# 第7章 信道容量

当说到“A与B通信”时，我们的真实意思是什么？我们的意思是A的物理行为使B产生一种需要的物理状态。信息的传输是一个物理过程，因此，必然受到无法控制的周边噪声以及信号处理本身缺陷的影响。如果接收者B与传输者A就所传输的内容是一致的，那么说这次通信是成功的。

在本章中，在 $n$ 次使用信道下，将计算出可区分的信号的最大数目。该数与 $n$ 成指数增长关系，这个指数就是所说的信道容量。信道容量(可区别的信号数目的对数值)被特征化为最大互信息，是信息论的中心问题，也是信息论中最著名的成就。

在图7-1中给出一个物理发送信号系统的数学模拟。来自某个有限字母表的信源字符被映射成一系列信道字符串，系统就得到信道的输出序列。输出序列虽然是随机的，但它的分布由输入序列决定。我们试图凭借着这些输出序列来恢复出传输的消息。

每个可能的输入序列将导出关于输出序列的概率分布。由于两个不同的输入序列可以产生相同的输出序列，于是根据输出序列不知道输入序列到底是哪个。在下面的几节中，我们将证明能够以很高的概率从输入序列中挑选出一个“不会混淆”的子集，使得对于每一个特定的输出序列，只存在惟一的一个输入最有可能导致该输出。于是，在不计较可以忽略的误差概率的情况下，可以在输出端重构输入序列。将信源映射到适合于输入信道的“足够分散的”输入序列集合，我们能够以非常低的误差概率传输一条消息，并且在信道的输出端重构出这个信源消息。可实现的最大的码率称作该信道的容量。

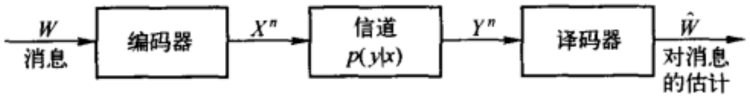


图7-1 通信系统

183 **定义** 离散信道(discret channel)是由输入字母表$\mathcal{X}$，输出字母表$\mathcal{Y}$和概率转移矩阵 $p(y|x)$ 构成的系统，其中 $p(y|x)$表示发送字符 $x$ 的条件下收到输出字符$y$ 的概率。如果输出的概率分布仅依赖于它所对应的输入，而与先前信道的输入或者输出条件独立，就称这个信道是无记忆的(memoryless)。

**定义** 离散无记忆信道的“信息”信道容量(channel capacity)定义为

$$C=\max_{p(x)} I(X;Y) \tag{7-1}$$

这里的最大值取自所有可能的输入分布 $p(x)$。

我们稍后将给出信道容量的一个可操作性的定义，也就是将信道容量定义为信道的最高码率(单位为比特/信道使用)，在此码率下，信息能够以任意小的误差概率被传输。香农第二定理表明，信息信道容量等于这个可操作的信道容量。于是，在大多数情况下，讨论信道容量时总是略去信息(information)这个字眼。

在数据压缩与数据传输问题之间存在对偶性。在压缩过程中，去除数据中所有的冗余以使其得到最大程度的压缩；而在数据传输过程中，以一种受控方式加入冗余以抵抗信道传输中可能

发生的错误。在 7.13 节中，我们将证明一般的通信系统可以分成两部分，而且数据压缩与数据传输问题可以分开考虑。

## 7.1 信道容量的几个例子

### 7.1.1 无噪声二元信道

假定有如图 7-2 所示的信道，它的二元输入在输出端能精确地重现。

在这种情况下，任何一个传输的比特都能被无误差地接收到。因此，每次使用该信道，都可以毫无误差地传输一个比特，信道容量就是 1 比特。当然，也可以计算得到信息容量 $C=\max I(X;Y)=1$ 比特，且在 $p(x)=\left(\frac{1}{2},\frac{1}{2}\right)$时达到。 184

### 7.1.2 无重叠输出的有噪声信道

这个信道对于两个输入中的每一个，均有两个可能的输出，如图 7-3 所示。这个信道看起来有噪声，其实不然。即使信道的输出是输入的随机结果，但输入也可以根据输出确定，于是每个传输的比特都可以准确无误地得到恢复。因此，该信道的容量仍然是 1 比特/传输。也可以计算出该信道的信息容量 $C=\max I(X;Y)=1$ 比特，且在 $p(x)=\left(\frac{1}{2},\frac{1}{2}\right)$时达到。 185

### 7.1.3 有噪声的打字机信道

在此情形中，信道输入以概率 1/2 在输出端无改变地被接收，或以概率 1/2 转变为下一个字母（如图 7-4 所示）。若输入端有 26 个字符，并以间隔的方式使用输入字符，那么在每次传输过程中，可以毫无误差地传输其中的 13 个字符。因此，该信道的容量为 log13 比特/传输。也可计算得到信道的容量 $C=\max I(X;Y)=\max[H(Y)-H(Y|X)]=\max H(Y)-1=\log 26-1=\log 13$ 比特，且当 $p(x)$为整个输入字母表上的均匀分布时达到该容量。


图 7-2 无噪声二元信道。$C=1$ 比特

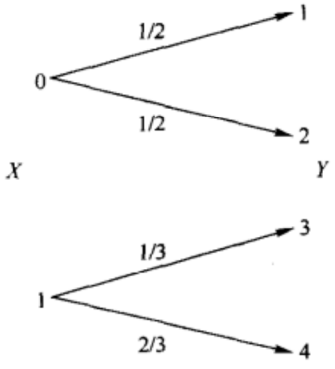


图 7-3 无重叠输出的有噪声信道。$C=1$ 比特

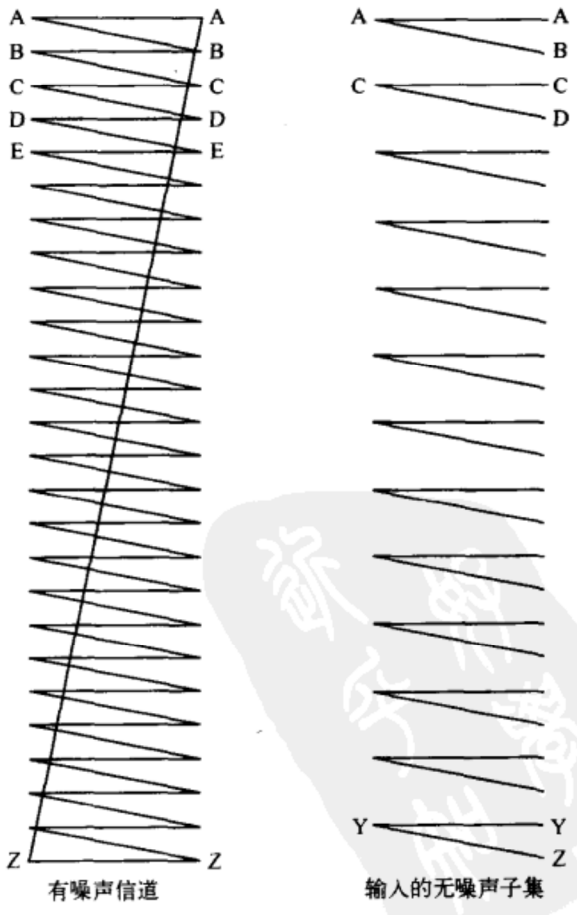


图 7-4 噪声打字机信道。$C=\log 13$ 比特 186

### 7.1.4 二元对称信道

考虑如图 7-5 所示的二元对称信道(Binary Symmetric Channel, BSC)。这个二元信道的输入字符以概率 $p$ 互补。这是一个有误差信道的最简单模型，然而，它反映出了有误差信道问题的复杂度的普遍特点。

