



信息论

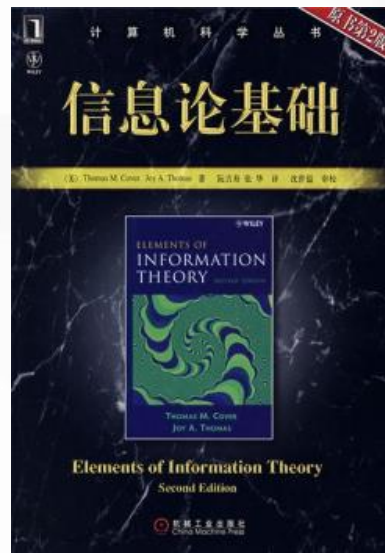
THEORY
INFORMATION

孙妍

计算机学院805

yansun@shu.edu.cn

参考书籍



课程信息

- 课程时间地点：1-10 周，周三 6-8节，材218
- 考核方式：平时成绩（出勤，作业） 30%；考试成绩 70%
- 答疑时间及地点：周三 10-11节 计算机楼805
- 作业：每一章节附有练习题，将在同步教学过程中发布
- 课上不准吃东西、不准大声喧哗
- 课后复习，独立作业，不要考试突击

»»» 什么是信息



信息论的应用

数据压缩

- 压缩制式
- 压缩极限在哪？

ZIP



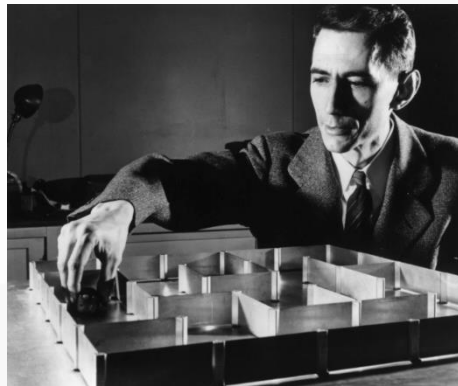
视频编码

- 编码标准的差异
- 平衡清晰度和对于带宽的要求



信息论与人工智能

- 信息熵：损失函数
- 决策树
- 维特比算法：用于自然语言处理
- 自动编码解码技术：语言翻译
- 信息安全/密码学



»»» 为什么要学习信息论

- 通过理论进行创新
- 如果系统复杂到一定程度，需要理论来解决
- 科技发展需要理论指导



目录 CONTENTS

第一章 绪论

第二章 信息的度量

第三章 信源及信息熵

第四章 信道及信道容量

第五章 无失真信源编码

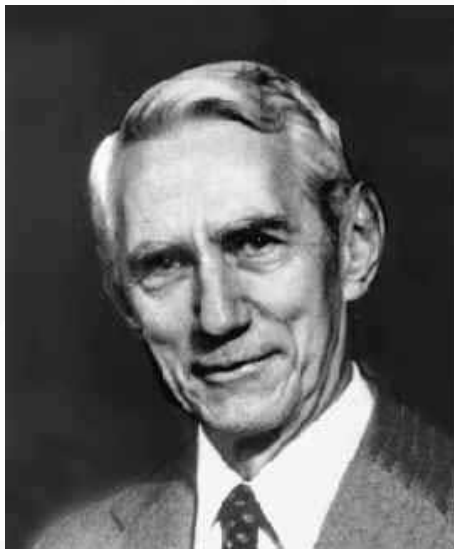
第六章 有噪信道编码

第七章 限失真信源编码

第一章 绪论



1.1 信息的概念



香农

信息论是通信的数学基础，它是随着通信技术的发展而形成和发展起来的一门新兴横断学科。信息论创立标志是1948年Claude Shannon(**香农**)发表论文“A Mathematical Theory of Communication”。在这篇文章中香农创造性的采用**概率论**的方法来研究通信中的问题，并且对信息给予了科学的定量描述，第一次提出了**信息熵**的概念。

1928年，**哈特莱**(Hartley)首先提出了用对数度量信息的概念。一个消息所含有的**信息量**用它的可能值的个数的对数来表示。

1.1 信息的概念

- 信息就是对事物运动状态或者存在方式的不确定性的描述
- 自信息：某个消息出现的不确定性，也是这个消息所包含的信息量

$$I(x_i) = -\log p(x_i) = \log \frac{1}{p(x_i)}$$

- 信息熵：信源所含有的信息量，也就是自信息的统计平均，也被称为平均自信息

$$H(X) = E[I(x_i)] = -\sum_{i=1}^q p(x_i) \log p(x_i)$$

➤➤➤ 1.2 信息论研究对象、目的和内容

➤ 信息论研究对象

- 广义的通信系统，它把所有的信息流通系统都抽象成如下模型：

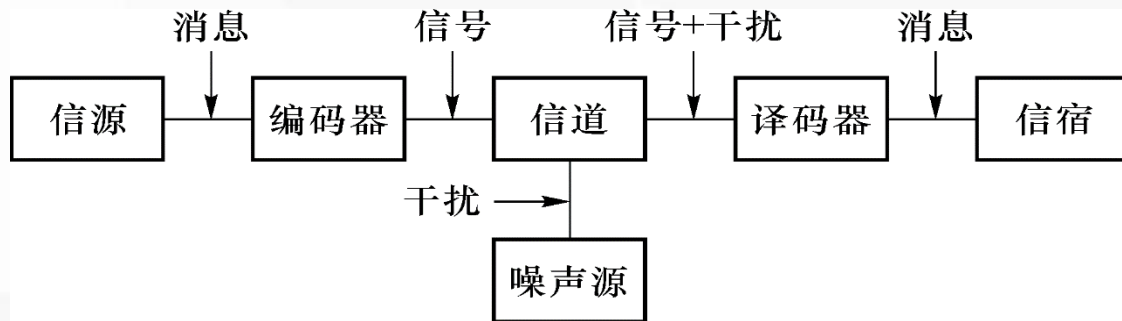


图1.1 通信系统模型

1.2 信息论研究对象、目的和内容

这个通信系统主要分成五个部分：

- **(1)信源** 顾名思义，信源是产生消息和消息序列的源。
- **(2)编码器** 编码就是把消息变成适合在信道传输的物理量——**信号**。
- 编码器可分为信源编码器和信道编码器。信源编码目的是提高通信系统**有效性**和提高信息 传输的**可靠性**。实际通信系统中，可靠性和有效性常常相互矛盾。
- **(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。
- **(4)译码器** 译码就是把信道输出的已迭加了干扰的编码信号进行反变换，变成信宿能够接受的消息。译码器也可分成信源译码器和信道译码器。
- **(5)信宿** 信宿是消息传送的对象，即接受消息的人或机器。

1.2 信息论研究对象、目的和内容

信息论研究的是关于这个通信系统的最根本、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. 为了能够无失真的传输信源信息，对信源编码时所需的最少的码符号数是多少？

(无失真信源编码即**香农第一定理**)

- 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？

(有噪信道编码即**香农第二定理**)

- 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？

(限失真信源编码即**香农第三定理**)

1.2 信息论研究对象、目的和内容

目前，对信息论的研究内容一般有三种理解：

(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：

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领域，如模式识别、计算机翻译、心理学、遗传学、神经生理学、语言学、语义学甚至包括社会学中有关信息的问题。



第二章 信息的度量



»»» 第二章 信息的度量

关于信息的度量有几个重要的概念：

(1) 自信息：一个事件(消息)本身所包含的信息量，它是由事件的不确定性决定的。比如抛掷一枚硬币的结果是正面这个消息所包含的信息量

(2) 互信息：一个事件所给出关于另一个事件的信息量，比如下天下雨所给出关于明天下雨的信息量。

(3) 平均自信息（信息熵）：事件集（用随机变量表示）所包含的平均信息量，它表示信源的平均不确定性。比如抛掷一枚硬币的试验所包含的信息量。

(4) 平均互信息：一个事件集所给出关于另一个事件集的平均信息量，比如今天的天气所给出关于明天的天气的信息量。

2.1.1 自信息

定义2.1 随机事件的自信息量定义为该事件发生概率的对数的负值。

事件 x_i 的概率为 $p(x_i)$, 则它的自信息定义为:

$$I(x_i) \stackrel{\text{def}}{=} -\log p(x_i) = \log \frac{1}{p(x_i)}$$

从图2.1种可以看到上述信息量的定义正是满足上述公理性条件的函数形式。 $I(x_i)$ 代表两种含义：当事件发生以前，等于事件发生的不确定性的的大小；当事件发生以后，表示事件所含有或所能提供的信息量。

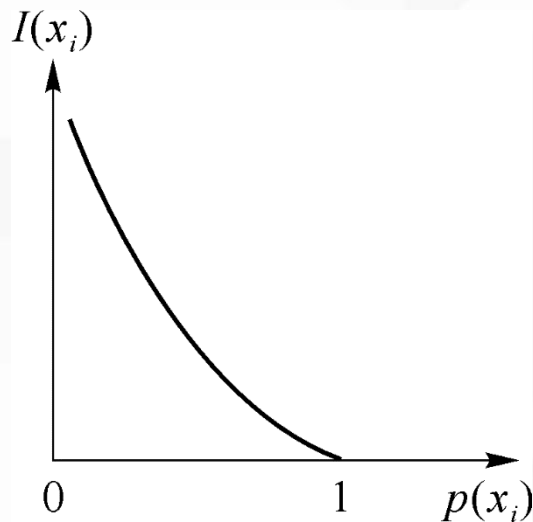


图2.1 自信息量

2.1.1 自信息

自信息量的单位与所用对数的底有关。

(1) 常取对数的底为2，信息量的单位为比特(bit, binary unit)。当

$p(x_i) = 1/2$ 时， $I(x_i) = 1$ 比特，即概率等于1/2的事件具有1比特的自信息量。

(2) 若取自然对数（对数以e为底），自信息量的单位为奈特（nat,

natural unit）。1奈特 = $\log_2 e$ 比特 = 1.443比特

(3) 工程上用以10为底较方便。若以10为对数底，则自信息量的单位

为哈特莱（Hartley）。1哈特莱 = $\log_2 10$ 比特 = 3.322比特

(4) 如果取以 r 为底的对数（ $r > 1$ ），则 $I(x_i) = -\log_r p(x_i)$ 进制单位

1 r 进制单位 = $\log_2 r$

2.1.1 自信息

【例2.1】

- 1) 英文字母中“a”出现的概率为0.0064, “c”出现的概率为0.022, 分别计算它们的自信息量。
- 2) 假定前后字母出现时互相独立的, 计算“ac”的自信息量。
- 3) 假定前后字母出现不是互相独立的, 当“a”出现以后, “c”出现的概率为0.04, 计算“a”出现以后, “c”出现的自信息量。

2.1.2 互信息

定义2.2 一个事件 y_j 所给出关于另一个事件 x_i 的信息定义为**互信息**，用 $I(x_i; y_j)$ 表示。

$$I(x_i; y_j) \stackrel{def}{=} I(x_i) - I(x_i | y_j) = \log \frac{p(x_i | y_j)}{p(x_i)}$$

互信息 $I(x_i; y_j)$ 是已知事件 y_j 后所消除的关于事件 x_i 的不确定性，它等于事件 x_i 本身的不确定性 $I(x_i)$ 减去已知事件 y_j 后对 x_i 仍然存在的不确定性 $I(x_i | y_j)$ 。

互信息的引出，使信息得到了定量的表示，是信息论发展的一个重要的里程碑。

2.1.2 互信息

【例2.2】

某地二月份天气出现的概率分别为晴 $1/2$ ，阴 $1/4$ ，雨 $1/8$ ，雪 $1/8$ 。某一天有人告诉你：“今天不是晴天”，把这句话作为收到的消息 y_1 ，求收到 y_1 后， y_1 与各种天气的互信息量。

➤➤➤ 2.2 平均自信息

2.2.1 平均自信息(信息熵)的概念

自信息量是信源发出某一具体消息所含有的信息量，发出的消息不同所含有的信息量不同。因此自信息量不能用来表征整个信源的不确定度。

我们定义**平均自信息量**来表征整个信源的不确定度。平均自信息量又称为**信息熵、信源熵**，简称**熵**。

因为信源具有不确定性，所以我们把信源用随机变量来表示，用随机变量的概率分布来描述信源的不确定性。通常把一个随机变量的所有可能的取值和这些取值对应的概率 $[X, P(X)]$ 称为它的**概率空间**。

2.2.1 平均自信息(信息熵)的概念

定义2.3 随机变量X的每一个可能取值的自信息的统计平均值定义为随机变量X的**平均自信息量**:

$$H(X) = E[I(x_i)] = -\sum_{i=1}^q p(x_i) \log p(x_i)$$

这里q为的所有X可能取值的个数。

熵的单位也是与所取的对数底有关，根据所取的对数底不同，可以是比特/符号、奈特/符号、哈特莱/符号或者是r进制单位/符号。通常用比特/符号为单位。

一般情况下，信息熵并不等于受信者平均获得的信息量，受信者不能全部消除信源的平均不确定性，获得的信息量将小于信息熵。

➤➤➤ 2.2.1 平均自信息(信息熵)的概念

【例2.3】

假设随机变量X的概率分布为 $p(x^i) = 2^{-i}$, $i = 1, 2, 3, 4 \dots$, 求 $H(X)$ 。

2.2.2 熵函数的性质

信息熵 $H(X)$ 是随机变量 X 的概率分布的函数，所以又称为**熵函数**。如果把概率分布 $p(x_i), i=1, 2, \dots, q$ ，记为 p_1, p_2, \dots, p_q ，则熵函数又可以写成概率矢量 $\mathbf{p} = (p_1, p_2, \dots, p_q)$ 的函数的形式，记为 $H(\mathbf{p})$ 。

$$H(X) = -\sum_{i=1}^q p_i \log p_i = H(p_1, p_2, \dots, p_q) = H(\mathbf{p})$$

熵函数 $H(\mathbf{p})$ 具有以下性质：

1. 对称性：

$$H(p_1, p_2, \dots, p_q) = H(p_2, p_1, \dots, p_q) = \dots = H(p_q, p_1, \dots, p_{q-1})$$

