（1）哈夫曼树

为介绍哈夫曼编码，我们先复习一下有关数据结构里“树”的概念。

所谓树，就是既有树根、树枝，又有节点的一种非线性数据结构，如图5-1所示。图中， A、B、C、D、E皆为节点，最上端A为根节点，E为终端节点（叶子节点），A、B、C、D为中间节点。

若每个节点最多有r个孩子节点，则称为r叉树。如果r=2，就叫做二叉树。因为编码大多采用二进制编码，所以在编码里应用最多的是二叉树。

从根节点到最下端节点所走过的树枝个数，叫做树的高度。如图5-1中，从根节点A到最下端的叶子节点E需要走过四个树枝，所以该树的高度为4.

r叉树，如果每个中间节点都有r个孩子节点，则称为满树。否则称为非满树。

可以用树的概念来进行哈夫曼编码，这样的树叫做哈夫曼树。 二进制编码对应一棵二叉树，二叉树的左分支写“0”，右分支写“1”（反之亦可）。若中间节点不安排码字，只在终端节点安排码字，每个终端节点所对应的码字就是从根节点出发到终端节点走过的树枝上所对应的符号序列。例如图5-1中的终端节点E，走过的路径为A-B-C-D-E，所对应的码符号分别为0、0、0、1，则E对应的码字为0001。

可以看出，按码树法构成的码一定是即时码，是非前缀的唯一可译码。



图5-1 码树结构图

当码字长度给定后，用码树法安排的即时码不是唯一的。如图5-1中，如果把左树枝写为“1”，右树枝写为“0”，则得到不同的编码结果，但其平均码长是相同的。

，如果码树上有中间节点安排了码字，则该码一定不是即时码。这是最简单的即时码判断方法。

即时码的码树同样可以用来译码。当收到一串码符号序列后，首先从根节点出发，根据接收到的第一个码符号来选择应走的路径（如图5-1所示，码符号为“0”，则往左分支走，为“1”则往右分支走），再根据收到的第二个符号来选择接下来应走的路径，直到走到一个终端节点为止，就可以根据终端节点，立即判断出所接收的码字。然后从树根继续下一个码字的判断。这样，就可以将接收到的一串码符号序列译成对应的信源符号序列。

（2）哈夫曼编码

1952年，哈夫曼提出了一种构造最佳码的方法。设信源符号集，对应的概率为，且所有的P(x)>0，其编码步骤如下：

1. 以q个信源符号的概率为权值，构造q个带权值的节点。每个节点作为一棵树，构造一棵具有q棵树的森林；
2. 在森林中选出两棵根节点权值最小的树作为一棵新树的左、右子树，且置新树的附加根节点权值为其左、右子树上根节点权值之和；
3. 从森林中删除这两棵树，同时把新树加入到森林中；
4. 重复步骤（2）、（3），直到森林中只有一棵树为止，此树便是哈夫曼树，需要编码的q个符号为该哈夫曼树的q个终端节点；
5. 将哈夫曼树中指向左子树的分支标记为”0”，指向右子树的分支标记为“1”，从根节点到每个终端节点路径上的“0”、“1”组成的序列作为各个终端节点对应的编码，即得哈夫曼编码。

也可用图表的方法进行哈夫曼编码（与哈夫曼树结果相同）：

（1）将q个信源符号按概率分布的大小，以递减次序排列起来，设



(2) 用“0”和“1”码符号分别代表概率最小的两个信源符号，并将两个概率最小的符号合并成一个符号，合并的符号概率为两个符号概率之和，从而得到只包含q-1个符号的新信源，称为缩减信源。

(3)把缩减信源的符号仍旧按概率大小以递减次序排列，再将其概率最小的两个信源符号分别用“0”和“1”表示，并将其合并成一个符号，概率为两符号概率之和，这样又形成了q-2个符号的缩减信源。

(4)依此继续下去，直至信源只剩下两个符号为止。将这最后两个信源符号分别用“0”和“1”表示。

(5)然后从最后一级缩减信源开始，向前返回，就得出各信源符号所对应的码符号序列，即对应的码字。

由于哈夫曼树中任何终端节点（待编码节点）不可能是别的终端节点的祖先节点，故哈夫曼编码是非前缀码，是即时码。

**例5-6**： 设给定信源，对应的概率为的哈夫曼编码过程如图5-2和表5-2所示。

0.01

0.10

0.11

0.15

0.26

0.17

0.18

0.35

0.61

0.19

0.20

0.39

1

0

0

1

0

1

0

1

0

1

0

1

图5-2 树图法哈夫曼编码过程

表5-2哈夫曼编码过程



平均码长为



信息传输率为



可见，该信源的哈夫曼编码效率达约96%.