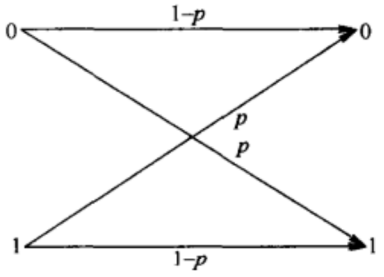

图 7-5 二元对称信道。$C=1-H(p)$比特

在出现错误时，0 作为 1 收到，或者正好相反。从接收到的比特中我们并不能看出哪里发生了错误。从某种意义上说，所有接收到的比特都不可靠。稍后将证明，我们仍然可以使用这样的通信信道以非 0 的传输码率发送信息，并且误差概率任意小。

给出互信息的一个界

$$I(X;Y)=H(Y)-H(Y|X) \tag{7-2}$$
$$=H(Y)-\sum p(x)H(Y|X=x) \tag{7-3}$$
$$=H(Y)-\sum p(x)H(p) \tag{7-4}$$
$$=H(Y)-H(p) \tag{7-5}$$
$$\leqslant 1-H(p) \tag{7-6}$$

其中最后一个不等式成立是因为 $Y$ 是一个二元随机变量。当输入分布是均匀分布时等号成立。因此，参数为 $p$ 的二元对称信道的信息容量是

187

$$C=1-H(p) \quad \text{比特} \tag{7-7}$$

### 7.1.5 二元擦除信道

有一种信道类似于二元对称信道，会损失一些比特(不是被损坏)，这种信道称作二元擦除信道(binary erasure channel)。在二元擦除信道中，比例为 $\alpha$ 的比特被擦除掉，并且接收者知道是哪些比特已经被擦除掉了。如图 7-6 所示，二元擦除信道有两个输入和三个输出。

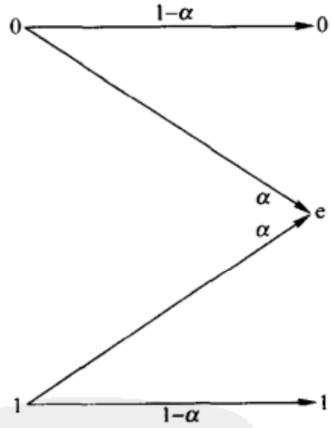

图 7-6 二元擦除信道

计算二元擦除信道的容量如下：

$$C=\max_{p(x)} I(X;Y) \tag{7-8}$$
$$=\max_{p(x)}(H(Y)-H(Y|X)) \tag{7-9}$$
$$=\max_{p(x)} H(Y)-H(\alpha) \tag{7-10}$$

初看，似乎 $H(Y)$的最大值是 log3，但无论选择什么输入分布 $p(x)$，都无法达到这个值。设 $E$ 代表事件$\{Y=e\}$，并使用表达式

$$H(Y)=H(Y,E)=H(E)+H(Y|E) \tag{7-11}$$

设 $\Pr(X=1)=\pi$，我们有

188

$$H(Y)=H((1-\pi)(1-\alpha),\alpha,\pi(1-\alpha))=H(\alpha)+(1-\alpha)H(\pi) \tag{7-12}$$

因此

$$C=\max_{p(x)} H(Y)-H(\alpha) \tag{7-13}$$
$$=\max_{\pi}(1-\alpha)H(\pi)+H(\alpha)-H(\alpha) \tag{7-14}$$
$$=\max_{\pi}(1-\alpha)H(\pi) \tag{7-15}$$
$$=1-\alpha \tag{7-16}$$

其中，当 $\pi=1/2$ 时，达到该信道容量。

这个信道容量的表达式有其直观的意义：由于比例为 $\alpha$ 的比特在信道中损失，因而我们(至多)能够恢复比例为 $1-\alpha$ 的比特。因此，容量至多为 $1-\alpha$。但码率是否真的可以达到这个值并不十分明显，这可以从香农第二定理推出。

对于许多实际的信道，发送者会从接收者那里收到一些反馈。如果二元擦除信道中存在反馈，那么很清楚下一步该做什么：如果一个比特损失了，那么重新传输它，直到其顺利通过为止。由于所有比特以概率 $1-\alpha$ 通过，所以传输的有效码率就是 $1-\alpha$。在这种方式下，通过反馈可以容易地达到容量 $1-\alpha$。

在本章后面的部分中，将证明，无论有无反馈，$1-\alpha$ 都是信道可以达到的最高码率。这个事实令人惊讶，也就是说反馈并不能增加离散无记忆信道的容量。

## 7.2 对称信道

二元对称信道的容量是 $C=1-H(p)$比特/传输，二元擦除信道的容量是 $C=1-\alpha$ 比特/传输。下面考虑具有如下转移矩阵的信道：

$$p(y|x)=\begin{bmatrix}0.3 & 0.2 & 0.5\\ 0.5 & 0.3 & 0.2\\ 0.2 & 0.5 & 0.3\end{bmatrix} \tag{7-17}$$

上述矩阵中的第 $x$ 行第 $y$ 列的元素表示条件概率 $p(y|x)$，即传输 $x$ 收到 $y$ 的概率。在该信道中，概率转移矩阵中所有的行都可以通过其他行置换得到，每一列也如此。这样的信道称为对称的(symmetric)。另一个对称信道的例子如

$$Y=X+Z \pmod c \tag{7-18}$$

189

其中 $Z$ 服从整数集$\{0,1,2,\cdots,c-1\}$上的某个分布，$X$ 与 $Z$ 拥有相同的字母表，并且 $Z$ 独立于 $X$。

在上述两种情况中，我们能够容易地求得信道容量的显表达式。设 $\mathbf{r}$ 表示转移矩阵的一行，则有

$$I(X;Y)=H(Y)-H(Y|X) \tag{7-19}$$

$$=H(Y)-H(\mathbf{r}) \tag{7-20}$$

$$\leqslant \log|\mathcal{Y}|-H(\mathbf{r}) \tag{7-21}$$

当输出是均匀分布时等号成立。而且，$p(x)=1/|\mathcal{X}|$可以使 $Y$ 达到均匀分布，这可由如下式子看出

$$p(y)=\sum_{x\in\mathcal{X}}p(y|x)p(x)=\frac{1}{|\mathcal{X}|}\sum p(y|x)=c\frac{1}{|\mathcal{X}|}=\frac{1}{|\mathcal{Y}|} \tag{7-22}$$

其中 $c$ 是概率转移矩阵的一列中所有元素之和。

于是，式(7-17)中的信道容量为

$$C=\max_{p(x)}I(X;Y)=\log 3-H(0.5,0.3,0.2) \tag{7-23}$$

并且当输入分布为均匀时达到上述容量 $C$。

如上定义的对称信道的转移矩阵是双随机的。在计算信道容量时，我们用到了转移矩阵中行与行互为置换以及各列元素之和都相等的性质。

基于这些性质，可以对对称信道的概念进行如下的推广：

**定义** 如果信道转移矩阵 $p(y|x)$的任何两行互相置换；任何两列也互相置换，那么称该信道是对称的(symmetric)。如果转移矩阵的每一行 $p(\cdot|x)$都是其他每行的置换，而所有列的元素和$\sum_x p(y|x)$相等，则称这个信道是弱对称的(weakly symmetric)。

例如，转移矩阵为

$$p(y|x)=\begin{pmatrix}\frac{1}{3} & \frac{1}{6} & \frac{1}{2}\\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6}\end{pmatrix} \tag{7-24}$$