对称性说明熵函数仅与信源的总体统计特性有关。

2.2.2 熵函数的性质

2. 确定性: $H(1, 0) = H(1, 0, 0) = H(1, 0, 0, 0) = \dots = H(1, 0, \dots, 0) = 0$

在概率矢量中，只要有一个分量为1，其它分量必为0，它们对熵的贡献均为0，因此熵等于0。也就是说确定信源的不确定度为0。

3. 非负性: $H(\mathbf{p}) = H(p_1, p_2, \dots, p_q) \geq 0$

对确定信源，等号成立。信源熵是自信息的数学期望，自信息是非负值，所以信源熵必定是非负的。

4. 扩展性: $\lim_{\varepsilon \rightarrow 0} H_{q+1}(p_1, p_2, \dots, p_q - \varepsilon, \varepsilon) = H_q(p_1, p_2, \dots, p_q)$

这个性质的含义是增加一个基本不会出现的小概率事件，信源的熵保持不变。

5. 连续性: $\lim_{\varepsilon \rightarrow 0} H(p_1, p_2, \dots, p_{q-1} - \varepsilon, p_q + \varepsilon) = H(p_1, p_2, \dots, p_q)$

即信源概率空间中概率分量的微小波动，不会引起熵的变化。

2.2.2 熵函数的性质

6. 递推性:

$$H(p_1, p_2, \dots, p_{n-1}, q_1, q_2, \dots, q_m) = H(p_1, p_2, \dots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}, \dots, \frac{q_m}{p_n}\right)$$

这性质表明，假如有一信源的 n 个元素的概率分布为 (p_1, p_2, \dots, p_n) ，其中某个元素 x_n 又被划分成 m 个元素，这 m 个元素的概率之和等于元素 x_n 的概率，这样得到的新信源的熵增加，熵增加了一项是由于划分产生的不确定性。

7. 极值性:

$$H(p_1, p_2, \dots, p_n) \leq H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right) = \log n$$

式中 n 是随机变量 X 的可能取值的个数。

极值性表明离散信源中各消息等概率出现时熵最大，这就是最大离散熵定理。

2.2.2 熵函数的性质

【例2.4】

利用递推性计算 $H(1/2, 1/8, 1/8, 1/8, 1/8)$ 。

2.2.2 熵函数的性质

8. 上凸性: 是严格的上凸函数, 设 $H(\mathbf{p})$

$$\mathbf{p} = (p_1, p_2, \dots, p_q), \mathbf{p}' = (p_1', p_2', \dots, p_q'), \sum_{i=1}^q p_i = 1, \sum_{i=1}^q p_i' = 1$$

则对于任意小于1的正数 α , ($0 < \alpha < 1$) 有以下不等式成立:

$$H[\alpha \mathbf{p} + (1-\alpha) \mathbf{p}'] > \alpha H(\mathbf{p}) + (1-\alpha) H(\mathbf{p}')$$

凸函数在定义域内的极值必为极大值, 可以利用熵函数的这个性质

可以证明熵函数的极值性。

2.2.2 熵函数的性质

直观来看，随机变量的不确定程度并不都是一样的。**香农**指出，存在这样的不确定性度量，它是随机变量的概率分布的函数，而且必须满足三个公理性条件

- 1. 连续性条件：** $f(p_1, p_2, \dots, p_n)$ 应是 $p_i, i=1, 2, \dots, n$ 的连续函数；
- 2. 等概时为单调函数：** $f(1/n, 1/n, \dots, 1/n)$ 应是 n 的增函数；
- 3. 递增性条件：** 当随机变量的取值不是通过一次试验而是若干次试验才最后得到时，随机变量在各次试验中的不确定性应该可加，且其和始终与通过一次试验取得的不确定程度相同，即：

$$\begin{aligned} & f(p_1, p_2, \dots, p_n) \\ &= f[(p_1 + p_2 + \dots + p_k), p_{k+1}, \dots, p_n] + (p_1 + p_2 + \dots + p_k) f(p_1', p_2', \dots, p_k') \end{aligned}$$

其中 $p_k' = p_k / (p_1 + p_2 + \dots + p_k)$

2.2.3 联合熵与条件熵

一个随机变量的不确定性可以用熵来表示，这一概念可以方便地推广到多个随机变量。

定义2.4 二维随机变量 XY 的概率空间表示为

$$\begin{bmatrix} XY \\ P(XY) \end{bmatrix} = \begin{bmatrix} x_1 y_1 & \cdots & x_i y_j & \cdots & x_n y_m \\ p(x_1 y_1) \cdots p(x_i y_j) \cdots p(x_n y_m) \end{bmatrix}$$

其中： $p(x_i y_j)$ 满足概率空间的非负性和完备性：

$$0 \leq p(x_i y_j) \leq 1, \sum_{i=1}^n \sum_{j=1}^m p(x_i y_j) = 1$$

2.2.3 联合熵与条件熵

二维随机变量 XY 的**联合熵**定义为联合自信息的数学期望，它是二维随机变量 XY 的不确定性的度量。

$$H(XY) \stackrel{\text{def}}{=} \sum_{i=1}^n \sum_{j=1}^m p(x_i y_j) I(x_i y_j) = - \sum_{i=1}^n \sum_{j=1}^m p(x_i y_j) \log p(x_i y_j)$$

定义2.5 给定 X 时, Y 的**条件熵**:

$$\begin{aligned} H(Y|X) &= \sum_i p(x_i) H(Y|x_i) = - \sum_i \sum_j p(x_i) p(y_j|x_i) \log p(y_j|x_i) \\ &= - \sum_i \sum_j p(x_i y_j) \log p(y_j|x_i) \end{aligned}$$

其中, $H(Y|X)$ 表示已知 X 时, Y 的平均不确定性。

»»» 各类熵之间关系:

1. 联合熵与信息熵、条件熵的关系:

$$H(XY) = H(X) + H(Y|X)$$

这个关系可以方便地推广到 N 个随机变量的情况:

$$H(X_1X_2\cdots X_N) = H(X_1) + H(X_2|X_1) + \cdots + H(X_N|X_1X_2\cdots X_{N-1})$$

称为**熵函数的链规则**。

推论: 当二维随机变量 X, Y 相互独立时, 联合熵等于 X, Y 各自熵之和:

$$H(XY) = H(X) + H(Y)$$

2. 条件熵与信息熵的关系:

$$H(X|Y) \leq H(X), H(Y|X) \leq H(Y)$$

3. 联合熵和信息熵的关系:

$$H(XY) \leq H(X) + H(Y) \quad \text{当 } X, Y \text{ 相互独立时等号成立。}$$

➤➤➤ 各类熵之间关系:

【例2.5】

随机变量 X, Y 的联合概率分布如下表所示, 求联合熵 $H(XY)$ 和条件熵 $H(Y|X)$ 。

$X \backslash Y$	0	1	$p(x_i)$
0	1/4	1/4	1/2
1	1/2	0	1/2
$p(y_j)$	3/4	1/4	1

2.3 平均互信息

2.3.1 平均互信息的概念

为了从整体上表示从一个随机变量 Y 所给出关于另一个随机变量 X 的信息量，我们定义互信息 $I(x_i; y_j)$ 在的 XY 联合概率空间中的统计平均值为随机变量 X 和 Y 间的**平均互信息**：

定义2.6

$$\begin{aligned} I(X; Y) &= \sum_{i=1}^n \sum_{j=1}^m p(x_i y_j) I(x_i; y_j) = \sum_{i=1}^n \sum_{j=1}^m p(x_i y_j) \log \frac{p(x_i | y_j)}{p(x_i)} \\ &= \sum_{i=1}^n \sum_{j=1}^m p(x_i y_j) \log \frac{1}{p(x_i)} - \sum_{i=1}^n \sum_{j=1}^m p(x_i y_j) \log \frac{1}{p(x_i | y_j)} \\ &= H(X) - H(X | Y) \end{aligned}$$

2.3.2 平均互信息的性质

1.非负性: $I(X;Y) \geq 0$

平均互信息是非负的，说明给定随机变量 Y 后，一般来说总能消除一部分关于 X 的不确定性。

2.互易性(对称性): $I(X;Y) = I(Y;X)$

对称性表示从 Y 中获得的关于 X 信息量等于从 X 中获得的关于 Y 信息量

3.平均互信息和各类熵的关系:

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(XY) \end{aligned}$$

当 X, Y 统计独立时, $I(X;Y) = 0$

➤➤➤ 2.3.2 平均互信息的性质

4. 极值性: $I(X;Y) \leq H(X), I(X;Y) \leq H(Y)$

极值性说明从一个事件提取关于另一个事件的信息量，至多只能是另一个事件的平均自信息量，不会超过另一事件本身所含的信息量。

5. 凸函数性:

定理2.1 当条件概率分布 $\{p(y_j | x_i)\}$ 给定时，平均互信息 $I(X;Y)$ 是输入分布 $\{p(x_i)\}$ 的上凸函数。

定理2.2 对于固定的输入分布 $\{p(x_i)\}$ ，平均互信息量 $I(X;Y)$ 是条件概率分布 $\{p(y_j | x_i)\}$ 的下凸函数。

➤➤➤ 2.3 平均互信息

【例2.6】

掷骰子，若结果是1, 2, 3或4，则抛一次硬币；如果结果是5或者6，则抛两次硬币，试计算从抛硬币的结果可以得到多少掷骰子的信息量。

2.3.3 数据处理定理

为了表述数据处理定理，需要引入三元随机变量 X, Y, Z 的平均条件互信息和平均联合互信息的概念。

定义2.7 平均条件互信息

$$I(X; Y | Z) = E\{I(xy | z)\} = \sum_x \sum_y \sum_z p(xyz) \log \frac{p(x | yz)}{p(x | z)}$$

它表示随机变量 Z 给定后，从随机变量 Y 所得到的关于随机变量 X 的信息量

定义2.8 平均联合互信息

$$I(X; YZ) = E\{I(x; yz)\} = \sum_x \sum_y \sum_z p(xyz) \log \frac{p(x | yz)}{p(x)}$$

它表示从二维随机变量 YZ 所得到的关于随机变量 X 的信息量。

»» 2.3.3 数据处理定理

定理2.3 (数据处理定理)

如果随机变量 X, Y, Z 构成一个马尔可夫链，则有以下关系成立：

$$I(X; Z) \leq I(X; Y), I(X; Z) \leq I(Y; Z)$$

等号成立的条件是对于任意的 X, Y, Z ，有：

$$p(x|yz) = p(x|z) \quad p(z|xy) = p(z|x)$$

数据处理定理再一次说明，在任何信息传输系统中，最后获得的信息至多是信源所提供的信息，如果一旦在某一过程中丢失一些信息，以后的系统不管如何处理，如不触及丢失信息的输入端，就不能再恢复已丢失的信息，这就是信息不增性原理，它与热熵不减原理正好对应，反映了信息的物理意义。



第三章 信源及信息熵



»» 第三章 信源及信息熵

信源(Information Source)是信息的来源，是产生消息(符号)、时间离散的消息序列(符号序列)以及时间连续的消息的来源。

信源输出的消息都是随机的，因此可用概率来描述其统计特性。

在信息论中，用随机变量 X 、随机矢量 X 、随机过程 $\{X(e, t)\}$ 分别表示产生消息、消息序列以及时间连续消息的信源。

信源的主要问题：

1. 如何描述信源（信源的数学建模问题）
2. 怎样计算信源所含的信息量
3. 怎样有效的表示信源输出的消息，也就是信源编码问题

3.1 信源的分类及其数学模型

信源分类有多种方法，根据信源输出的消息在时间和取值上是离散或连续进行分类：

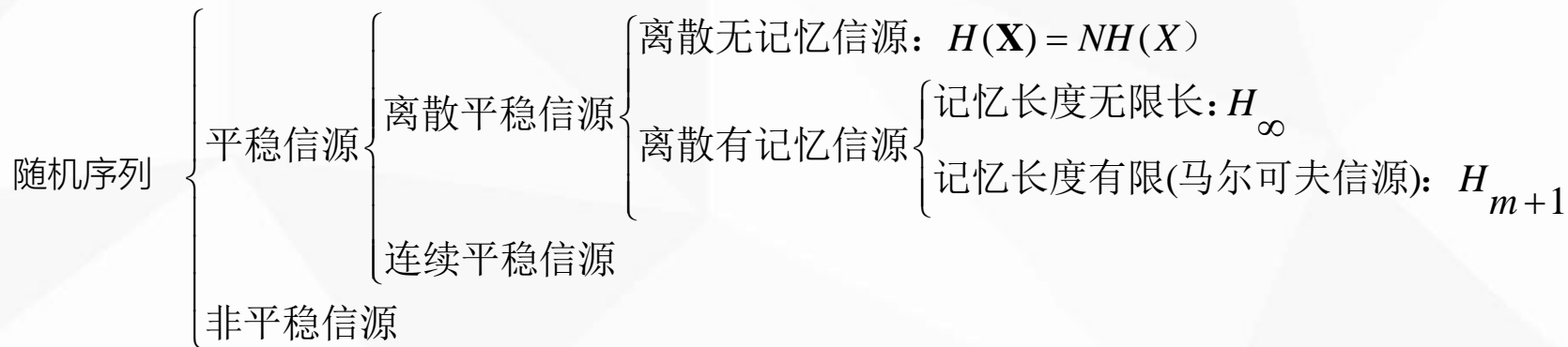
时间（空间）	取值	信源种类	举例	数学描述
离散	离散	离散信源 （数字信源）	文字、数据、 离散化图象	离散随机变量序列
离散	连续	连续信号	跳远比赛的结果、语 音信号抽样以后	连续随机变量序列
连续	连续	波形信源 （模拟信源）	语音、音乐、热噪声、 图形、图象	随机过程
连续	离散		不常见	

表3.1 信源的分类

3.1 信源的分类及其数学模型

我们还可以根据各维随机变量的概率分布是否随时间变化将信源分为**平稳信源**和**非平稳信源**，根据随机变量间是否统计独立将信源分为**有记忆信源**和**无记忆信源**。

一个实际信源的统计特性往往是相当复杂的，要想找到精确的数学模型很困难。实际应用时常常用一些可以处理的数学模型来近似。随机序列，特别是离散平稳随机序列是我们研究的主要内容。



3.2 离散单符号信源

输出单个离散取值的符号的信源称为**离散单符号信源**。它是最简单、最基本的信源，是组成实际信源的基本单元。用离散随机变量表示。

信源所有可能输出的消息及其对应的概率组成的二元序对 $[X, P(X)]$ 称为信源的**概率空间**：

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} X = x_1 \cdots X = x_i \cdots X = x_q \\ p(x_1) \cdots p(x_i) \cdots p(x_q) \end{bmatrix}$$

信源输出的所有消息自信息的统计平均值定义为信源的**平均自信息量（信息熵）**，它表示离散单符号信源的平均不确定性：

$$H(X) = E[-\log p(x_i)] = -\sum_{i=1}^q p(x_i) \log p(x_i)$$

3.2 离散单符号信源

【例3.1】

二元信源 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ p & q \end{bmatrix}$, $p + q = 1$, 求 $H(X)$.