哈夫曼编码结果并不是唯一的。这是因为：1. 哈夫曼树的左右分支表示“0”或者“1”，是可以互换的；2. 当有多个节点具有相同的概率时，选择哪些节点进行合并是任意的。

需要注意的是，如果缩减信源中有两个以上的节点概率相同，则应优先选择未被合并过（或合并次数少）的节点进行合并，以便尽量减小编码方差，从而减小编码设备复杂度。

算术编码

香农编码和哈夫曼编码，都是基于分组的块编码。分组编码方法没有考虑到组间符号的相关性，因此编码效率会有所损失。增加码长可以考虑更多符号间的相关性，但会增加编码复杂度及编码时延。

跳出分组码的思路，研究非分组码，可能会解决这一问题。算术编码是一种非分组编码方法，其基本思路是借鉴香农编码的区间映射思想。

在介绍香农编码时，我们说过：“香农编码的基本做法，是把长度为1的整个累积概率区间，按照信源符号集中q个信源符号的概率大小，分成q份不同长度的子区间（每份子区间的长度正比于对应符号的概率大小），将每中信源符号xi，映射到其对应子区间上的一个点。这样，每个信源符号xi映射区间都是不重叠的，从而可以保证唯一可译码”。

算术编码从全序列出发，将不同的信源序列的累计概率映射到[0，1]区间上，使每个序列对应区间上的一点，也就是说，把区间[0，1]分成许多互不重叠的小区间，不同的信源序列对应不同的小区间，可以证明，只要这些小区间互不重叠，就可以编得即时码。

上述找位置的过程，是根据原始符号序列的一个个顺序给出，在（0，1）区间上一步步细化对应位置的区间，相当于编码的过程；

根据位置一层层确定其所在的各层区间，从而报出对应的长度数字，相当于解码过程。当然，序列必须有一个结束标志（结束符），或者可以从其他途径知道数据流的结束。否则译码过程会一直持续下去。

这种编码方法，并不需要实现对原始符号序列进行分组，而是随着序列中符号的顺序到来，一步步细化其映射到的位置范围，所以是非分组码。

而且，这种编码方法无需计算出所有信源序列的概率分布及编出码表，只要知道所有单符号的先验概率就可以了。

算术编码的过程与上述米尺的例子相似。不同点主要有两个：

1. 区间的划分，不是等分，而是根据概率划分成长度不等的区间；
2. 用当前累积概率分布代表当前信息符号序列。

设P（S）表示序列S的累积概率，A（S）为序列S对应的区间长度，C（S）表示序列S的码字，表示表示m进制单符号的概率，Pr（r=0，1，2，…，m-1）表示*m*进制单符号的累积概率。可以把把上述算术编码寻找对应子区间的过程写成如下一般形式：

（1）设起始状态为空序列S=ϕ，A(ϕ) ＝1，P(ϕ)＝0；

（2）设在序列S后又来一个新符号r，从而得到新序列Sr后，更新累积概率分布函数P(Sr)及区间长度A(Sr):



（3）重复（2）直至序列结束。

由于累积分布函数和子区间宽度都是递推公式，因此在实际应用中，只需要两个存储器，存储A(s)和P(s)，随着符号的输入，不断地更新两个存储器中的数值。因为在编码过程中，每输入一个符号要进行乘法和加法运算，所以称这种编码方法为算术编码。

通过上述信源符号序列累计分布函数的计算，可以看出，P(S)可以把区间[0，1）分割成许多小区间，不同的信源符号序列对应于不同的区间：

[P(S),P(S)+A(S)） （5-10）

可取小区间内的一点来代表这序列。如何选择这个点,即如何由累积概率P(S)得到码字C(S)？

C(S)可按下述方法得到：

将符号序列的累计分布函数写成二进制小数，取小数点后L位，若后面有尾数，则进位到第L位，并使L满足：

 （5-11）

这样得到的一个数，就作为码字C(S)。

设按上述方法得到的P（S）=0.z1z2…zL，取0或者1，得符号S的码字C(S)为z1z2…zL。根据二进制小数截去位数的影响，可知：



即：

 （5-12）

当P(s)在L位后没有尾数时，C（S）=P(s).

由于，故

所以



因此，式（5-12）可以写为

 （5-13）

式（5-13）表明，用上述方法得到的码字C（S），位于区间[P(S),P(S)+A(S)），与式（5-10）相符合。

由此可知，不同的信源序列对应的不同区间（左封右开的区间）是不重叠的，所以编得的码是即时码。

**例5-7**：有四个符号a，b，c，d构成简单序列S＝abda，各符号及其对应概率如下表，

|  |  |  |
| --- | --- | --- |
| 符号 | 符号概率pi | 符号累积概率Pj |
| a | 0.100(1/2) | 0.000 |
| b | 0.010(1/4) | 0.100 |
| c | 0.001(1/8) | 0.110 |
| d | 0.001(1/8) | 0.111 |

试给出其算术编码编解码过程。

**解**：算术编码过程如下：

1. 设起始状态为空序列*ϕ*，则A(*ϕ*) ＝1，P(*ϕ*)＝0；
2. 
3. 
4. 
5. 
6. 

故编码后的码字C(abda)即为序列abda的累积概率P(abda)（二进制数）小数点后的7位：0101110.