190 的信道是弱对称的，但不对称。

上面关于对称信道的一些结论同样适用于弱对称信道。除此之外，对于弱对称信道，我们还有下列定理：

**定理** 7.2.1 对于弱对称信道，

$$C=\log|\mathcal{Y}|-H(\text{转移矩阵的行}) \tag{7-25}$$

当输入字母表上的分布为均匀时达到该容量。

## 7.3 信道容量的性质

1. 由于 $I(X;Y)\geqslant 0$，所以 $C\geqslant 0$。
2. 由于 $C=\max I(X;Y)\leqslant \max H(X)\leqslant \log|\mathcal{X}|$，所以 $C\leqslant \log|\mathcal{X}|$。
3. $C\leqslant \log|\mathcal{Y}|$，理由同上。
4. $I(X;Y)$是关于 $p(x)$的一个连续函数。
5. $I(X;Y)$是关于 $p(x)$的凹函数(定理 2.7.4)。由于 $I(X;Y)$是闭凸集上的凹函数，因而局部最大值也是全局最大值。由上述性质 2 和 3 可以看出，最大值是有限的，这证实了在容量的定义中使用 max 而不用 sup 记号是合理的。最大值可以利用标准的非线性最优化技术(如梯度搜索)求解。下面这些方法都可以考虑：

- 利用微积分和库恩－塔克条件求解带约束的最大化问题。
- Frank-Wolfe 梯度搜索算法。
- 由 Arimoto[25]和 Blahut[65]开发的迭代算法。在 10.8 节中详细叙述该算法。

一般得不到信道容量的解析解(closed-form solution)，但对于很多简单的信道，可以利用它们的特性(如对称性)来计算出信道容量。前面例子中提到过的那些信道就具有解析解。

## 7.4 信道编码定理预览

191 到现在为止，我们已经给出了离散无记忆信道的信息容量定义。在下一节中，我们将证明香农第二定理，它给出了容量定义的可操作性解释，即容量可以视为能够在该信道中可靠传输的比特数。但首先将尝试给出一个直观思路，解释为什么能通过信道来传输 $C$ 比特的信息。基本思路是，对于大的分组长度，每个信道可以看作是有噪声打字机信道(图 7-4)，由此每个信道都有一个输入子集，使得在输出端接收到的序列基本上互不相交。

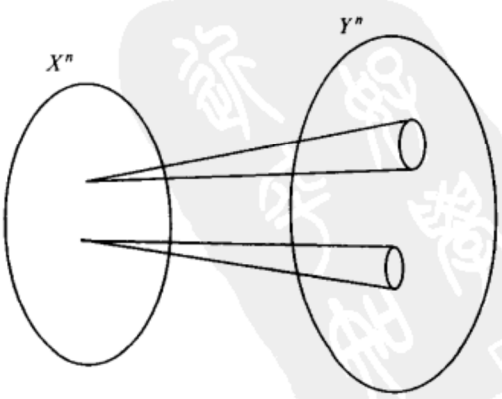

图 7-7 $n$ 次使用下的信道

对于输入的每个(典型的)$n$ 长序列，会有大约 $2^{nH(Y|X)}$个可能的 $Y$ 序列与之对应，并且所有这些序列是等可能的(如图 7-7)。我们希望确保没有两个 $X$ 序列能够产生相同的 $Y$ 输出序列。否则，将无法判断到底传输的是哪个 $X$ 序列。

所有可能的(典型的)$Y$ 序列的总数约等于 $2^{nH(Y)}$。对应于不同的输入 $X$ 序列，这个集合分割成大小为

$2^{nH(Y|X)}$的许多个小集合。所以不相交集的总数小于等于 $2^{n(H(Y)-H(Y|X))}=2^{nI(X;Y)}$。因此，我们至多可以传输$\approx 2^{nI(X;Y)}$个可区分的 $n$ 长序列。

虽然以上讨论只是大致描述了容量的上界，在下一节中，将用更加严格的语言来证明码率 $I$ 是可达到的，而且误差概率可以任意低。

在开始香农第二定理的证明之前，我们需要一些定义。

## 7.5 定义

我们分析如图 7-8 所示的通信系统。

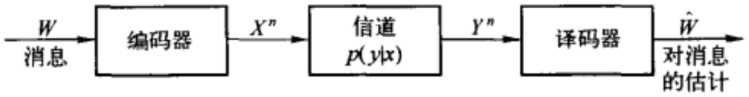


图 7-8 通信信道

取自下标集$\{1,2,\cdots,M\}$的消息 $W$，产生信号 $X^n(W)$，这个信号以随机序列 $Y^n\sim p(y^n|x^n)$ 192 的方式被接收者收到。然后，接收者使用适当的译码规则 $\hat{W}=g(Y^n)$猜测消息 $W$。如果 $\hat{W}$与所传输的消息 $W$ 不同，则表明接受者出错。下面我们严格定义这些思路。

**定义** 用$(\mathcal{X},p(y|x),\mathcal{Y})$表示的离散信道由两个有限集$\mathcal{X}$和 $\mathcal{Y}$以及一簇概率密度函数 $p(y|x)$ $(x\in\mathcal{X})$构成，其中对任意 $x$ 与$y$，有 $p(y|x)\geqslant 0$，以及对任意的 $x$，有 $\sum_y p(y|x)=1$，而 $X$ 和 $Y$ 分别看作信道的输入与输出。

**定义** 离散无记忆信道(DMC)的 $n$ 次扩展是指信道$(\mathcal{X}^n,p(y^n|x^n),\mathcal{Y}^n)$，其中

$$p(y_k|x^k,y^{k-1})=p(y_k|x_k),k=1,2,\cdots,n \tag{7-26}$$

**注释** 如果信道不带反馈，也就是说，如果输入字符不依赖于过去的输出字符，即 $p(x_k|x^{k-1},y^{k-1})=p(x_k|x^{k-1})$，那么离散无记忆信道的 $n$ 次扩展的信道转移函数就简化为

$$p(y^n|x^n)=\prod_{i=1}^{n}p(y_i|x_i) \tag{7-27}$$

在讨论离散无记忆信道时，除非明确指出，一般都是指不带反馈的离散无记忆信道。

**定义** 信道$(\mathcal{X},p(y|x),\mathcal{Y})$的$(M,n)$码由以下部分构成：

1. 下标集$\{1,2,\cdots,M\}$。

2. 编码函数 $X^n:\{1,2,\cdots,M\}\to\mathcal{X}^n$，生成码字 $x^n(1),x^n(2),\cdots,x^n(M)$。所有码字的集合称作码簿(codebook)。 193

3. 译码函数

$$g:\mathcal{Y}^n\to\{1,2,\cdots,M\} \tag{7-28}$$

它是一个确定性规则，为每个收到的字符向量指定一个猜测。

**定义(条件误差概率)** 设

$$\lambda_i=\Pr(g(Y^n)\neq i\mid X^n=x^n(i))=\sum_{y^n}p(y^n\mid x^n(i))I(g(y^n)\neq i) \tag{7-29}$$