3.3 离散多符号信源

定义3.1: 对于随机变量序列 $X_1, X_2, \dots, X_n, \dots$, 在任意两个不同时刻 i 和 j (i 和 j 为大于1的任意整数)信源发出消息的概率分布完全相同

即对于任意的 $N = 0, 1, 2, \dots$, $X_i X_{i+1} \dots X_{i+N} \dots$ 和 $X_j X_{j+1} \dots X_{j+N} \dots$ 具有相同的概率分布。也就是

$$P(X_i) = P(X_j)$$

$$P(X_i X_{i+1}) = P(X_j X_{j+1})$$

$$\vdots$$

$$P(X_i X_{i+1} \dots X_{i+N}) = P(X_j X_{j+1} \dots X_{j+N})$$

即各维联合概率分布均与时间起点无关的信源称为**离散平稳信源**。

3.2 离散单符号信源

【例3.2】

一平稳信源 X 的符号集 $A = \{0, 1\}$, 产生随机序列, 其中 $P(x_1=0) = p$

(1) 求 $P(X_n=1)$ ($n>1$)的概率。

(2) 若 $P(x_1=0, x_2=1) = b$, 求 $P(x_4=1 \mid x_3=0)$ 的概率。

3.3 离散多符号信源

对于离散多符号信源, 我们引入**熵率**的概念, 它表示信源输出的符号序列中, 平均每个符号所携带的信息量。

定义3.2 随机变量序列中, 对前 N 个随机变量的联合熵求平均:

$$H_N(\mathbf{X}) = \frac{1}{N} H(X_1 X_2 \cdots X_N)$$

称为**平均符号熵**。如果当 $N \rightarrow \infty$ 时上式极限存在, 则 $\lim_{N \rightarrow \infty} H_N(\mathbf{X})$ 称为**熵率**, 或称为**极限熵**, 记为:

$$H_\infty \stackrel{def}{=} \lim_{N \rightarrow \infty} H_N(\mathbf{X})$$

»» 3.3.1 离散平稳无记忆信源

离散平稳无记忆信源输出的符号序列是平稳随机序列，并且符号之间是无关的，即是统计独立的。

假定信源每次输出的是 N 长符号序列，称为离散平稳无记忆信源的 N 次扩展信源，则它的数学模型是 N 维离散随机变量序列： $\mathbf{X} = X_1 X_2 \cdots X_N$ ，并且每个随机变量之间统计独立。

同时，由于是平稳信源，每个随机变量的统计特性都相同，还可把一个输出 N 长符号序列的信源记为：

$$\mathbf{X} = X_1 X_2 \cdots X_N = X^N$$

根据统计独立的多维随机变量联合熵与信息熵之间的关系，可推出：

$$H(\mathbf{X}) = H(X^N) = N H(X)$$

离散平稳无记忆信源的熵率：

$$H_{\infty} = \lim_{N \rightarrow \infty} H_N(\mathbf{X}) = \lim_{N \rightarrow \infty} \frac{1}{N} N H(X) = H(X)$$

3.2 离散单符号信源

【例3.3】

给定离散无记忆信源模型 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$, 求熵率及其二次扩展信源的熵。

»» 3.3.2 离散平稳有记忆信源

实际信源往往是有记忆信源。对于相互间有依赖关系的 N 维随机变量的联合熵存在以下关系（**熵函数的链规则**）：

$$\begin{aligned} H(\mathbf{X}) &= H(X_1 X_2 \cdots X_N) \\ &= H(X_1) + H(X_2 | X_1) + H(X_3 | X_1 X_2) + \cdots + H(X_N | X_1 X_2 \cdots X_{N-1}) \end{aligned}$$

定理3.1 对于离散平稳信源，有以下几个结论：

- (1) 条件熵 $H(X_N | X_1 X_2 \cdots X_{N-1})$ 随 N 的增加是递减的；
- (2) N 给定时平均符号熵大于等于条件熵，即
$$H_N(\mathbf{X}) \geq H(X_N | X_1 X_2 \cdots X_{N-1})$$
- (3) 平均符号熵 $H_N(\mathbf{X})$ 随 N 的增加是递减的；
- (4) 如果 $H(X_1) < \infty$ ，则 $H_\infty = \lim_{N \rightarrow \infty} H_N(\mathbf{X})$ 存在，并且

$$H_\infty = \lim_{N \rightarrow \infty} H_N(\mathbf{X}) = H(X_N | X_1 X_2 \cdots X_{N-1})$$

3.3.2 离散平稳有记忆信源

有一类信源，信源在某时刻发出的符号仅与在此之前发出的有限个符号有关，而与更早些时候发出的符号无关，这称为**马尔可夫性**，这类信源称为**马尔可夫信源**。马尔可夫信源可以在 N 不很大时得到 H_{∞} 。如果信源在某时刻发出的符号仅与在此之前发出的 m 个符号有关，则称为 m 阶马尔可夫信源，它的熵率：

$$H_{\infty} = \lim_{N \rightarrow \infty} H(X_N | X_1 X_2 \cdots X_{N-1}) \quad (\text{马尔可夫性})$$

$$= \lim_{N \rightarrow \infty} H(X_N | X_{N-m} X_{N-m+1} \cdots X_{N-1})$$

$$= H(X_{m+1} | X_1 X_2 \cdots X_m) \quad (\text{平稳性})$$

$$H(X_{m+1} | X_1 X_2 \cdots X_m) \text{ 通常记作 } H_{m+1}$$

3.3.2 离散平稳有记忆信源

【例3.4】

信源 X 的模型 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ \frac{1}{4} & \frac{4}{9} & \frac{11}{36} \end{bmatrix}$ 输出符号序列中，只有前后两个符号有记忆，条件概率 $P(X_2|X_1)$ 列于表3.2。求熵率，并比较 $H(X_2|X_1)$ 、 $1/2H(X_1X_2)$ 和 H 的大小。

3.3.3 马尔可夫信源

马尔可夫信源是一类相对简单的有记忆信源，信源在某一时刻发出某一符号的概率除与该符号有关外，只与此前发出的有限个符号有关。因此我们把前面若干个符号看作一个状态，可以认为信源在某一时刻发出某一符号的概率除了与该符号有关外只与该时刻信源所处的状态有关，而与过去的状态无关。信源发出一个符号后，信源所处的状态即发生改变，这些状态的变化组成了马氏链。

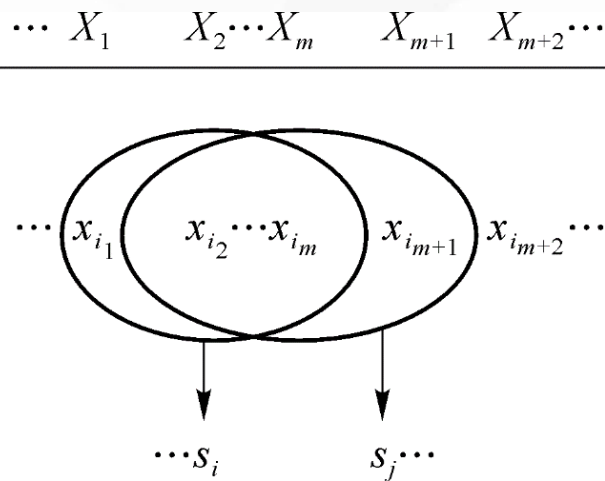


图3.1 马尔可夫信源

»» 3.3.3 马尔可夫信源

【例3.5】

设一个二元一阶马尔可夫信源，信源符号集为 $X = \{0,1\}$ ，信源输出符号的条件概率为：

$$p(0|0) = 0.25, \quad p(0|1) = 0.5, \quad p(1|0) = 0.75, \quad p(1|1) = 0.5$$

求状态转移概率。

3.3.3 马尔可夫信源

【例3.6】

设两个二元二阶马尔可夫信源，信源符号集为 $X = \{0,1\}$ ，信源输出符号的条件概率为 $p(0|00) = p(1|11) = 0.8$, $p(0|01) = p(0|10) = p(1|01) = p(1|10) = 0.5$, $p(1|00) = p(0|11) = 0.2$

求状态转移概率矩阵。

»» 3.3.3 马尔可夫信源

对于一般的 m 阶马尔可夫信源，它的概率空间可以表示成：

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_i & \cdots & x_q \\ p(x_{i_{m+1}} | x_{i_1} x_{i_2} \cdots x_{i_m}) \end{bmatrix}$$

令 $s_i = x_{i_1} x_{i_2} \cdots x_{i_m}$, $i_1, i_2, \cdots, i_m \in \{1, 2, \cdots, q\}$, 从而得到马尔可夫信源的状态空间：

$$\begin{bmatrix} s_1 & \cdots & s_i & \cdots & s_{q^m} \\ p(s_j | s_i) \end{bmatrix}$$

状态空间由所有状态及状态间的状态转移概率组成。通过引入状态转移概率，可以将对马尔可夫信源的研究转化为对马尔可夫链的研究。

3.3.3 马尔可夫信源

下面计算遍历的 m 阶马尔可夫信源的熵率。

当时间足够长后，遍历的马尔可夫信源可以视作平稳信源来处理，又因为 m 阶马尔可夫信源发出的符号只与最近的 m 个符号有关，所以极限熵 H_∞ 等于条件熵 H_{m+1} 。

对于齐次遍历的马尔可夫链，其状态 s_i 由 $x_{i_1} x_{i_2} \cdots x_{i_m}$ 唯一确定，因此有

$$p(s_j | s_i) = p(x_{i_{m+1}} | x_{i_1} x_{i_2} \cdots x_{i_m}) = p(x_{i_{m+1}} | s_i)$$

所以

$$\begin{aligned} H_{m+1} &= H(X_{m+1} | X_1 X_2 \cdots X_m) = E \left[p(x_{i_{m+1}} | x_{i_1} x_{i_2} \cdots x_{i_m}) \right] \\ &= E \left[p(x_{i_{m+1}} | s_i) \right] = - \sum_{i=1}^{q^m} \sum_{i_{m+1}=1}^q p(s_i) p(x_{i_{m+1}} | s_i) \log p(x_{i_{m+1}} | s_i) \\ &= \sum_i p(s_i) H(X | s_i) = - \sum_i \sum_j p(s_i) p(s_j | s_i) \log p(s_j | s_i) \end{aligned}$$

»» 3.3.3 马尔可夫信源

【例3.7】

求图3.3中的二阶马尔可夫信源的极限熵。

3.3.4 信源的相关性和剩余度

根据定理3.1可得 $\log q = H_0 \geq H_1 \geq H_2 \geq \cdots \geq H_{m+1} \geq \cdots \geq H_\infty$

由此看出，由于信源输出符号间的依赖关系也就是信源的相关性使信源的实际熵减小。信源输出符号间统计约束关系越长，信源的实际熵越小。当信源输出符号间彼此不存在依赖关系且为等概率分布时，信源的实际熵等于最大熵 H_0 。

定义3.3 一个信源的熵率(极限熵)与具有相同符号集的最大熵的比值称为熵的相对率：

$$\eta = \frac{H_\infty}{H_0}$$

信源剩余度为：

$$\gamma = 1 - \eta = 1 - \frac{H_\infty}{H_0} = 1 - \frac{H_\infty}{\log q}$$

3.3.4 信源的相关性和剩余度

信源的剩余度来自两个方面，一是信源符号间的相关性，相关程度越大符号间的依赖关系越长，信源的 H_∞ 越小，另一方面是信源符号分布的不均匀性使信源的 H_∞ 越小。

为更经济有效地传送信息，需要尽量压缩信源的剩余度，压缩剩余度的方法就是尽量减小符号间相关性，且尽可能的使信源符号等概率分布。

从提高信息传输效率的观点出发，人们总是希望尽量去掉剩余度。但从提高抗干扰能力角度来看，却希望增加或保留信源的剩余度，因为剩余度大的消息抗干扰能力强。

信源编码是减少或消除信源的剩余度以提高信息的传输效率，而信道编码则通过增加冗余度来提高信息传输的抗干扰能力。

3.3.4 信源的相关性和剩余度

英文字母概率表

字母	概率	字母	概率	字母	概率
空格	0.1859	I	0.0575	R	0.0484
A	0.0642	J	0.0008	S	0.0514
B	0.0127	K	0.0049	T	0.0796
C	0.0218	L	0.0321	U	0.0228
D	0.0317	M	0.0198	V	0.0083
E	0.1031	N	0.0574	W	0.0175
F	0.0208	O	0.0632	X	0.0013
G	0.0152	P	0.0152	Y	0.0164
H	0.0467	Q	0.0008	Z	0.0005

3.3.4 信源的相关性和剩余度

汉字近似概率表

类别	汉字个数	所占概率P	每个汉字的概率 P_i
I	140	0.5	0.5/140
II	625-140=485	$(0.85-0.5) = 0.35$	0.35/485
III	2400-625=1775	$(0.997-0.85) = 0.147$	0.147/1775
IV	7600	0.003	0.003/7600



*3.4 连续信源

3.4.1 连续信源的微分熵

连续随机变量的取值是连续的，一般用概率密度函数描述其统计特征

单变量连续信源的数学模型为 $X: \begin{bmatrix} R \\ p(x) \end{bmatrix}$ ，并满足 $\int_R p(x)dx = 1$ ，