编码过程如下图所示：

解码过程如下：

C(*abda*)=0.010111<0.1∈[0,0.1]，第一个符号为*a；*

放大至[0,1](×*p*a-1)：

C(*abda*)×21＝0.10111∈[0.1,0.110]，第二个符号为*b；*

去掉累积概率Pb： 0.10111-0.1=0.00111

放大至[0,1](×*p* b-1)：

0.00111×22=0.111 ∈[0.111,1]，第三个符号为*d；*

去掉累积概率Pd： 0.111-0.111=0

放大至[0,1](×p d-1)：

0×24＝0 ∈[0,0.1]，第四个符号为*a*

因此得到序列为abda，解码完毕。

据此可知，算术编码的过程如下：

设信源符号集，其相应的概率分布为（假定已按照从小到大排序）。

定义信源符号的累积分布函数为：



则



对二元序列有：



下面，通过计算某信源序列的累积分布函数来做说明。只讨论二元无记忆信源，结果可推广到一般情况。

(1) 初始时，在[0，1）区间内由F(1)划分成二个子区间[0，F(1)）和[F(1)，1），F(1)=p(0)。子区间[0，F(1)]的宽度为A(0)=p(0)，子区间[F(1)，1）的宽度为A(1)=p(1)。子区间[0，F(1) ）对应信源符号“0”，子区间[F(1)，1）对应信源符号“1”。

若输入符号序列的第一个符号为“0”，即落入相应的区间为[0，F(1)），得F(s=“0”)=F(0)=0。即某序列累积概率分布函数为该序列所对应区间的下界值；

(2) 当输入第二个符号时，其所对应的区间是在[0，F(1)）中进行分割。若第二个符号为“0”，则符号序列为“00”，对应新分割的第一个小区间，即区间[0，F(s=“01”)），区间宽度为A(00)=A(0)p(0)=p(0)p(0)；若第二个符号为“1”，则符号序列为“01”，符号序列“01”对应新分割的第二个小区间，即区间[F(s=“01”)，F(1)），区间宽度为A(01)=A(0)p(1)=p(0)p(1)=p(01)。其中F(s=“01”)是符号序列“01”区间的下界值，可见，F(s=“01”)=p(0)p(0)正是符号序列s=“01”的累计分布函数。

后续过程，可以此类推。

蓝色骨头 - 崔健

词：崔健

曲：崔健

并不可惜 也并不可气

我经过了基本的努力

接受了基本的教育

我就是一个春天的花朵

正好长在一个春天里

我爸爸当初告诉我要想有出息

就得好好学习拿出好成绩

可是我曾经不太相信这个

我现在还是不太相信这个

我说人活着要痛快

加独立才算是有意义

所以我学校还没毕业

就开始找了个工作

我要干我最喜欢干的

不管挣钱多少

所以我的工作就是一个写字的

一开始我就是想用笔发发牢骚

可是谁知道这一开始

就一发不可收拾

俗话说活人不能被\*憋死

只要我有笔 谁都拦不住我

这就是我的事业 更是我的兴趣

还能有什么工作比这更来兴趣呢

钱虽然不多 所以我并不太忙

正好剩下了时间

让我琢磨活着的意义

三角架有三条腿才稳定

少了任何一条都要不停的运动

我的生活也要有三大要素才幸福

就是为了得到幸福人们才忙活着

第一 就是事业像我上面说的

能高高兴兴工作挣钱养活自己

有话就说 有话就写 而且要彻底

因为每次彻底之后

才会出现美妙的空虚

第二 就是身体一定要健康

因为身体要是不舒服

什么都是白给

所以我一周三次跑步加上一次游泳

在运动中想事儿越想越起劲儿

第三 当然就是一个爱情了

其实姑娘们不知道

小伙子心中的虚弱

没有爱情的日子自然哥儿们多

就像男人越是闲着 越是人缘儿好

哥儿们之间谈论爱情认真也是假的

只有在姑娘面前动感情才算是真的

当你真的爱的时候理论都是虚的

只有分手的时候疼痛才是实的

为什么没有人告诉我

没有人告诉我 有人在追求我

是不是我的工作太多了

感情也变坏了

还是身体一独立欲望就变野了

反正这三条腿儿的原则

听起来有点简单

可是现实中得到两个就不容易

如今美女都需要好的身体

谁能告诉我爱情

到底要我使出多大的力气

红色 黄色和蓝色

分别代表人的心 身体和智慧

如今这三个颜色

统统被泥土盖了起来

就像眼前这个社会的大酱缸

多年的政治运动使人们厌倦了红色

周围黄色的肉体已经把灵魂埋没

只有扭曲一下我自己

抬头看看上面

原来是少有的一片蓝蓝的天空

红色已经把鲜血污染了

真不知血和心到底哪个是热的

阳光和灯光同时照着我的身体

要么我选择孤独 要么我选择堕落

蓝色的天空给了我无限的理性

看起来却像是忍受

只有无限的感觉

才能给我无穷的力量

爸爸 我就是一个春天的花朵

正好长在一个春天里

因为我的骨头却是蓝色