为已知下标 $i$ 被发送的条件下的条件误差概率(conditional probability of error)，其中 $I(\cdot)$为示性函数。

**定义** $(M,n)$码的最大误差概率 $\lambda^{(n)}$(maximum probability of error)定义为

$$\lambda^{(n)}=\max_{i\in\{1,2,\cdots,M\}}\lambda_i \tag{7-30}$$

**定义** $(M,n)$码的(算术)平均误差概率 $P_e^{(n)}$(average probability of error)定义为

$$P_e^{(n)} = \frac{1}{M}\sum_{i=1}^{M}\lambda_i \tag{7-31}$$

注意，如果下标 $W$ 是从集合$\{1,2,\cdots,M\}$中的均匀分布中选出的，以及 $X^n = x^n(W)$，则

$$P_e^{(n)} \triangleq \Pr(W \neq g(Y^n)) \tag{7-32}$$

(即 $P_e^{(n)}$ 为误差概率。)显然，有

$$P_e^{(n)} \leqslant \lambda^{(n)} \tag{7-33}$$

人们一般期望，最大误差概率与平均误差概率的性质有相当大的差异。然而，在下一节中我们将
194 证明，在相同的码率下，平均误差概率很小可以推出它的最大误差概率也很小。

值得注意的是，式(7-32)中定义的 $P_e^{(n)}$ 仅是条件误差概率 $\lambda_i$ 的一种数学构造，它本身成为误差概率只有当消息均匀取自消息集$\{1,2,\cdots,2^M\}$时才成立。然而，不论是在可达性的证明中，还是其逆命题中，都选取 $W$ 上的均匀分布来界定误差概率。这使我们能够确定 $P_e^{(n)}$ 以及最大误差概率$\lambda^n$ 的行为，从而，不论信道是如何使用的，也能刻画出信道的行为(即不考虑 $W$ 的分布是什么)。

**定义** $(M,n)$码的**码率** $R$(rate)为

$$R = \frac{\log M}{n} \quad \text{比特/传输} \tag{7-34}$$

**定义** 如果存在一个$(\lceil 2^{nR}\rceil, n)$码序列，满足当 $n\to 0$ 时，最大误差概率 $\lambda^{(n)}\to 0$，则称码率 $R$ 是可达的(achievable)。

为简化记号，以下我们将用$(2^{nR},n)$码来表示$(\lceil 2^{nR}\rceil,n)$码。

**定义** 信道的容量定义为所有可达码率的上确界。

于是，对于充分大的分组长度，小于信道容量的码率对应的误差概率可以任意小。

## 7.6 联合典型序列

粗略地说，如果码字 $X^n(i)$与接收到的信号 $Y^n$ 是“联合典型”的话，就将信道输出 $Y^n$ 译为第 $i$ 个下标。现在来定义联合典型这一重要的概念，并且计算当 $Y^n$ 确实由 $X^n(i)$产生与不是由 $X^n(i)$产生时，这两种情况所对应的联合典型概率。

**定义** 服从分布 $p(x,y)$的联合典型序列$\{(x^n,y^n)\}$所构成的集合 $A_\varepsilon^{(n)}$ 是指其经验熵与真实熵 $\varepsilon$ 接近的 $n$ 长序列构成的集合，即：

$$A_\varepsilon^{(n)} = \{(x^n,y^n)\in \mathcal{X}^n\times\mathcal{Y}^n:$$

$$\left|-\frac{1}{n}\log p(x^n) - H(X)\right| < \varepsilon \tag{7-35}$$

195

$$\left|-\frac{1}{n}\log p(y^n) - H(Y)\right| < \varepsilon \tag{7-36}$$

$$\left|-\frac{1}{n}\log p(x^n,y^n) - H(X,Y)\right| < \varepsilon\Big\} \tag{7-37}$$

其中

$$p(x^n,y^n) = \prod_{i=1}^{n} p(x_i,y_i) \tag{7-38}$$

**定理 7.6.1(联合 AEP)** 设 $(X^n,Y^n)$ 为服从 $p(x^n,y^n) = \prod_{i=1}^{n} p(x_i,y_i)$ 的 i.i.d 的 $n$ 长序列，那么：

1. 当 $n\to\infty$ 时，$\Pr((X^n,Y^n)\in A_\varepsilon^{(n)})\to 1$。

2. $|A_\varepsilon^{(n)}| \leqslant 2^{n(H(X,Y)+\varepsilon)}$。

3. 如果$(\tilde{X}^n, \tilde{Y}^n) \sim p(x^n)p(y^n)$，即 $\tilde{X}^n$ 与 $\tilde{Y}^n$ 是独立的且与 $p(x^n, y^n)$有相同的边际分布，那么

$$\Pr((\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)}) \leqslant 2^{-n(I(X;Y)-3\varepsilon)} \tag{7-39}$$

而且，对于充分大的 $n$，

$$\Pr((\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)}) \geqslant (1-\varepsilon)2^{-n(I(X;Y)+3\varepsilon)} \tag{7-40}$$

**证明：**

1. 首先证明，包含在典型集中的序列具有很高的概率。由弱大数定律，

$$-\frac{1}{n}\log p(X^n) \to -E[\log p(X)] = H(X) \quad \text{依概率} \tag{7-41}$$

因此，给定 $\varepsilon > 0$，存在 $n_1$，使得对于任意 $n > n_1$，

$$\Pr\left(\left| -\frac{1}{n}\log p(X^n) - H(X) \right| \geqslant \varepsilon\right) < \frac{\varepsilon}{3} \tag{7-42}$$

类似地，由弱大数定律，

$$-\frac{1}{n}\log p(Y^n) \to -E[\log p(Y)] = H(Y) \quad \text{依概率} \tag{7-43}$$ 196

以及

$$-\frac{1}{n}\log p(X^n, Y^n) \to -E[\log p(X, Y)] = H(X, Y) \quad \text{依概率} \tag{7-44}$$

从而，存在 $n_2$ 和 $n_3$，使得对于任意 $n \geqslant n_2$，

$$\Pr\left(\left| -\frac{1}{n}\log p(Y^n) - H(Y) \right| \geqslant \varepsilon\right) < \frac{\varepsilon}{3} \tag{7-45}$$

以及对任意的 $n \geqslant n_3$，

$$\Pr\left(\left| -\frac{1}{n}\log p(X^n, Y^n) - H(X, Y) \right| \geqslant \varepsilon\right) < \frac{\varepsilon}{3} \tag{7-46}$$

选取 $n > \max(n_1, n_2, n_3)$，则式(7-42)、(7-45)和式(7-46)中的集合之并的概率必定小于 $\varepsilon$。因此，对于充分大的 $n$，集合 $A_\varepsilon^{(n)}$ 的概率大于 $1-\varepsilon$，从而证明了定理的第一部分。

2. 为证明定理的第二部分，我们注意到

$$1 = \sum p(x^n, y^n) \tag{7-47}$$

$$\geqslant \sum_{A_\varepsilon^{(n)}} p(x^n, y^n) \tag{7-48}$$

$$\geqslant |A_\varepsilon^{(n)}| \, 2^{-n(H(X,Y)+\varepsilon)} \tag{7-49}$$

因此

$$|A_\varepsilon^{(n)}| \leqslant 2^{n(H(X,Y)+\varepsilon)} \tag{7-50}$$

3. 现在，如果 $\tilde{X}^n$ 与 $\tilde{Y}^n$ 相互独立，但是与 $X^n$ 和 $Y^n$ 分别具有相同的边际分布，那么

$$\Pr((\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)}) = \sum_{(x^n, y^n) \in A_\varepsilon^{(n)}} p(x^n)p(y^n) \tag{7-51}$$

$$\leqslant 2^{n(H(X,Y)+\varepsilon)} 2^{-n(H(X)-\varepsilon)} 2^{-n(H(Y)-\varepsilon)} \tag{7-52}$$

$$= 2^{-n(I(X;Y)-3\varepsilon)} \tag{7-53}$$ 197

对充分大的 $n$，$\Pr(A_\varepsilon^{(n)}) \geqslant 1-\varepsilon$，因此

$$1 - \varepsilon \leqslant \sum_{(x^n, y^n) \in A_\varepsilon^{(n)}} p(x^n, y^n) \tag{7-54}$$

$$\leqslant |A_\varepsilon^{(n)}| 2^{-n(H(X,Y)-\varepsilon)} \tag{7-55}$$

以及

$$|A_\varepsilon^{(n)}| \geqslant (1-\varepsilon)2^{n(H(X,Y)-\varepsilon)} \tag{7-56}$$