R 是实数域，表示 X 的取值范围。

对于取值范围有限的连续信源还可以表示成： $X: \begin{bmatrix} (a,b) \\ p(x) \end{bmatrix}$ ，并满足 $\int_a^b p(x)dx = 1$ ， (a,b) 是 X 的取值范围。

通过对连续变量的取值进行量化分层，可以将连续随机变量用离散随机变量来逼近。量化间隔越小，离散随机变量与连续随机变量越接近。当量化间隔趋于0时，离散随机变量就变成了连续随机变量。通过这种方式，我们来推导连续随机变量的熵。



*3.4 连续信源

定义连续信源的**微分熵**为： $h(X) = -\int_{\mathcal{R}} p(x) \log p(x) dx$

微分熵又称为**差熵**。虽然已不能代表连续信源的平均不确定性，也不能代表连续信源输出的信息量，但是它具有和离散熵相同的形式，也满足离散熵的主要特性，比如可加性，但是不具有非负性。

同样，我们可以定义两个连续随机变量的**联合熵**：

$$h(XY) = -\iint_{\mathcal{R}^2} p(xy) \log p(xy) dx dy$$

$$h(Y | X) = -\iint_{\mathcal{R}^2} p(xy) \log p(y | x) dx dy$$

以及**条件熵**

$$h(X | Y) = -\iint_{\mathcal{R}^2} p(xy) \log p(x | y) dx dy$$

3.4.2 连续信源的最大熵

离散信源当信源符号为等概分布时有最大熵。连续信源微分熵也有极大值，但是与约束条件有关，当约束条件不同时，信源的最大熵不同。我们一般关心的是下面两种约束下的最大熵。

定理3.1

在输出幅度受限时，服从均匀分布的随机变量 X 具有最大输出熵。

定理3.2

对于平均功率受限的连续随机变量，当服从高斯分布时具有最大熵。

(对于均值为 m ，方差为 σ^2 的连续随机变量，平均功率

$$p = \text{直流功率} + \text{交流功率} = m^2 + \sigma^2 \quad)$$

3.4.3 连续信源的熵功率

均值为零，平均功率限定为 p 的连续信源当服从高斯分布时达到最大熵

$$h(X) = \log \sqrt{2\pi e \sigma^2} = \log \sqrt{2\pi e P}$$

也就是说，高斯信源的熵值与 p 有确定的对应关系： $P = \frac{1}{2\pi e} e^{2h(X)}$

现在假定限定的平均功率为 p ，某连续信源的熵为 $h(X)$ ，则与它具有相同熵的高斯信源的平均功率 \bar{P} 定义为**熵功率**即

$$\bar{P} = \frac{1}{2\pi e} e^{2h(X)}$$

所以， $\bar{P} \leq P$ ，当该连续信源为高斯信源时等号成立。

\bar{P} 的大小可表示连续信源剩余的大小。如果熵功率等于信源平均功率表示信源没有剩余；熵功率和信源平均功率相差越大，说明信源的剩余越大。信源平均功率和熵功率之差($p - \bar{p}$) 称为连续信源的**剩余度**。



第四章 信道及其信道容量



»»» 第四章 信道及其信道容量

信道是指信息传输的通道，包括**空间传输**和**时间传输**。我们在实际通信中所利用的各种物理通道是空间传输信道的最典型的例子，时间传输是指将信息保存，在以后读取，如磁带、光盘等在时间上将信息进行传输的信道。有时我们把为了某种目的而使信息不得不经过的通道也看作信道，这里最关键的是信道有一个输入以及一个与输入有关的输出。至于信道本身的物理结构可能是千差万别的，信息论研究的信道其输入点和输出点在一个实际物理通道中所处位置的选择完全取决于研究的目的。关于信道的主要问题有：

- 1.信道的建模（信道的统计特性的描述）
- 2.信道容量的计算
- 3.在有噪信道中能不能实现可靠传输？怎样实现可靠传输？

»»» 4.1 信道的分类

信息论不研究信号在信道中传输的物理过程，假定信道的传输特性已知，这样信道就可以用图4.1所示的抽象的数学模型来描述。在信息论中，信道通常表示成：

$$\{X, P(Y | X), Y\}$$

即信道输入随机变量 X 、输出随机变量 Y 以及在输入已知的情况下，输出的条件概率分布 $P(Y | X)$ 。

根据实际应用的需要，信道有几种分类方法：

(1) 按其输入/输出信号在幅度和时间上的取值是离散或连续来划分，如表4.1所示：

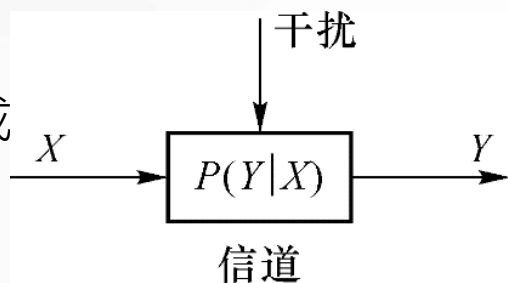


图4.1 信道模型

»»» 4.1 信道的分类

表4.1 按其输入/输出信号在幅度和时间上的取值是离散或连续来划分

幅度	时间	信道名称
离散	离散	离散信道（数字信道）
连续	离散	连续信道
连续	连续	模拟信道（波形信道）
离散	连续	（理论和实用价值均很小）

(2) 按其输入/输出信号之间关系的记忆特性分为**有记忆信道**和**无记忆信道**。

(3) 按输入/输出信号之间的关系是否确定关系分为**有噪声信道**和**无噪声信道**。

»»» 4.1 信道的分类

(4) 另外，根据信道输入和输出的个数可分为

① **两端信道（单用户信道）**：只有一个输入端和一个输出端的单向通信的信道。

② **多端信道（多用户信道）**：双向通信或三个或更多个用户之间相互通信的情况。

本课程主要研究两端信道的情况。

(5) 根据信道的统计特性是否随时间变化分为：

① **恒参信道（平稳信道）**：信道的统计特性不随时间变化。

卫星通信信道在某种意义下可以近似为恒参信道。

② **随参信道（非平稳信道）**：信道的统计特性随时间变化。

如短波通信中，其信道可看成随参信道。

本课程主要研究恒参信道的情况。

»» 4.2 离散单符号信道及其信道容量

4.2.1 离散单符号信道的数学模型

信道的输入、输出都取值于离散符号集，且都用一个随机变量来表示的信道就是**离散单符号信道**。

设离散单符号信道的输入随机变量为 $X, X = \{x_1, x_2, \dots, x_r\}$ ，输出随机变量为 $Y, Y = \{y_1, y_2, \dots, y_s\}$

由于信道中存在干扰，因此输入符号在传输中将产生错误，这种信道干扰对传输的影响可用传递概率 $p(y_j | x_i)$ 描述：

$$p(y_j | x_i) = P\{Y = y_j | X = x_i\} \quad i = 1, 2, \dots, r ; j = 1, 2, \dots, s$$

信道传递概率实际上是一个传递概率矩阵，称为**信道矩阵**，记为：

»»» 4.2 离散单符号信道及其信道容量

$$\mathbf{P} = \begin{matrix} & y_1 & y_2 & \cdots & y_s \\ \begin{matrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{matrix} & \begin{bmatrix} p(y_1 | x_1) & p(y_2 | x_1) & \cdots & p(y_s | x_1) \\ p(y_1 | x_2) & p(y_2 | x_2) & \cdots & p(y_s | x_2) \\ \vdots & \vdots & & \vdots \\ p(y_1 | x_r) & p(y_2 | x_r) & \cdots & p(y_s | x_r) \end{bmatrix} \end{matrix}$$

为了表述简便，常常写成 $\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rs} \end{bmatrix}$

»» 4.2 离散单符号信道及其信道容量

下面推导一般离散单符号信道的一些概率关系：

(1) 输入输出随机变量的联合概率分布为 $P(X = x_i, Y = y_j) = p(x_i y_j)$

则有 $p(x_i y_j) = p(x_i)p(y_j | x_i) = p(y_j)p(x_i | y_j)$

其中, $p(y_j | x_i)$ 是信道传递概率, 即输入为 x_i , 通过信道传输输出 y_j 的概率, 通常称为**前向概率**。它是由于信道噪声引起的, 所以通常用它描述信道噪声的特性。而 $p(x_i | y_j)$ 是已知信道输出符号 y_j , 输入符号为 x_i 的概率, 称为**后向概率**。有时把 $p(x_i)$ 称为输入符号的**先验概率**。而对应的把 $p(x_i | y_j)$ 称为输入符号的**后验概率**。

4.2 离散单符号信道及其信道容量

(2) 由全概率公式, 可从先验概率和信道传递概率求输出符号概率:

$$p(y_j) = \sum_{i=1}^r p(x_i) p(y_j | x_i)$$

写成向量的形式: $[p(y_1) \ p(y_2) \ \cdots \ p(y_s)] = [p(x_1) \ p(x_2) \ \cdots \ p(x_r)] \cdot \mathbf{P}$

或记成 $\mathbf{P}_Y = \mathbf{P}_X \mathbf{P}_{Y|X}$

(3) 根据贝叶斯公式, 可由先验概率和信道的传递概率求后向概率:

$$p(x_i | y_j) = \frac{p(x_i y_j)}{p(y_j)} = \frac{p(x_i) p(y_j | x_i)}{\sum_{i=1}^r p(x_i) p(y_j | x_i)} \quad i = 1, 2, \dots, r ; j = 1, 2, \dots, s$$

»»» 4.2.2 信道容量的概念

平均互信息 $I(X;Y)$ 是接收到输出符号集 Y 后所获得的关于输入符号集 X 的信息量。信源的不确定性为 $H(X)$ 。由于干扰的存在，接收端收到 Y 后对信源仍然存在的不确定性为 $H(Y | X)$ ，又称为**信道疑义度**。信宿所消除的关于信源的不确定性，也就是获得的关于信源的信息为 $I(X;Y)$ ，它是平均意义上每传送一个符号流经信道的信息量，从这个意义上来说，平均互信息又称为信道的**信息传输率**，通常用 R 表示。即

$$R = I(X;Y) = H(X) - H(X | Y) \quad \text{比特/符号}$$

有时我们所关心的是信道在单位时间内平均传输的信息量。如果平均传输一个符号为 t 秒，则信道平均每秒钟传输的信息量为

$$R_t = \frac{1}{t} I(X;Y) \quad \text{比特/秒}$$

一般称为**信息传输速率**。

»» 4.2.2 信道容量的概念

对于固定的信道，总存在一种信源（某种输入概率分布），使信道平均传输一个符号接收端获得的信息量最大，也就是说对于每个固定信道都有一个最大的信息传输率，这个最大的信息传输率即为**信道容量**，而相应的输入概率分布称为**最佳输入分布**。

定义4.1 信道容量为平均互信息对于输入概率分布的最大值：

$$C \stackrel{\text{def}}{=} \max_{p(x)} \{I(X;Y)\}$$

单位依所用的对数底不同可以是比特/符号、奈特/符号等。

若平均传输一个符号需要 t 秒钟，则信道在单位时间内平均传输的最大信息量

$$C_t \stackrel{\text{def}}{=} \max_{p(x)} \{I(X;Y)\}$$

信道容量是信道传送信息的最大能力的度量，信道实际传送的信息量必然不大于信道容量。

»» 4.2.2 信道容量的概念

【例4.1】

输入概率分布 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \omega & \bar{\omega} \end{bmatrix}$, 信道矩阵为 $P = \begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix}$, p 为信道错误传递概率。求二元对称信道的信道容量。

4.2.2 信道容量的概念

图4.3 表示不同的二元对称信道，其传递概率 p 不同，信道容量也不同。

当 $p = 1/2$ 时，是一种最坏的信道，这时 $C = 0$ ，即该信道不能传递任何信息，信息全部损失在信道中了。

而当 $p = 0$ 或 $p = 1$ 时， $C = 1$ ，这是最好的情况，信道能够无失真传送信源信息。

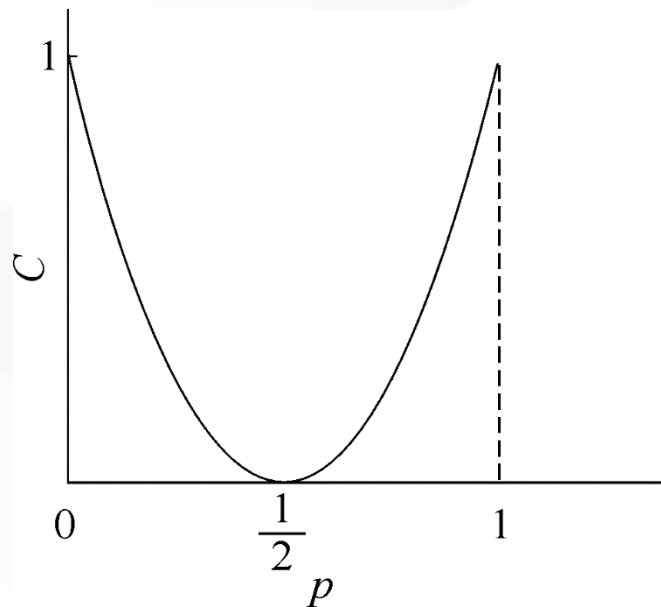


图4.3 BSC的信道容量

4.2.3 几种特殊信道的信道容量

1. 具有扩展性能的无损信道

无损信道是一个输入对应多个输出。

如图4.4所示，信道矩阵为

$$\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

无损信道信道矩阵中每一列只有一个非零元素，接收到信道输出符号后对输入符号将不存在不确定性。

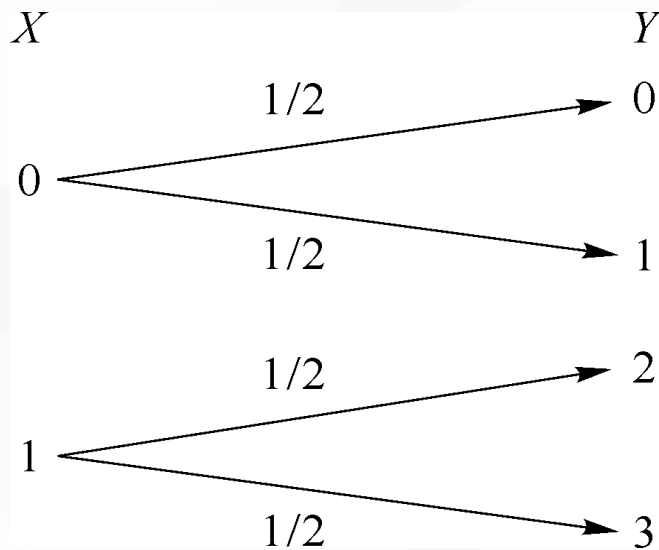


图4.4 无损信道

4.2.3 几种特殊信道的信道容量

2. 具有归并性能的无噪信道

无噪信道是一个输出对应多个输入。

如图4.5所示，信道矩阵为

$$\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

无噪信道每一行只有一个非零元素1，信道矩阵元素非零即1。已知信道输入符号，必能确定输出符号。

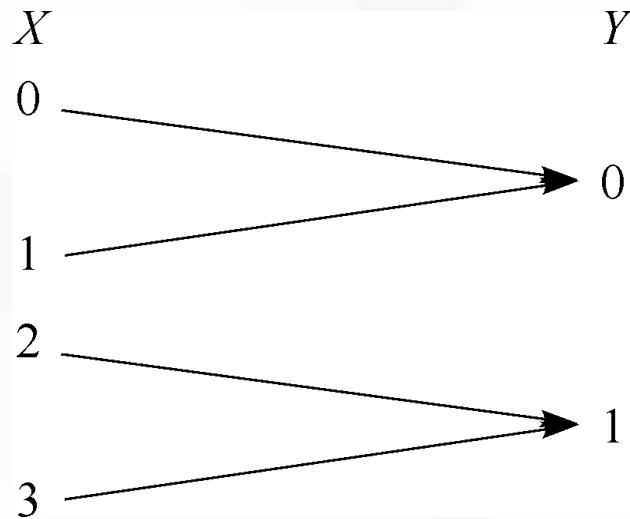


图4.5 无噪信道

4.2.3 几种特殊信道的信道容量

3.具有一一对应关系的无噪无损信道

无噪无损信道输入、输出之间有确定的一一对应关系，即 $y = f(x)$ 。

信道传递概率为：

$$p(y_j | x_i) = \begin{cases} 1 & y_j = f(x_i) \\ 0 & y_j \neq f(x_i) \end{cases}$$

如图4.6所示，信道矩阵为：
$$\mathbf{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

无噪无损信道每一行、每一列只有一个“1”已知 X 后对 Y 不存在不确定性，收到 Y 后对 X 也不存在不确定性。

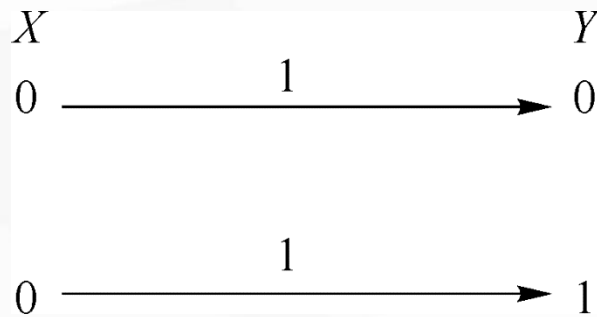


图4.6 无噪无损信道

4.2.4 离散对称信道的信道容量

离散信道中有一类特殊的信道，其特点是信道矩阵具有行对称性，利用这个对称性我们可以简化信道容量的计算。

定义4.2 若信道矩阵中，每行都是第一行元素的不同排列，则称此类信道为**行对称信道**。

定义4.3 若信道矩阵中，不仅每行都是第一行元素的不同排列，而且每列都是第一列元素的不同排列，这类信道称为**对称信道**。

定义4.4 若信道矩阵中，每行都是第一行元素的不同排列，每列并不都是第一列元素的不同排列，但是可以按照信道矩阵的列将信道矩阵划分成若干对称的子矩阵，则称这类信道为**准对称信道**。

4.2.4 离散对称信道的信道容量

定义4.5 若对称信道中输入符号和输出符号个数相同，且信道中总的错误概率为 p ，对称地平均分配给 $r-1$ 个输出符号， r 为输入输出符号的个数，即信道矩阵为

$$\mathbf{P} = \begin{bmatrix} \frac{1-p}{r-1} & \frac{p}{r-1} & \frac{p}{r-1} & \cdots & \frac{p}{r-1} \\ \frac{p}{r-1} & \frac{1-p}{r-1} & \frac{p}{r-1} & \cdots & \frac{p}{r-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{p}{r-1} & \frac{p}{r-1} & \frac{p}{r-1} & \cdots & \frac{1-p}{r-1} \end{bmatrix}$$

则称此信道为**强对称信道**或**均匀信道**。

»»» 4.2.4 离散对称信道的信道容量

二元对称信道就是 $r=2$ 的均匀信道。一般信道的信道矩阵中各行之和为1，但各列之和不一定等于1，而均匀信道中各列之和亦等于1。

定理4.1 对于对称信道，当输入分布为等概分布时，输出分布必能达到等概分布。

定理4.2 若一个离散对称信道具有 r 个输入符号， s 个输出符号，则当输入为等概分布时达到信道容量，且

$$C = \log s - H(p'_1, p'_2, \dots, p'_s)$$

其中 p'_1, p'_2, \dots, p'_s 为信道矩阵中的任一行。

推论：均匀信道的信道容量为 $C = \log r - p \log(r-1) - H(p)$

4.2.4 离散对称信道的信道容量

当输入为等概分布时，输出为等概分布，信道达到信道容量。当 $r=2$ 时的均匀信道常称为**二元对称信道**，这时 $C = 1 - H(p)$ 。

对于一般的离散行对称信道，信道容量 C 仍然可以写成：

$$C = \max_{p(x)} \{H(Y)\} - H(p'_1, p'_2, \dots, p'_s)$$

但不一定存在一种输入分布能使输出达到等概分布，此时信道容量

$$C \leq \log s - H(p'_1, p'_2, \dots, p'_s)$$

而离散对称信道的信道矩阵中每一列都是由同一组元素的不同排列组成所以保证了当输入符号 X 为等概分布时，输出符号 Y 一定也是等概分布输出随机变量熵可以达到 $\log s$ 。

对于离散准对称信道， $C \leq \log s - H(p'_1, p'_2, \dots, p'_s)$

但是可以证明当输入为等概分布时，可以达到信道容量

4.2.4 离散对称信道的信道容量

【例4.2】

某离散对称信道的信道矩阵为 $\mathbf{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \end{bmatrix}$, 求信道容量。

»» 4.2.5 一般离散信道的信道容量

平均互信息 $I(X;Y)$ 是输入概率分布 $p(x)$ 的上凸函数，因此必存在极大值。在信道固定的条件下，平均互信息是 r 个变量 $p(x_i)$, $i=1,2,\dots,r$ 的多元函数，且满足约束条件 $\sum_{i=1}^r p(x_i) = 1$ ，故可用拉格朗日乘子法来求这个条件极值。即在

$$\begin{cases} p(x_i) \geq 0 & i=1,2,\dots,r \\ \sum_i p(x_i) = 1 \end{cases} \quad \text{的条件下求 } I(X;Y) \text{ 的极值。}$$

设辅助函数： $F = I(X;Y) - \lambda \sum_i p(x_i)$ ，当 $\frac{\partial F}{\partial p(x_i)} = 0$ 时求得的 $I(X;Y)$ 的值即为信道容量。

通过计算可得平均互信息的极大值 C ，即 $C = \log e + \lambda$

»» 4.2.5 一般离散信道的信道容量

这样得到的信道容量有一个参数 λ 。在某些情况下可以消去 λ 得到信道容量值。

1. 当输入概率分布只有一个变量时，例如 $r=2$ ，可以设输入概率分布为 α 和 $1-\alpha$ ，因此输入概率分布只有一个变量，这时我们可以直接对 $I(X;Y)$ 求导求出，从而得出 $I(X;Y)$ 的极大值 C 。
2. 对于信道矩阵为可逆矩阵的信道，我们可以采用解方程组的方法。

在一般信道的信道容量的推导中我们推出了下式：

$$\sum_j p(y_j | x_i) \log \frac{p(y_j | x_i)}{p(y_j)} = \log e + \lambda = C \quad i = 1, 2, \dots, r$$

»» 4.2.5 一般离散信道的信道容量

$$\begin{aligned}\text{移项得 } \sum_j p(y_j | x_i) \log p(y_j | x_i) &= \sum_j p(y_j | x_i) \log p(y_j) + C \\ &= \sum_j p(y_j | x_i) [\log p(y_j) + C]\end{aligned}$$

$$\text{令 } \beta_j = \log p(y_j) + C$$

$$\text{则 } \sum_j p(y_j | x_i) \log p(y_j | x_i) = \sum_j p(y_j | x_i) \beta_j$$

当 $r=s$, 且信道矩阵是可逆矩阵时, 该方程组有唯一解。这时就可以出 β_j , 然后根据 $\sum_j p(y_j) = 1$ 求出信道容量:

$$\sum_j 2^{\beta_j - C} = 1 \quad \text{所以} \quad C = \log \sum_j 2^{\beta_j}$$

»»» 4.2.5 一般离散信道的信道容量

由 β_j 和 C 就可以求得输出概率分布: $p(y_j) = 2^{\beta_j - C}$

再根据 $p(y_j) = \sum_i p(x_i) p(y_j | x_i)$ 列方程组求 $p(x_i)$

将计算步骤总结如下:

(1) 由 $\sum_j p(y_j | x_i) \beta_j = \sum_j p(y_j | x_i) \log p(y_j | x_i)$ 列方程组求出 β_j ;

(2) 由 $C = \log \sum_j 2^{\beta_j}$ 求出 C ;

(3) 由 $p(y_j) = 2^{\beta_j - C}$ 求出 $p(y_j)$;

(4) 由 $p(y_j) = \sum_i p(x_i) p(y_j | x_i)$ 列方程组求 $p(x_i)$ 。

➤➤➤ 4.2.5 一般离散信道的信道容量

【例4.4】 $P = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$, 求信道容量。

➤➤➤ 4.2.5 一般离散信道的信道容量

【例4.5】 $\mathbf{P} = \begin{bmatrix} 0.5 & 0.3 & 0.2 \\ 0.3 & 0.5 & 0.2 \end{bmatrix}$, 求信道容量。

»» 4.2.6 信道容量定理

从以上的讨论可知，求信道容量的问题实际上是在约束条件下求多元函数极值问题，在通常情况下，计算量是非常大的。

下面我们介绍一般离散信道的平均互信息 $I(X;Y)$ 达到信道容量的充要条件，在某些情况下它可以帮助我们较快地找到极值点。

定理4.3 设有一般离散信道，它有 r 个输入信号， s 个输出信号。当且仅当存在常数 C 使输入分布 $p(x_i)$ 满足：

- (1) $I(x_i;Y) = C$ 当 $p(x_i) \neq 0$
- (2) $I(x_i;Y) \leq C$ 当 $p(x_i) = 0$ 时， $I(X;Y)$ 达到最大值。

其中，

$$I(x_i;Y) = \sum_j p(y_j | x_i) \log \frac{p(y_j | x_i)}{p(y_j)}$$

它表示信道输入 x_i 时，所给出关于输出 Y 的信息量。常数 C 即为所求的信道容量。

»»» 4.2.6 信道容量定理

信道容量定理告诉我们，平均互信息 $I(X;Y)$ 取到极大值也就是达到信道容量时，对于任意 x_i ，只要它出现的概率大于0， $I(x_i;Y)$ 都相等。

信道容量定理只给出了达到信道容量时，最佳输入概率分布应满足的条件，并没有给出最佳输入概率分布值，也没有给出信道容量的数值

另外，定理本身也隐含着达到信道容量的最佳分布不一定是唯一的，只要输入概率分布满足充要条件式，就是信道的最佳输入分布。在一些特殊情况下，我们常常利用这一定理寻求输入分布和信道容量值。

4.3 离散多符号信道及其信道容量

实际离散信道的输入和输出常常是随机变量序列，用随机矢量来表示称为**离散多符号信道**，如图4.8所示。实际离散信道往往是有记忆信道、为了简化起见，我们主要研究离散无记忆信道。

定义4.6 若在任意时刻信道的输出只与此时刻信道的输入有关，而与其他时刻的输入和输出无关，则称之为**离散无记忆信道**，简称为DMC(discrete memoryless channel)。

输入、输出随机序列的长度为 N 的离散无记忆平稳信道通常称为离散无记忆信道的 N 次扩展信道。

N 次扩展信道的信道矩阵是一个 $r^N \times s^N$ 的矩阵。

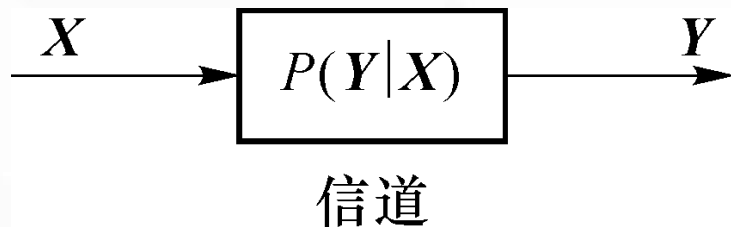


图4.8 离散多符号信道

»» 4.3 离散多符号信道及其信道容量

离散无记忆信道的数学模型仍然表示为： $\{\mathbf{X}, P(\mathbf{Y}|\mathbf{X}), \mathbf{Y}\}$ ，注意：这时输入、输出均为随机矢量。

根据信道无记忆的特性，其转移概率

$$\begin{aligned} P(\mathbf{Y} | \mathbf{X}) &= P(Y_1 Y_2 \cdots Y_N | X_1 X_2 \cdots X_N) \\ &= P(Y_1 | X_1) P(Y_2 | X_2) \cdots P(Y_N | X_N) = \prod_{k=1}^N P(Y_k | X_k) \end{aligned}$$