类似上界估计的讨论，也可以证明，对充分大的 $n$，

$$\Pr((\tilde{X}^n, \tilde{Y}^n) \in A_\varepsilon^{(n)}) = \sum_{A_\varepsilon^{(n)}} p(x^n)p(y^n) \tag{7-57}$$

$$\geqslant (1-\varepsilon)2^{n(H(X,Y)-\varepsilon)}2^{-n(H(X)+\varepsilon)}2^{-n(H(Y)+\varepsilon)} \tag{7-58}$$

$$= (1-\varepsilon)2^{-n(I(X;Y)+3\varepsilon)} \tag{7-59}$$□

图 7-9 是关于联合典型集的示意图。大约有 $2^{nH(X)}$ 个典型的 $X$ 序列和大约 $2^{nH(Y)}$ 个典型的 $Y$ 序列。但是，联合典型序列只有 $2^{nH(X,Y)}$ 个，所以并不是所有典型的 $X^n$ 与典型的 $Y^n$ 构成的序列对都是联合典型的。随机选取的序列对是联合典型的概率大约为 $2^{-nI(X;Y)}$。因此，我们很可能需要考虑约 $2^{nI(X;Y)}$ 个这样的序列对，才可能遇到一个联合典型对。这表明存在大约 $2^{nI(X;Y)}$ 个可区分的信号 $X^n$。

198

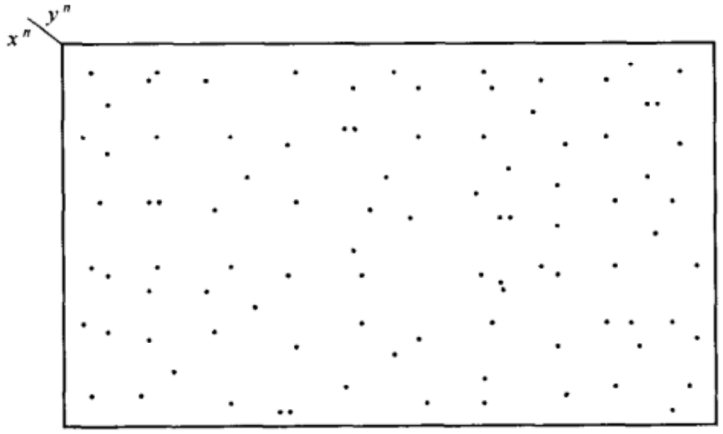


图 7-9 联合典型序列

着眼上述问题的另一种方式是考虑固定输出序列 $Y^n$ 下的联合典型序列集，这里假定该输出序列来自真实的输入信号 $X^n$。对于序列 $Y^n$，大约存在 $2^{nH(X|Y)}$ 个条件典型的输入信号。某个随机选取的(其他)输入信号 $X^n$ 与 $Y^n$ 为联合典型的概率大约等于 $2^{nH(X|Y)}/2^{nH(X)} = 2^{-nI(X;Y)}$。这再次表明，我们可能选取出大约 $2^{nI(X;Y)}$ 个码字 $X^n(W)$，才能使其中的一个码字与产生输出 $Y^n$ 的对应码字混淆起来。

## 7.7 信道编码定理

我们现在证明信道容量的可达性，这也许是信息论中最基本的定理。最初的证明由香农在1948年的开创性论文中给出。该结果与直观感觉正好相反。如果在信道传输过程中存在误差，那么如何纠正所有误差？任何纠错过程本身也要受到误差的影响，这样将无穷无尽地进行下去。

为了证明只要码率小于信道容量，信息就可以通过该信道可靠地传输，香农使用了许多新的思想。这些思想包括：

- 允许任意小的非 0 误差概率存在，
- 连续使用信道许多次，以保证可以使用大数定律，
- 在随机选择的码簿上计算平均误差概率，这样可以使概率对称，而且可以用来证明至少存在一个好的编码。

香农的概述性证明基于典型序列的思想，其严格的证明直到很晚才给出。下面将要给出的证明利用了典型序列的性质，而且可能也是至今为止给出的最简单的证明。在所有的证明中，都使用了相同的基本思想——随机码选择，计算随机选择的码字的平均误差概率，等等。主要的差别在于译码规则。在这个证明中，我们使用联合典型性译码规则，即寻找一个与收到的序列是联合典型的码字。如果找到惟一满足该性质的码字，我们则认为这就是被发送的码字。依据前面 199
所述的联合典型性的性质，由于发送的码字与接收到的序列是概率相关的，所以它们以很高的概率成为联合典型。并且，任意其他码字与接收到的序列是联合典型的概率是 $2^{-nI}$。因此，如果码字个数小于 $2^{nI}$，那么可以以很高的概率断定不会有其他的码字能够与被传输的码字相混淆，并且误差概率很小。

虽然联合典型译码仅是次优的，但它便于分析而且可以达到小于信道容量的任何码率。

下面就给出香农第二定理的完整叙述及其证明：

**定理** 7.7.1（信道编码定理） 对于离散无记忆信道，小于信道容量 $C$ 的所有码率都是可达的。具体来说，对任意码率 $R<C$，存在一个$(2^{nR},n)$码序列，它的最大误差概率为 $\lambda^{(n)}\to 0$。

反之，任何满足 $\lambda^{(n)}\to 0$ 的$(2^{nR},n)$码序列必定有 $R\leqslant C$。

**证明**：证明小于 $C$ 的码率$R$ 是可达的，而将逆定理的证明放在 7.9 节。

可达性：固定 $p(x)$，根据分布 $p(x)$随机生成一个$(2^{nR},n)$码。具体来说，根据分布

$$p(x^n)=\prod_{i=1}^{n}p(x_i) \tag{7-60}$$

独立生成 $2^{nR}$个码字。将 $2^{nR}$个码字展开为矩阵的行：

$$C=\begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_n(1) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(2^{nR}) & x_2(2^{nR}) & \cdots & x_n(2^{nR}) \end{bmatrix} \tag{7-61}$$

该矩阵中的每一项都是依据 i.i.d 服从 $p(x)$而生成的。因此，我们生成一个特定码$C$的概率就是

$$\Pr(C)=\prod_{w=1}^{2^{nR}}\prod_{i=1}^{n}p(x_i(w)) \tag{7-62}$$ 200

考虑下面的系列事件：

1．如式(7-62)中所述，服从分布 $p(x)$的随机码$C$生成。

2．然后将码$C$告知给发送者和接收者，并且假定二者都知道该信道的信道转移矩阵 $p(y|x)$。

3．依如下的均匀分布选取一条消息 $W$

$$\Pr(W=w)=2^{-nR},\ w=1,2,\cdots,2^{nR} \tag{7-63}$$

4．第 $w$ 个码字 $X^n(w)$是$C$的第 $w$ 行，通过该信道被发送。

5．接收者收到的序列 $Y^n$ 服从分布

$$P(y^n\mid x^n(w))=\prod_{i=1}^{n}p(y_i\mid x_i(w)) \tag{7-64}$$

6．接收者猜测所发送的消息是什么。(使误差概率达到最小的最优方法是最大似然译码，也就是说，接收者应该选择后验($a$ posteriori)概率最大的消息。但是这个过程很难分析。取而代之，使用下面描述的联合典型译码(jointly typical decoding)。这种方法易于分析而且是渐近最优的。)如果满足下面的两个条件，则接收者认为$\hat{W}$就是所发送的下标。