定理4.4 若信道的输入和输出分别是 N 长序列 \mathbf{X} 和 \mathbf{Y} ，且信道是无记忆的，则存在

$$I(\mathbf{X}; \mathbf{Y}) \leq \sum_{k=1}^N I(X_k; Y_k)$$

这里 X_k 、 Y_k 分别是序列 \mathbf{X} 和 \mathbf{Y} 中第 k 位随机变量。

»» 4.3 离散多符号信道及其信道容量

【例4.8】

求二元对称信道的二次扩展信道的信道矩阵。

4.3 离散多符号信道及其信道容量

对于离散无记忆 N 次扩展信道，当信源是平稳无记忆信源时，其平均互信息 $I(\mathbf{X};\mathbf{Y})$ 等于单符号信道的平均互信息的 N 倍。

离散无记忆信道的 N 次扩展信道的信道容量为

$$C^N = \max_{P(\mathbf{X})} I(\mathbf{X};\mathbf{Y}) = \max_{P(\mathbf{X})} \sum_{k=1}^N I(X_k;Y_k) = \sum_{k=1}^N \max_{P(X_k)} I(X_k;Y_k) = \sum_{k=1}^N C_k$$

因为现在输入随机序列在同一信道中传输，所以任何时刻通过离散无记忆信道传输的最大信息量都相同，即 $C_k = C$ ($k = 1, 2, \dots, N$)

所以 $C^N = NC$

当信源也是无记忆信源并且每一时刻的输入分布各自达到最佳输入分布时，才能达到这个信道容量 NC 。

一般情况下，消息序列在离散无记忆 N 次扩展信道中传输时，其平均互信息量 $I(\mathbf{X};\mathbf{Y}) \leq NC$

»»» 4.4 组合信道及其信道容量

前面我们分析了离散单符号信道和离散无记忆信道的扩展信道。实际应用中常常会遇到两个或更多个信道组合在一起使用的情况。例如，待发送的消息比较多时，可能要用两个或更多个信道并行发送，这种组合信道称为**并联信道**；有时消息会依次地通过几个信道串联发送，例如无线电中继信道，数据处理系统，这种组合信道称为**级联信道**。在研究较复杂信道时，为使问题简化，往往可以将它们分解成几个简单的信道的组合。这一节我们将讨论这两种组合信道的信道容量与其组成信道的信道容量间关系。

4.4.1 独立并联信道

一般独立并联信道如图4.10所示。

可以把定理4.4的结论推广到 N 个独立并联信道中来：

$$C_{\text{并}} = \max_{P(X_1 \cdots X_N)} I(X_1 \cdots X_N; Y_1 \cdots Y_N) \leq \sum_{k=1}^N C_k$$

只有当每个输入随机变量的概率分布均达到各自信道的最佳输入分布时，独立并联信道的信道容量才等于各信道容量之和：

$$C_{\text{并}} = \sum_{k=1}^N C_k$$

当 N 个独立并联信道信道容量都相同时，

$$C_{\text{并}} = NC$$

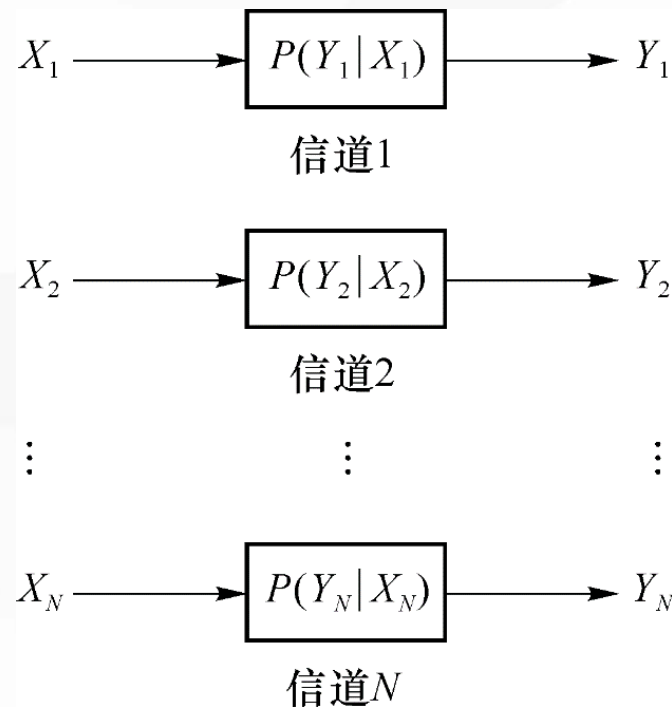


图4.10 独立并联信道

4.4.2 级联信道

级联信道是信道最基本的组合形式，许多实际信道都可以看成是其组成信道的级联。图4.11是由两个单符号信道组成的最简单的级联信道

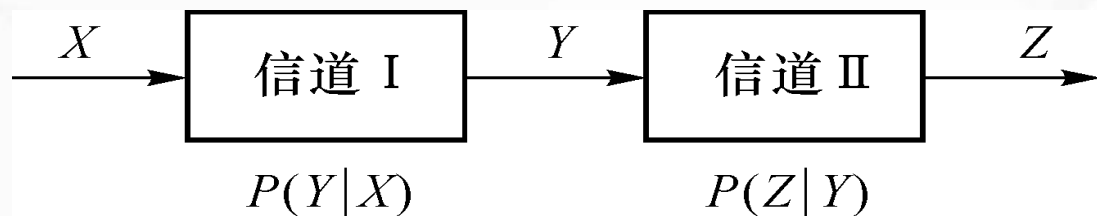


图4.11 级联信道

$X \rightarrow Y \rightarrow Z$ 组成一个马尔可夫链。根据马尔可夫链的性质，级联信道的总的信道矩阵等于这两个串接信道的信道矩阵的乘积。求得级联信道的总的信道矩阵后，级联信道的信道容量就可以用求离散单符号信道的信道容量的方法计算。

➤➤➤ *4.5 连续信道及其信道容量

4.5.1 连续随机变量的互信息

连续随机变量和之间的平均互信息定义为

$$I(X;Y) = \iint_{R^2} p(xy) \log \frac{p(xy)}{p(x)p(y)} dx dy$$

连续随机变量的平均互信息具有和离散随机变量的平均互信息一样的性质：

- 1. 对称性：** $I(X;Y) = I(Y;X) = h(X) - h(X|Y)$
 $= h(Y) - h(Y|X) = h(X) + h(Y) - h(XY)$
- 2. 非负性：** $I(X;Y) \geq 0$ 当且仅当随机变量和统计独立时等号成立

4.5.2 高斯加性信道的信道容量

高斯信道是最差的信道，实际应用中往往把噪声视为高斯噪声。

噪声源为高斯白噪声的加性信道。其信道容量

$$C = \max_{p(x)} I(X; Y) = \max_{p(x)} [h(Y) - h(Y | X)] = \max_{p(x)} [h(Y) - h(N)]$$

如果噪声 N 是均值为0、方差为 σ^2 的高斯随机变量，即

$$\int_{-\infty}^{+\infty} p(n) dn = 1$$

$$\int_{-\infty}^{+\infty} np(n) dn = 0$$

$$\int_{-\infty}^{+\infty} n^2 p(n) dn = \sigma_N^2 = P_N$$

P_N 表示噪声 N 的平均功率。这种信道称为**高斯加性连续信道**。

4.5.2 高斯加性信道的信道容量

当输入随机变量 X 的概率密度是均值为0、方差为 σ_X^2 的高斯随机变量，加性信道的噪声 N 是均值为0、方差为 σ_N^2 的高斯随机变量时，输出随机变量 Y 也是一个高斯随机变量，其均值为0、方差为 $\sigma_Y^2 = \sigma_X^2 + \sigma_N^2 = P_Y$ 此时输出随机变量的熵 $h(Y)$ 达到最大，而信道达到信道容量：

$$\begin{aligned} C &= \max_{p(x)} h(Y) - h(N) = \frac{1}{2} \log 2\pi e(\sigma_X^2 + \sigma_N^2) - \frac{1}{2} \log 2\pi e\sigma_N^2 \\ &= \frac{1}{2} \log \frac{\sigma_X^2 + \sigma_N^2}{\sigma_N^2} = \frac{1}{2} \log \left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right) = \frac{1}{2} \log \left(1 + \frac{P_X}{P_N}\right) \end{aligned}$$

其中， $\frac{P_X}{P_N}$ 称为信道的信噪比。

4.5.3 多维高斯加性信道的信道容量

当信道为多维高斯加性信道时，由于加性噪声信道必然是一个无记忆

信道，所以 $I(\mathbf{X}; \mathbf{Y}) \leq \sum_{i=1}^n I(X_i; Y_i) \leq \frac{1}{2} \sum_{i=1}^n \log(1 + \frac{P_{X_i}}{P_{N_i}})$

$$\text{因此 } C = \max_{p(\mathbf{x})} I(\mathbf{X}; \mathbf{Y}) = \max_{p(\mathbf{x})} \sum_{i=1}^n I(X_i; Y_i) = \frac{n}{2} \log(1 + \frac{P_{X_i}}{P_{N_i}}), i = 1, 2, \dots, n$$

当且仅当输入随机矢量 \mathbf{X} 中各分量统计独立，并且均为高斯变量时达到信道容量。

如果在每个抽样时刻信源和噪声是均值为0、方差分别为 σ_N^2 和 σ_X^2 的高斯随机变量，

$$\text{则 } C = \frac{n}{2} \log(1 + \frac{\sigma_X^2}{\sigma_N^2}) \quad \text{比特/n个样值}$$

*** *4.6 波形信道及其信道容量

波形信道通常根据抽样定理转化成多维连续信道进行处理。

一般来说，信道的带宽总是有限的。假设某信道的频带限于 $(0, B)$ 则其输入、输出信号和噪声都是限频的随机过程，频带限于 $(0, B)$ 根据抽样定理，可把一个时间连续的信道变换成时间离散的随机序列信道来处理，即用每隔 $\frac{1}{2B}$ 秒时间的采样值来表示输入、输出信号和噪声。我们把一次采样看成信道的 $\frac{1}{2B}$ 次传输，由于每秒传送 $2B$ 个样值，所以单位时间的信道容量为

$$C_t = B \log\left(1 + \frac{\sigma_X^2}{\sigma_N^2}\right) \quad \text{比特/秒}$$

当噪声是双边功率谱密度为 $\frac{N_0}{2}$ 的高斯白噪声时， $C_t = B \log\left(1 + \frac{\sigma_X^2}{N_0 B}\right)$

*** *4.6 波形信道及其信道容量

这就是著名的**香农公式**，它适用于加性高斯白噪声信道。从前面的讨论可知，只有当输入信号为功率受限的高斯白噪声信号时，才能达到该信道容量。

香农公式说明，当信道容量一定时，增大信道的带宽，可以降低对信噪功率比的要求；反之，当信道频带较窄时，可以通过提高信噪功率比来补偿。

当 $B \rightarrow \infty$ 时，则

$$C_t = \lim_{B \rightarrow \infty} B \log\left(1 + \frac{\sigma_x^2}{N_0 B}\right) = \lim_{B \rightarrow \infty} \frac{\sigma_x^2}{N_0} \frac{N_0 B}{\sigma_x^2} \log\left(1 + \frac{\sigma_x^2}{N_0 B}\right) = \frac{\sigma_x^2}{N_0} \cdot \log e = 1.44 \frac{\sigma_x^2}{N_0}$$

上式表明当频带很宽时，信道容量正比于信号功率与噪声谱密度之比
上式是加性高斯噪声信道信息传输率的极限值。



第五章 无失真信源编码



»» 5.1 信源编码的相关概念

5.1.1 编码器

信源输出的符号序列，需要变换成适合信道传输的符号序列，一般称为**码序列**，对信源输出的原始符号按照一定的数学规则进行的这种变换称为**编码**，完成编码功能的器件，称为**编码器**。接收端有一个**译码器**完成相反的功能。

信源编码器的输入是信源符号集 $S = \{s_1, s_2, \dots, s_q\}$ ，共有 q 个信源符号。同时存在另一个符号集 $X = \{x_1, x_2, \dots, x_r\}$ ，称为码符号集，共有 r 个码符号，码符号集中的元素称为**码元**或**码符号**，编码器的作用就是将信源符号集 S 中的符号 $s_i, i = 1, 2, \dots, q$ 变换成由 l_i 个码符号组成的一一对应的码符号序列。编码器输出的码符号序列称为**码字**，并用 $w_i, i = 1, 2, \dots, q$ 来表示，它与信源符号 $s_i, i = 1, 2, \dots, q$ 之间是一一对应的关系，如图5.1所示。

»» 5.1 信源编码的相关概念

码字的集合 C 称为**码**，即 $C = \{w_1, w_2, \dots, w_q\}$ 。信源符号 s_i 对应的码字 w_i 包含 L_i 个码符号， L_i 称为**码字长度**，简称**码长**。

所以，信源编码就是把信源符号序列变换到码符号序列的一种映射。若要实现**无失真编码**，那么这种映射必须是一一对应的、可逆的。一般来说，人们总是希望把信源所有的信息毫无保留地传递到接收端，即实现无失真传递，所以首先要对信源实现无失真编码。

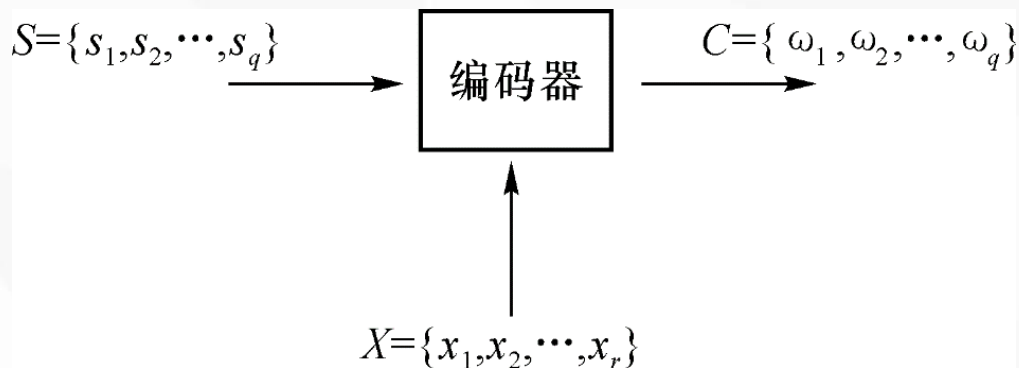


图5.1 信源编码器

»»» 5.1 信源编码的相关概念

信源编码有以下3种主要方法：

(1) 匹配编码 根据信源符号的概率不同，编码的码长不同：概率大的信源符号，所编的代码短；概率小的信源符号所编的代码长，这样使平均码长最短。将要讲述的香农编码、哈夫曼编码、费诺码都是概率匹配编码，都是无失真信源编码。