- $(X(\hat{W}),Y^n)$是联合典型的。

• 不存在其他的下标 $W' \neq \hat{W}$ 满足 $(X^n(W'), Y^n) \in A_\varepsilon^{(n)}$。

如果这样的 $\hat{W}$ 不存在，或者有超过一个这样的 $\hat{W}$，则断言发生了错误(在这种情况下，假定接收者给出一个哑下标，例如 0)。

7. 如果 $\hat{W} \neq W$，则说明译码错误，设 $\mathcal{E}$ 代表事件 $\{\hat{W} \neq W\}$。

*误差概率分析*

*概述*：我们首先简要分析一下。我们计算所有随机生成的码(服从式(7-62)的分布)的平均误
201 差概率，而不是某一个码的误差概率。根据编码构造的对称性，平均误差概率不依赖于被发送的具体下标。对一个典型码字，在使用联合典型译码时，存在两种不同的误差源：输出 $Y^n$ 与被传输的码字并不是联合典型的，或者存在其他码字与 $Y^n$ 是联合典型的。正如证明联合 AEP，被传输的码字与接收到的序列是联合典型的概率趋于 1。对任意一个竞争码字，它与接收到的序列是联合典型的概率大约为 $2^{-nI}$，因此，可以使用大约 $2^{nI}$ 个码字，并且仍然保持很低的误差概率。稍后我们会推广这个论述来寻求一个码使得最大误差概率很低。

*误差概率的具体计算*：设 $W$ 服从 $\{1,2,\cdots,2^{nR}\}$ 上的均匀分布，并且利用步骤 6 中描述的联合典型译码 $\hat{W}(y^n)$。设 $\mathcal{E} = \{\hat{W}(Y^n) \neq W\}$ 表示误差事件。现在计算平均误差概率，这里的平均取自码簿中的所有码字以及所有码簿。也就是计算

$$\Pr(\mathcal{E}) = \sum_C \Pr(C) P_e^{(n)}(C) \tag{7-65}$$

$$= \sum_C \Pr(C) \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \lambda_w(C) \tag{7-66}$$

$$= \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \sum_C \Pr(C) \lambda_w(C) \tag{7-67}$$

其中 $P_e^{(n)}(C)$ 是针对联合典型译码定义的。根据码构造的对称性，取自所有码上的平均误差概率并不依赖于发送的具体下标，也就是说，$\sum_C \Pr(C)\lambda_w(C)$ 不依赖于 $w$。于是，不失一般性，可以假定发送的消息是 $W = 1$，这是由于

$$\Pr(\mathcal{E}) = \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \sum_C \Pr(C) \lambda_w(C) \tag{7-68}$$

$$= \sum_C \Pr(C) \lambda_1(C) \tag{7-69}$$

$$= \Pr(\mathcal{E} \mid W = 1) \tag{7-70}$$

定义下列事件：

202 $$E_i = \{(X^n(i), Y^n) \text{ 在 } A_\varepsilon^{(n)} \text{ 中}\},\ i \in \{1,2,\cdots,2^{nR}\} \tag{7-71}$$

其中 $E_i$ 表示第 $i$ 个码字与 $Y^n$ 为联合典型的这一事件。回忆一下，$Y^n$ 是在信道上发送第一个码字 $X^n(1)$ 而得到的结果。

如果 $E_1^c$ 发生(当传输的码字与接收到的序列是非联合典型时)，或者 $E_2 \cup E_3 \cup \cdots \cup E_{2^{nR}}$ 发生(当一个错误的码字与接收到的序列是联合典型时)，则在译码时会出现错误。因此，设 $P(\mathcal{E})$ 表示 $\Pr(\mathcal{E} \mid W = 1)$，根据事件之并，我们有

$$\Pr(\mathcal{E} \mid W = 1) = P(E_1^c \cup E_2 \cup E_3 \cup \cdots \cup E_{2^{nR}} \mid W = 1) \tag{7-72}$$

$$\leqslant P(E_1^c \mid W = 1) + \sum_{i=2}^{2^{nR}} P(E_i \mid W = 1) \tag{7-73}$$

由联合 AEP 的性质，$P(E_1^c)\to 0$，因而

$$P(E_1^c \mid W=1)\leqslant \varepsilon \quad \text{对充分大的 } n \tag{7-74}$$

从编码的生成过程可以看出，$X^n(1)$与 $X^n(i)(i\neq 1)$是独立的，所以 $Y^n$ 与$X^n(i)$也是独立的。因此，根据联合 AEP 的性质，$X^n(i)$与 $Y^n$ 是联合典型的概率$\leqslant 2^{-n(I(X;Y)-3\varepsilon)}$。从而，如果 $n$ 充分大且$R<I(X;Y)-3\varepsilon$ 时，

$$\Pr(\mathcal{E}) = \Pr(\mathcal{E} \mid W=1) \leqslant P(E_1^c \mid W=1) + \sum_{i=2}^{2^{nR}} P(E_i \mid W=1) \tag{7-75}$$

$$\leqslant \varepsilon + \sum_{i=2}^{2^{nR}} 2^{-n(I(X;Y)-3\varepsilon)} \tag{7-76}$$

$$= \varepsilon + (2^{nR}-1)2^{-n(I(X;Y)-3\varepsilon)} \tag{7-77}$$

$$\leqslant \varepsilon + 2^{3n\varepsilon}2^{-n(I(X;Y)-R)} \tag{7-78}$$

$$\leqslant 2\varepsilon \tag{7-79}$$

因此，如果 $R<I(X;Y)$，可以选取适当的 $\varepsilon$ 和 $n$，使得取自所有码簿和码字上的平均误差概率小于 $2\varepsilon$。

为了完成这个证明，通过选取一系列码来加强该结论。

1. 将证明中的 $p(x)$变为 $p^*(x)$，即达到信道容量时关于 $X$ 的分布。此时，条件 $R<I(X;Y)$ 可由可达性条件 $R<C$ 所替代。 203
2. 去除码簿上的平均。由于在所有码簿上的平均误差概率比较小($\leqslant 2\varepsilon$)，所以至少存在一个码簿$\mathcal{C}^*$具有小的平均误差概率。于是，$P_e^n(\mathcal{E}|\mathcal{C}^*)\leqslant 2\varepsilon$。若想找到$\mathcal{C}^*$可以穷举搜索所有的$(2^{nR},n)$码。注意到

$$\Pr(\mathcal{E} \mid \mathcal{C}^*) = \frac{1}{2^{nR}}\sum_{i=1}^{2^{nR}} \lambda_i(\mathcal{C}^*) \tag{7-80}$$

这是因为我们以式(7-63)中给定的均匀分布选取 $\hat{W}$。

3. 抛弃最佳码簿$\mathcal{C}^*$中最差的一半码字。由于这个码的算术平均误差概率 $P_e^n(\mathcal{C}^*)$小于 $2\varepsilon$，我们有

$$\Pr(\mathcal{E}|\mathcal{C}^*)\leqslant \frac{1}{2^{nR}}\sum\lambda_i(\mathcal{C}^*)\leqslant 2\varepsilon \tag{7-81}$$

这说明至少有一半的下标 $i$ 及其对应的码字$X^n(i)$的条件误差概率 $\lambda_i$ 小于$4\varepsilon$(否则，这些码字本身的和就将大于 $2\varepsilon$)。因此，所有码字中最佳的一半的最大误差概率必定小于 $4\varepsilon$。如果重新检索这些码字，会有 $2^{nR-1}$个码字。抛弃一半码字使得码率由 $R$ 变为$R-\frac{1}{n}$，当 $n$ 充分大时，这是可忽略的。

结合所有这些改进，我们已经构造了一个码率为 $R'=R-\frac{1}{n}$的码，它的最大误差概率 $\lambda^{(n)}\leqslant 4\varepsilon$。这就证明了任何小于信道容量的码率都是可达的。 □

可以看出，随机编码是证明定理 7.7.1 的方法，而不是发送信号的方法。在证明中码被随机选择仅是为了达到数学上的对称性以及一个好的确定性码的存在性。我们证明了分组长度为 $n$ 的所有码上的平均有较小的误差概率。通过穷举搜索，也可以找到这个集合中的最佳码。顺便提及一下，这也表明了最佳码的科尔莫戈罗夫复杂度(见第 14 章)是一个小常数。这意味着将最佳码$\mathcal{C}^*$告知发送者和接收者(在步骤 2 中)并不需要使用信道。发送者与接收者仅需要同意在信道中使用最佳$(2^{nR},n)$码就可以了。

虽然这个定理说明了对于大的分组长度，存在误差概率任意小的好码，但它并没有提供一种
204 构造最佳码的方法。如果使用定理证明中的方法，根据适当的分布随机地生成一个码，那么对于充分大的分组长度，这样构造出来的编码可能是很好的。然而，由于该编码中缺乏某个结构，译码将是非常困难的(简单的查表方法也需要一个指数级大小的表)。因此，这个定理并不能提供一个实际的编码方案。自香农在信息论方面的开篇之作问世以来，研究者们试图发掘易于编和译的构造性编码。在7.11节将讨论一种最简单的代数纠错码——汉明(Hamming)码，它能在每个比特分组中纠正一个错。自香农的论文发表以来，各种各样的技术涌现出来用于构造纠错码，特别是turbo码接近了高斯信道容量。

## 7.8 零误差码

在允许完全无误差的情况下，审视上面定理的论证过程，显然可以极大地启发我们对于逆定理的简要证明。首先证明 $P_e^{(n)}=0$ 蕴含结论 $R\leqslant C$。假定有一个零误差概率的$(2^{nR},n)$码，也就是说，译码器输出的 $g(Y^n)$以概率1等于输入的下标 $W$。那么，输入下标 $W$ 完全由输出序列决定(即 $H(W|Y^n)=0$)。为了获得更强的界，随意假定 $W$ 服从$\{1,2,\cdots,2^{nR}\}$上的均匀分布，于是，$H(W)=nR$。从而，我们有如下的一串不等式：

$$nR=H(W)=\underbrace{H(W|Y^n)}_{=0}+I(W;Y^n) \tag{7-82}$$

$$=I(W;Y^n) \tag{7-83}$$

$$\overset{(a)}{\leqslant} I(X^n;Y^n) \tag{7-84}$$

$$\overset{(b)}{\leqslant} \sum_{i=1}^{n} I(X_i;Y_i) \tag{7-85}$$

$$\overset{(c)}{\leqslant} nC \tag{7-86}$$

其中(a)由数据处理不等式推出(由于 $W\to X^n(W)\to Y^n$ 形成马尔可夫链)，(b)会在引理7.9.2中借助离散无记忆假设得到证明，(c)直接由(信息)容量的定义推出。因此，对任何零误差的$(2^{nR},n)$码及所有的 $n$，

205

$$R\leqslant C \tag{7-87}$$

## 7.9 费诺不等式与编码定理的逆定理

下面将零误差码的证明过程推广到具有非常小误差概率的编码。证明中需要的新工具就是费诺不等式，它依据条件熵给出误差概率的下界。回忆一下费诺不等式的证明，为便于参考，将它重述如下。

先给出一些定义。下标 $W$ 服从集合$\mathcal{W}=\{1,2,\cdots,2^{nR}\}$上的均匀分布，序列 $Y^n$ 与 $W$ 是概率相关的。通过 $Y^n$ 来估计被发送的下标 $W$。设 $\hat{W}=g(Y^n)$为其估计，那么，$W\to X^n(W)\to Y^n\to \hat{W}$形成马尔可夫链。注意到误差概率为

$$\Pr(\hat{W}\neq W)=\frac{1}{2^{nR}}\sum_i \lambda_i = P_e^{(n)} \tag{7-88}$$

我们先给出下面的引理，它的证明在2.10节中。

**引理7.9.1(费诺不等式)** 设离散无记忆信道的码簿为 $C$，且输入消息 $W$ 服从$2^{nR}$上的均匀分布，则有

$$H(W|\hat{W})\leqslant 1+P_e^{(n)}nR \tag{7-89}$$

**证明**：由于 $W$ 服从均匀分布，则有 $P_e^{(n)}=\Pr(W\neq\hat{W})$。对大小为 $2^{nR}$ 的字母表中的 $W$ 应用费诺不等式(定理 2.10.1)，可得到引理的证明。 □

现在证明下面的引理，它说明如果多次使用离散无记忆信道，每次传输的容量并不增加。

**引理 7.9.2** 设 $Y^n$ 为 $X^n$ 经过容量 $C$ 离散无记忆信道传输所得到的输出信号。则

$$I(X^n;Y^n)\leqslant nC \quad \text{对于任意的 } p(x^n) \tag{7-90}$$

**证明**：由离散无记忆信道的定义，$Y_i$ 仅依赖于 $X_i$ 而与其他所有变量都是条件独立的。所以有

$$I(X^n;Y^n) = H(Y^n) - H(Y^n \mid X^n) \tag{7-91}$$

$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid Y_1,\cdots,Y_{i-1},X^n) \tag{7-92}$$

$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid X_i) \tag{7-93}$$ 206

继续该系列不等式，我们有

$$I(X^n;Y^n) = H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid X_i) \tag{7-94}$$

$$\leqslant \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i \mid X_i) \tag{7-95}$$

$$= \sum_{i=1}^{n} I(X_i;Y_i) \tag{7-96}$$

$$\leqslant nC \tag{7-97}$$

其中式(7-95)基于如下事实得到：一族随机变量的熵小于各自熵的和。式(7-97)直接由容量的定义推出。这样，就证明了多次使用信道并不增加每次传输的信息容量比特。 □