(2) 变换编码 先对信号进行变换，从一种信号空间变换为另一种信号空间，然后针对变换后的信号进行编码。

(3) 识别编码 识别编码主要用于印刷或打字机等有标准形状的文字符号和数据的编码，比如中文文字和语音的识别。

后两种信源编码均为有失真的信源编码。

无失真信源编码主要针对离散信源，连续信源在量化编码的过程中必然会有量化失真，所以对连续信源只能近似地再现信源的消息。

»» 5.1.2 码的分类

1. 分组码和非分组码

定义5.1 将信源符号集中的每个信源符号 s_i 固定地映射成一个码字 w_i ，这样的码称为**分组码**。

用分组码对信源符号进行编码时，为了使接收端能够迅速准确地将码译出，分组码必须具有有一些直观属性。与分组码对应的是**非分组码**，又称为**树码**、树码编码器输出的码符号通常与编码器的所有信源符号都有关。

2. 奇异码与非奇异码

定义5.2 若一种分组码中的所有码字都不相同，则称此分组码为**非奇异码**，否则称为**奇异码**。

»» 5.1.2 码的分类

3. 唯一可译码与非唯一可译码

定义5.3 任意有限长的码元序列，如果只能唯一地分割成一个个码字，便称为**唯一可译码**。

唯一可译码的物理含义是指不仅要求不同的码字表示不同的信源符号，而且还要求对由信源符号构成的符号序列进行编码时，在接收端仍能正确译码而不发生混淆。唯一可译码首先是非奇异码，且任意有限长的码字序列不会雷同。

4. 即时码与非即时码

定义5.4 无需考虑后续的码符号就可以从码符号序列中译出码字，这样的唯一可译码称为**即时码**。

»» 5.1.2 码的分类

下面讨论唯一可译码成为即时码的条件。

定义5.5 设 $w_i = x_{i_1}x_{i_2}\cdots x_{i_l}$ 为一码字，对于任意的 $1 \leq j \leq l$ ，称码符号序列的前 j 个元素 $x_{i_1}x_{i_2}\cdots x_{i_j}$ 为码字的前缀。按照上述的前缀的定义，有下述结论：

定理5.1 一个唯一可译码成为即时码的充要条件是其中任何一个码字都不是其他码字的前缀。

即时码可以用树图来构造。

图5.2是一个二元即时码的树图。

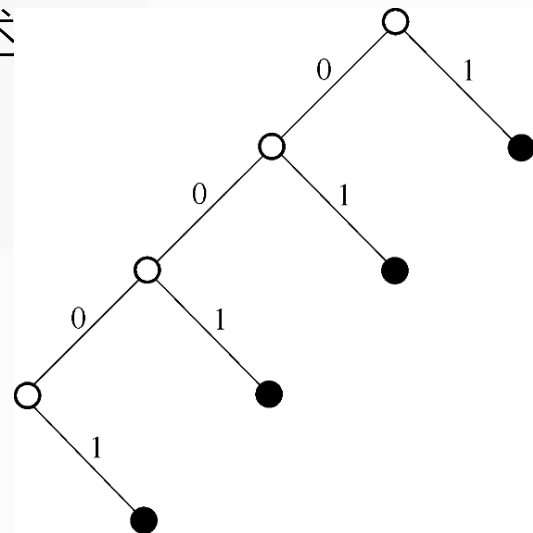


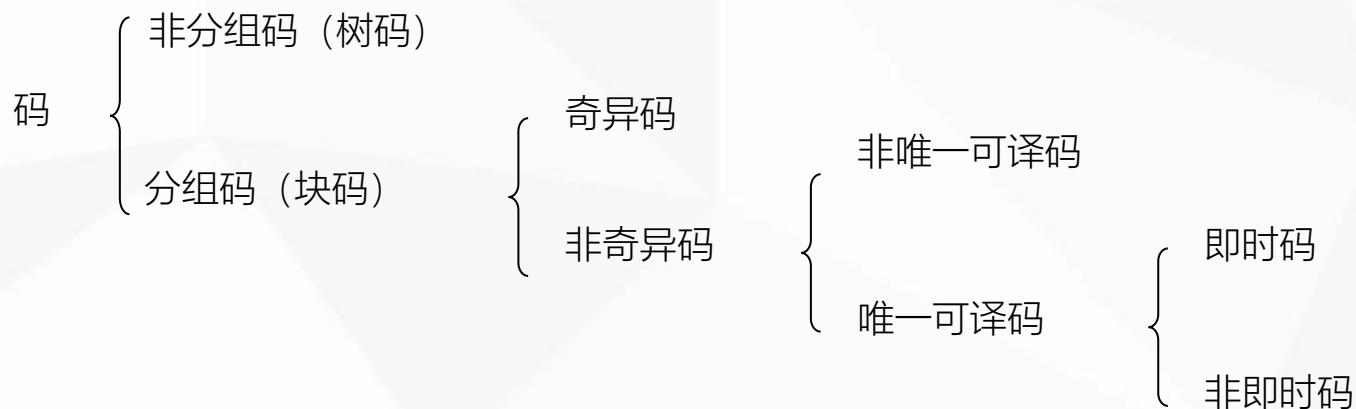
图5.2 二元即时码的树图

»» 5.1.2 码的分类

树是没有回路的图，所以它也是由节点和弧构成的。树中最顶部的节点称为**根节点**，没有子节点的节点称为**终端节点**。

所有根节点的子节点称为**一阶节点**，所有一阶节点的子节点称为**二阶节点**，依此类推。 n 阶节点最多有 r^n 个。节点的阶次又称为节点的**深度**。

综上所述，可将信源编码作如下分类：



»» 5.2 定长码及定长编码定理

若对一个有 q 个信源符号的信源 S 进行定长编码，那么信源 S 存在唯一可译定长码的条件是：

$$q \leq r^l \quad (5.1)$$

其中， r 是码符号集中的码元数， l 是定长码的码长。

如果对信源 S 的 N 次扩展信源 S^N 进行定长编码，若要编得的定长码是唯一可译码，则必须满足

$$q^N \leq r^l \quad (5.2)$$

其中， q 是信源 S 的符号个数， q^N 是信源 S 的 N 次扩展信源 S^N 的符号个数， r 是码符号集 X 的码符号数。

»» 5.2 定长码及定长编码定理

定长编码的信息传输效率是很低的。提高信息传输效率的方法有：

方法1：考虑符号之间的依赖关系，对信源 S 的扩展信源进行编码。

方法2：对于概率等于0 或非常小的符号序列不予编码。

定理 5.2 离散无记忆信源的熵为 $H(S)$ ，若对信源长为 N 的序列进行定长编码，码符号集 X 中有 r 个码符号，码长为 l ，则对于任意 $\varepsilon > 0$ ，只要满足：
$$\frac{l}{N} \geq \frac{H(S) + \varepsilon}{\log r}$$

则当 N 足够大时，可实现几乎无失真编码，即译码错误概率 δ 任小；

反之，如果 $\frac{l}{N} \leq \frac{H(S) - 2\varepsilon}{\log r}$ ，则不可能实现几乎无失真编码，即当 N 足够大时，译码错误概率为1。

»» 5.2 定长码及定长编码定理

定理5.2是在离散平稳无记忆信源的条件下证明的，但它同样适合于平稳有记忆信源，只要将式中 $H(S)$ 改为极限熵 $H_{\infty}(S)$ 即可，条件是有记忆信源的极限熵 $H_{\infty}(S)$ 和极限方差 $\sigma_{\infty}^2(S)$ 存在。

定长信源编码定理给出了定长编码时每个信源符号最少所需的码符号的理论极限，该极限值由 $H(S)$ 决定。

为了衡量编码效果，引进编码效率。

»» 5.2 定长码及定长编码定理

定义5.6 定义 $\eta = \frac{H(S)}{R'} = \frac{H(S)}{\frac{l}{N} \log r}$ 为**编码效率**。

由定理5.2可得最佳编码效率为 $\eta = \frac{H(S)}{H(S) + \varepsilon}, \varepsilon > 0$, 所以 $\varepsilon = \frac{1-\eta}{\eta} H(S)$

在已知方差和信源熵的条件下, 信源符号序列长度 N 与最佳编码效率 η 和允许错误概率 δ 的关系为:

$$N \geq \frac{D[I(s_i)]}{\varepsilon^2 \delta} = \frac{D[I(s_i)]}{H^2(S)} \frac{\eta^2}{(1-\eta)^2 \delta}$$

当允许错误概率越小, 编码效率又要高, 那么信源符号序列长度 N 必须越长. 在实际情况下, 要实现几乎无失真的定长编码, N 需要的长度将会大到难以实现。

»» 5.2 定长码及定长编码定理

【例5.2】 设有离散无记忆信源

$$\begin{bmatrix} S \\ P \end{bmatrix} = \begin{bmatrix} s_1 & s_2 & s_3 & s_4 & s_5 & s_6 & s_7 & s_8 \\ 0.4 & 0.18 & 0.1 & 0.1 & 0.07 & 0.06 & 0.05 & 0.04 \end{bmatrix}$$
, 如果对信源符号采用定长二元编码, 要求编码效率 $\eta = 90\%$ 。允许错误率 $\delta \leq 10^{-6}$, 求所需信源序列长度。

»» 5.3 变长码及变长编码定理

5.3.1 Kraft不等式和McMillan不等式

定理5.3 设信源符号集为 $S = \{s_1, s_2, \dots, s_q\}$, 码符号集为 $X = \{x_1, x_2, \dots, x_r\}$

对信源进行编码, 得到的码为 $C = \{w_1, w_2, \dots, w_q\}$, 码长分别为 l_1, l_2, \dots, l_q

即时码存在的充要条件是:

$$\sum_{i=1}^q r^{-l_i} \leq 1 \quad (5.4)$$

这称为**Kraft不等式**。

由Kraft不等式可知, 给定 r 和 q , 只要允许码字长度可以足够长, 则码长总可以满足Kraft不等式而得到即时码, Kraft不等式指出了即时码的码长必须满足的条件。

»» 5.3 变长码及变长编码定理

对于唯一可译码，该不等式又称为McMillan不等式。

定理5.4 唯一可译码存在的充要条件是
$$\sum_{i=1}^q r^{-l_i} \leq 1$$

r 为码符号个数，为码字长度， q 为信源符号个数。

定理5.4指出了唯一可译码中 r 、 q 、 l_i 之间的关系，如果满足这个不等式的条件，则一定能够构成至少一种唯一可译码，否则，无法构成唯一可译码。它给出了唯一可译变长码的存在性。

另外，从定理5.3和定理5.4可以得到一个重要的结论，即任何一个唯一可译码均可用一个相同码长的即时码来代替，因为即时码很容易用树图法构造，因此要构造唯一可译码，只要构造即时码就可以了。

5.3.2 唯一可译码的判别准则

A.A.Sardinas和G.W.Patterson于1957年提出下述算法用于判断码 C 的唯一可译性. 此算法的原理如下所示:

A_1	A_2	$A_3 \cdots A_m$
B_1	B_2	$B_3 \cdots B_n$

其中 A_i, B_i 都是码字. 可知, 当且仅当某个有限长的码符号序列能译成两种不同的码字序列时, 此码不是唯一可译码, 此时 B_1 一定是 A_1 的前缀, 而 A_1 的尾随后缀一定是另一码字 B_2 的前缀; 而 B_2 的尾随后缀又是其他码字的前缀. 最后, 码符号序列的尾部一定是一个码字。

»» 5.3.2 唯一可译码的判别准则

设 C 为码字集合，按以下步骤构造此码的尾随后缀集合 F ：

- (1) 考查 C 中所有的码字，若 W_i 是 W_j 的前缀，则将相应的后缀作为一个尾随后缀码放入集合 F_1 中；
- (2) 考查 C 和 F_i 两个集合，若 $W_i \in C$ 是 $W_i \in F_i$ 的前缀或 $W_i \in F_i$ 是 $W_j \in C$ 的前缀，则将相应的后缀作为尾随后缀码放入集合 F_{i+1} 中；
- (3) $F = \bigcup_i F_i$ 即为码 C 的尾随后缀集合；
- (4) 若 F 中出现了 C 中的元素，则算法终止，返回假(C 不是唯一可译码)，否则若 F_{i+1} 为空或者 F_{i+1} 中没有出现新的元素，则返回真。

»» 5.3.3 无失真变长编码定理

定义5.7 设有信源

$$\begin{bmatrix} S \\ P \end{bmatrix} = \begin{bmatrix} s_1 & s_2 & \cdots & s_q \\ p(s_1) & p(s_2) & \cdots & p(s_q) \end{bmatrix}$$

编码后的码字分别为 w_1, w_2, \dots, w_q ，各码字相应的码长分别为 l_1, l_2, \dots, l_q 因为是唯一可译码，信源符号 s_i 和码字 w_i 一一对应，则定义此码的**平均码长**为：

$$\bar{L} = \sum_{i=1}^q p(s_i) l_i \quad \text{码符号/信源符号}$$

其中， \bar{L} 表示每个信源符号平均需用的码元数。

当信源给定时，信源熵 $H(S)$ 就确定了，而编码后每个信源符号平均用 \bar{L} 个码元来表示，编码后平均每个码元载荷的信息量 $H(X)$ 定义为编码后的信息传输率。

»» 5.3.3 无失真变长编码定理

定义5.8 编码后信源的信息传输率为 $R = H(X) = \frac{H(S)}{\bar{L}}$ 比特/码符号

如果传输一个码符号平均需要 t 秒时间，则编码后信源每秒提供信息量为

$$R_t = \frac{H(S)}{\bar{L}_t} \quad \text{比特/秒}$$

可以看出， \bar{L} 越小，则 R 越大，信息传输率就越高，因此我们感兴趣的码是使平均码长 \bar{L} 最短的码。

定义5.9 对于给定的信源和码符号集，若有一个唯一可译码，其平均码长 \bar{L} 小于所有其他唯一可译码，则称这种码为**紧致码**或**最佳码**。无失真信源编码的核心问题就是寻找紧致码。

»» 5.3.3 无失真变长编码定理

下面的定理给出了紧致码的平均码长可能达到的理论极限。

定理5.5 若一个离散无记忆信源 S ，熵为 $H(S)$ ，码符号集 $X = \{x_1, x_2, \dots, x_r\}$ 对 S 进行无失真编码，则总可以找到一种唯一可译码，其平均码长满足：

$$\frac{H(S)}{\log r} \leq \bar{L} \leq 1 + \frac{H(S)}{\log r} \quad (5.5)$$

还可看到平均码长极限值与无失真定长信源编码定理中极限值是一致的

»» 5.3.4 香农第一编码定理

定理5.6 设离散无记忆信源 S 的信源熵 $H(S)$ ，它的 N 次扩展信源 $S^N = \{s_1, s_2, \dots, s_n\}$ ，其熵 $H(S^N)$ 。并用码符号 $X = \{x_1, x_2, \dots, x_r\}$ 对信源进行编码，总可以找到一种唯一可译码，使每个信源符号所需的平均码长满足：

$$\frac{H(S)}{\log r} \leq \frac{\bar{L}_N}{N} \leq \frac{H(S)}{\log r} + \frac{1}{N}$$

或者写成：

$$H_r(S) \leq \frac{\bar{L}_N}{N} \leq H_r(S) + \frac{1}{N}$$

定理5.5可以看作定理5.6在 $N=1$ 时的特殊情况。

定理5.6是香农信息论的主要定理之一。

»» 5.3.4 香农第一编码定理

编码后信道的**信息传输率**为：

$$R = \frac{H(S)}{\bar{L}} \frac{\text{比特/信源符号}}{\text{码符号/信源符号}} = \frac{H(S)}{\bar{L}} \text{ 比特/码符号}$$

这里， $\bar{L} = \frac{\bar{L}_N}{N} \geq \frac{H(S)}{\log r}$ ，代入上式，得到 $R \leq \log r$

无失真信源编码的实质就是对离散信源进行适当的变换，使变换后新的码符号信源(信道的输入)尽可能为等概率分布，使新信源的每个码符号平均所含的信息量达到最大，从而使信道的信息传输率 R 达到信道容量实现信源与信道理想的统计匹配。这就是香农第一定理的物理意义。

»» 5.3.4 香农第一编码定理

为了衡量各种编码是否已达到极限情况，我们定义变长码的编码效率。

定义5.12 设对信源 S 进行无失真编码所得到的平均码长为 \bar{L} ，则 $\bar{L} \geq H_r(S)$ ，
定义：

$$\eta = \frac{H_r(S)}{\bar{L}}$$

为**编码效率** $\eta \leq 1$

对同一信源来说，码的平均码长 \bar{L} 越短，越接近极限 $H_r(S)$ ，信道的信息传输率越高，越接近无噪无损信道的信道容量，这时 η 也越接近于1

所以用码的效率 η 来衡量各种编码的优劣。

»» 5.3.4 香农第一编码定理

另外，为了衡量各种编码与最佳码的差距，引入码的剩余度的概念。

定义5.11 定义：

$$\gamma = 1 - \eta = 1 - \frac{H_r(S)}{\bar{L}}$$

为码的**剩余度**。

在二元无噪无损信道中 $r = 2$, $\eta = \frac{H(S)}{\bar{L}}$ ，所以在二元无噪无损信道中信息传输率为

$$R = \frac{H(S)}{\bar{L}} = \eta$$

注意它们数值相同，单位不同。 η 是个无单位的比值，而 R 的单位是比特/码符号。

»» 5.3.4 香农第一编码定理

【例5.5】 有一离散无记忆信源

$$\begin{bmatrix} S \\ P \end{bmatrix} = \begin{bmatrix} s_1 & s_2 \\ 3/4 & 1/4 \end{bmatrix}$$

求其信息传输率及二次扩展信源的信息传输率。

»» 5.4 变长码的编码方法

本章介绍的变长码的常见编码方法，如香农编码、香农-费诺-埃利斯编码、霍夫曼编码、费诺编码均为匹配编码，也称**统计编码**，都是通过使用较短的码字来给出现概率较高的信源符号编码。而出现概率较小的信源符号用较长的码字来编码，从而使平均码长最短，达到最佳编码的目的。下面介绍这几种方法：

5.4.1 香农编码

香农码的方法是选择每个码字长度 l_i 满足：
$$l_i = \left\lceil \log \frac{1}{p(s_i)} \right\rceil \quad i = 1, 2, \dots, q$$

按香农编码方法构造的码，其平均码长 \bar{L} 不超过上界，即： $\bar{L} \leq H_r(S) + 1$

»» 5.4 变长码的编码方法

其具体方法如下：

(1) 将信源发出的 q 个消息符号按其概率的递减次序排列

$$p(s_1) \geq p(s_2) \geq \cdots \geq p(s_q)$$

(2) 计算出各个信源符号的累加概率 $F(s_i) = \sum_{k=1}^{i-1} p(s_k)$

(3) 计算第 i 个消息的二元代码组的码长 $l_i = \left\lceil \log \frac{1}{p(s_i)} \right\rceil$

(4) 将累加概率 $F(s_i)$ (十进制小数)变换成二进制小数。根据码长 l_i 取小数点后 l_i 个二进制符号作为第 i 个消息的码字。

5.4.2 香农-费诺-埃利斯编码

将香农编码中的累加概率换成修正累加概率即可得到相应的香农-费诺-埃利斯编码：

(1) 计算出各个信源符号的修正累加概率

$$\bar{F}(s_i) = \sum_{k=1}^{i-1} p(s_k) + \frac{1}{2} p(s_i)$$

(2) 按下式计算第*i*个消息的二元代码组的码长 l_i

$$l_i = \left\lceil \log \frac{1}{p(s_i)} \right\rceil + 1$$

(3) 将累加概率 $\bar{F}(s_i)$ (十进制小数)变换成二进制小数. 根据码长 l_i 取小数点后 l_i 个二进制符号作为第*i*个消息的码字.

»» 5.4.3 霍夫曼编码

这是霍夫曼于1952年提出的一种构造紧致码的方法。具体方法如下：

- (1) 将 q 个信源符号按概率大小递减排列； $p(s_1) \geq p(s_2) \geq \cdots \geq p(s_q)$
- (2) 用“0，1”码符号分别代表概率最小的两个信源符号，并将这两个概率最小的信源符号合并成一个，从而得到只包含 $q-1$ 个符号的新信源称为缩减信源 S_1 ；
- (3) 把缩减信源 S_1 的符号仍按概率大小递减次序排列，再将其最后两个概率最小的信源符号分别用“0”和“1”码符号表示，并且合并成一个符号，这样又形成了 $q-2$ 个信源符号的缩减信源 S_2 ；
- (4) 依次继续下去，直至信源最后只剩下两个信源符号为止，将这最后两个信源符号分别用二元码符号“0”和“1”表示；
- (5) 然后从最后一级缩减信源开始，进行回溯，就得到各信源符号所对应的码符号序列，即相应的码字。

»» 5.4.3 霍夫曼编码

霍夫曼编码方法得到的码并非是唯一的。造成非唯一的原因有两个：

- (1) 每次对信源缩减时，赋予最后两个概率最小的信源符号的码符号“0”和“1”是可以互换的，所以可得到不同的霍夫曼码；
- (2) 对信源进行缩减时，如果两个概率最小的信源符号合并后的概率与其他信源符号的概率相同，则在进行概率排序时的次序可以是任意的，因此会得到不同的霍夫曼码。

霍夫曼码是用概率匹配的方法进行信源编码。它有两个明显特点：

- (1) 霍夫曼码的编码方法保证了概率大的信源符号对应的码长小，概率小的信源符号对应的码长大，充分利用了短码；
- (2) 每次缩减信源的最长两个码字有相同的码长，并且仅仅只有最后一位码符号不同。

定理5.8 霍夫曼码是紧致码。

»» 5.4.4 r 元霍夫曼编码

二进制霍夫曼码的编码方法很容易推广到 r 进制的情况，只是编码过程中构成缩减信源时，每次都是将 r 个概率最小的信源符号合并，并分别用码符号 $0, 1, \dots, (r - 1)$ 表示。

为了充分利用短码，使霍夫曼码的平均码长最短，必须使最后一个缩减信源恰好有 r 个信源符号。因此对于 r 元霍夫曼编码，信源 S 符号个数 q 必须满足 $q = (r - 1)\theta + r$ 。 θ 表示信源缩减次数。

如果不满足上式，则可以在最后增补一些概率为 0 的信源符号，因此上式又可以写成 $q + i = (r - 1)\theta + r$ 。 i 为增加的信源符号数，并且是满足上式的最小正整数或 0 当 $r = 2$ 时的二源码。信源 S 的符号个数 q 满足 $q = \theta + 2$ 。

由于霍夫曼码是紧致码，所以编码后单个信源符号平均码长随 N 的增加很快接近于极限值——信源熵。

»» 5.4.5 费诺编码

费诺码编码过程如下：

(1) 将信源符号 $s_i, i = 1, 2, \dots, q$ 以概率递减次序排列，即

$$p(s_1) \geq p(s_2) \geq \dots \geq p(s_q)$$

(2) 将依次排列的信源符号以概率分为两组，使两个组的概率和基本相等，并对各组赋予二进制码符号“0”和“1”；

(3) 将每一大组的信源符号进一步再分成两组，使划分后的两个组的概率和近于相等，又分别赋予各组二进制码符号“0”和“1”；

(4) 如此重复，直至每组只剩下一个信源符号为止；

(5) 信源符号所对应的码符号序列即为费诺码。

»» 5.4.5 费诺编码

以上讨论了霍夫曼码和其他一些编码方法，并且证明了霍夫曼码是最佳码，当 N 不很大时，它能使无失真编码的效率接近于1，但是在实际使用时设备较复杂。

首先，每个信源符号所对应的码长不同，一般情况下，

信源符号以匀速输出，信道也是匀速传输的。

其次，差错扩散的问题。变长码的一个码元的差错可能造成译码错误，并且还会造成同步错误，结果后面一系列码字也会译错。

最后，霍夫曼码的编译码需要用查表的方法来进行。

编码	平均码长	信息传输率 R
香农码	3.14	0.831
霍夫曼	2.72	0.959
费诺码	2.74	0.953

表5.15 三种编码方法的比较

»»» 5.5 实用的无失真信源编码方法

5.5.1 游程编码

游程编码主要用于黑白二值文件、传真的数据压缩。由于传真文件中连“0”和连“1”较多，这些连“0”或连“1”的字符串称为游程。对游程长度进行霍夫曼编码或其他的编码处理就可以达到压缩数据的目的。图5.5是一幅10×50黑白二值图像。

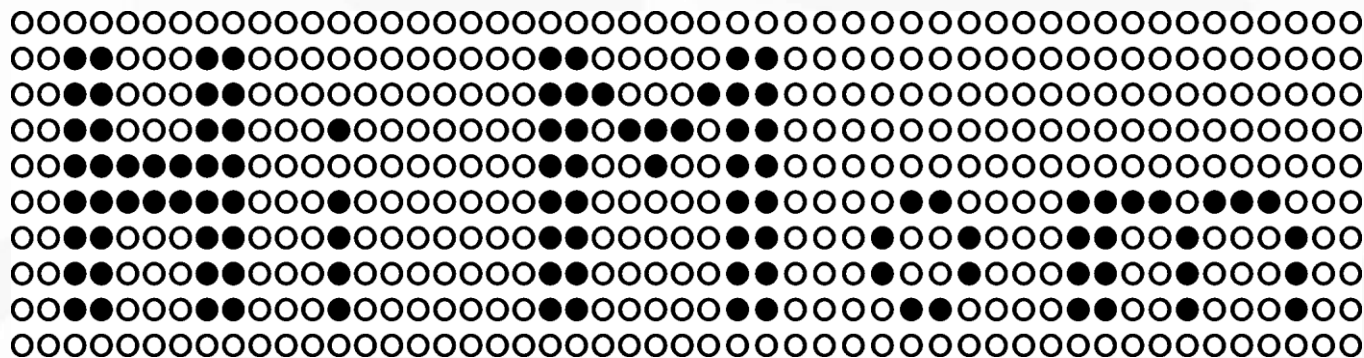


图5.5 二值图像

»» 5.5 实用的无失真信源编码方法

MH编码是一种实用的游程编码算法，应用于黑、白传真数据的压缩编码，根据不同的黑、白游程长度有两张结尾码表和两张组合码表其基本的编码规范为

- (1) 游程长度在0 ~ 63时，直接查表用相应的结尾码作为码字；
- (2) 游程长度在64 ~ 1728范围内时，用组合码加上结尾码作为相应的码字；
- (3) 每行的第一个游程规定为白游程(长度可以为零)，每行用一个结束码(EOL) 终止；
- (4) 在传输时，每页数据之前加一个结束码，每页尾部连续实用6个结束码。

»» 5.5.2 算术编码

算术式编码是一种非分组码，无需计算出所有 N 长信源序列的概率分布及码表，可以直接对输入的信源符号序列编码输出。这种方法是由香农-费诺-埃利斯编码直接扩展得到的。

算术式编码算法的中心思想是高效地计算 n 长信源符号序列 s 的分布概率 $p(s)$ 和累积概率 $F(s)$ ，然后用区间 $[F(s), F(s) + p(s)]$ 中的一个值来作为 s 的码字。

已知 N 长信源符号序列 s 的概率 $p(s)$ 和累加概率 $F(s)$ ， $N+1$ 长信源符号序列 sr 的概

率 $p(s)$ 和累加概率 $F(sr)$ 为：

$$\begin{cases} F(sr) = F(s) + p(s)F(r) \\ p(sr) = p(s)p(r) \end{cases}$$

r 为新添加的符号，信源符号序列 s 的码长： $l = \lceil -\log p(s) \rceil$