现在我们已经有充分的准备来证明信道编码定理中的逆定理。

**证明**：定理 7.7.1(信道编码定理)的逆定理。我们要证明，对任何满足 $\lambda^{(n)}\to 0$ 的 $(2^{nR},n)$ 码序列，必有 $R\leqslant C$。如果最大误差概率趋于 0，那么这个码序列的平均误差概率也趋于 0，即 $\lambda^{(n)}\to 0$ 蕴含 $P_e^{(n)}\to 0$，其中 $P_e^{(n)}$ 的定义见式(7-32)。对固定的编码规则 $X^n(\cdot)$ 和固定的译码规则 $\hat{W}=g(Y^n)$，我们有 $W\to X^n(W)\to Y^n\to\hat{W}$。对每个 $n$，设 $W$ 服从 $\{1,2,\cdots,2^{nR}\}$ 上的一个均匀分布。由于 $W$ 服从均匀分布，故 $\Pr(\hat{W}\neq W)=P_e^{(n)}=\frac{1}{2^{nR}}\sum_i \lambda_i$。因此，

$$nR \overset{(a)}{=} H(W) \tag{7-98}$$

$$\overset{(b)}{=} H(W\mid\hat{W}) + I(W;\hat{W}) \tag{7-99}$$

$$\overset{(c)}{\leqslant} 1 + P_e^{(n)}nR + I(W;\hat{W}) \tag{7-100}$$

$$\overset{(d)}{\leqslant} 1 + P_e^{(n)}nR + I(X^n;Y^n) \tag{7-101}$$ 207

$$\overset{(e)}{\leqslant} 1 + P_e^{(n)}nR + nC \tag{7-102}$$

其中，(a)由 $W$ 服从 $\{1,2,\cdots,2^{nR}\}$ 上的均匀分布假设推出，(b)是一个恒等式，(c)是由于 $W$ 至多可取 $2^{nR}$ 个值而获得的费诺不等式，(d)为数据处理不等式，而 (e)由引理 7.9.2 推出。两边同除 $n$，得到

$$R\leqslant P_e^{(n)}R + \frac{1}{n} + C \tag{7-103}$$

现在令 $n\to\infty$，则不等式右边的前两项趋于 0，因此

$$R \leqslant C \tag{7-104}$$

可以将式(7-103)改写为

$$P_e^{(n)} \geqslant 1 - \frac{C}{R} - \frac{1}{nR} \tag{7-105}$$

该式表明，当 $R>C$ 时，对充分大的 $n$，误差概率无法接近于0(从而对所有的 $n$ 都是成立的，因为如果对小的 $n$ 有 $P_e^{(n)}=0$，那么通过串联这些码来构造对大的 $n$ 也满足 $P_e^{(n)}=0$ 的码)。因此，当码率大于容量时，不可能达到任意低的误差概率。 □

上述逆定理有时称作信道编码定理的*弱逆定理*(weak converse)。也可以证明一个*强逆定理*(strong converse)，它说明当码率大于容量时，误差概率以指数级趋于1。因此，信道容量很明显是一个分界点——当码率小于容量时，以指数级有 $P_e^{(n)} \to 0$；而当码率大于容量时，以指数级有 $P_e^{(n)} \to 1$。

## 7.10 信道编码定理的逆定理中的等式

我们已经证明了信道编码定理和它的逆定理。从本质上讲，这些定理表明当 $R<C$ 时，可以以任意低的误差概率传输信息；而当 $R>C$ 时，误差概率将远离0。

探讨逆定理中的等式成立的结果是一件很有趣而且有价值的事情，这有望启发我们找出达到信道容量的编码。在 $P_e=0$ 的情况下，重复逆定理中的步骤，我们有

208

$$\begin{aligned}
nR &= H(W) && (7\text{-}106)\\
&= H(W \mid \hat{W}) + I(W;\hat{W}) && (7\text{-}107)\\
&= I(W;\hat{W})) && (7\text{-}108)\\
&\overset{(a)}{\leqslant} I(X^n(W);Y^n) && (7\text{-}109)\\
&= H(Y^n) - H(Y^n \mid X^n) && (7\text{-}110)\\
&= H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid X_i) && (7\text{-}111)\\
&\overset{(b)}{\leqslant} \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i \mid X_i) && (7\text{-}112)\\
&= \sum_{i=1}^{n} I(X_i;Y_i) && (7\text{-}113)\\
&\overset{(c)}{\leqslant} nC && (7\text{-}114)
\end{aligned}$$

只有当 $I(Y^n;X^n(W)\mid W)=0$ 以及 $I(X^n;Y^n\mid \hat{W})=0$ 时，数据处理不等式(a)中的等号才成立。如果所有码字都不同，而且 $\hat{W}$ 是译码的一个充分统计量，这是成立的。只有当 $Y_i$ 相互独立时，(b)中等式才能成立；只有当 $X_i$ 的分布是 $p^*(x)$ 时，即达到信道容量的 $X$ 上的分布时，(c)中等式才能成立。所以，只有当所有这些条件都满足时，才能得到逆定理中的等式。这说明对于达到信道容量的零误差码，其码字必须互不相同，且所有 $Y_i$ 的分布 i.i.d. 服从

$$p^*(y) = \sum_x p^*(x)p(y \mid x) \tag{7-115}$$

这是由 $X$ 的最优分布导出的 $Y$ 分布。在逆定理中涉及到的分布是由码字上的均匀分布诱导出的 $X$ 和 $Y$ 的经验分布，即

$$p(x_i,y_i) = \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} I(X_i(w)=x_i)p(y_i \mid x_i) \tag{7-116}$$

我们可以用一些达到信道容量的编码例子来检验这一结果：

1. 有噪声打字机信道。此时，输入字母表是由26个英文字母构成的，每一个字母能够正确地输出，或者变为下一个字母的概率均是1/2。达到信道容量(log13比特)的一个简单码是使用间隔的输入字母，这样就不会使两个字母相互混淆。此时，就有了13个分组长度为1的码字。如果挑选出其中一些码字的i.i.d服从{1,3,5,7,…,25}上的均匀分布，那么正如我们所期望的，这个信道的输出也是i.i.d.服从{1,2,…,26}上的均匀分布。 209
2. 二元对称信道。由于对给定任意输入序列，每一个可能的输出序列都具有正的概率，所以即使只有两个码字也不可能以零误差概率区分开它们。故BSC的零误差容量是0。然而，即使在这种情况下，还是可以得出一些有用的结论。有效码仍然可以导出关于 $Y$ 的分布，使得 $Y$ 看起来是i.i.d.服从Bernoulli$\left(\frac{1}{2}\right)$。并且，从逆定理的证明中也可以看出，当码率接近信道容量时，利用对应于码字的译码集，已经几乎完全覆盖了所有可能的输出序列的集合。当码率大于信道容量时，译码集变得相互重叠，并且误差概率不可能再任意小。

## 7.11 汉明码

信道编码定理使用分组码的方案。如果分组长度足够大的话，当码率小于信道容量时，可以用分组码以任意低的误差概率传输信息。自香农开创性的论文[471]问世以来，人们一直在寻找这样的码。除了要达到低的误差概率之外，实用的编码应该是“简单”的，以保证它们可以有效地编码和译码。

自香农1948年开创性的论文发表以来，为了寻找简单而优秀的编码工作已经持续了很长的时间。在寻找的过程中，人们发展出了一套完整的编码理论。我们无法逐一描述自从1948年以来所发明的众多精致而且复杂的编码方案。在这里仅介绍由汉明开发的一种最简单的方案[266]。它可以说明大多数码所共有的一些最基本的思想。

编码的目的是通过增加冗余使得在一些信息损失或者损坏的情况下仍可能由接收者恢复出原始的消息。最显而易见的一种编码方案是重复信息。例如，为发送一个1，我们发送11111，为发送一个0，我们发送00000。这一方案使用5个字符来传输1比特，因此码率为1/5比特/字符。如果在二元对称信道中使用这样的码，最优的译码方案就是将接收到的每个5比特分组译为其中占多数的比特。如果3个或者更多的比特是1，我们则将这个分组译为1；否则将其译为0。当且仅当超过3个比特发生改变时，才会出现错误。通过使用更长的重复码，可以达到任意小的误差概率。但是，随着分组长度的增加，码率也趋于0，因此，一个“简单的”编码，不一定是一个非常实用的编码。 210

替代这种简单的重复比特方法，可以用某种巧妙的方式将比特联合起来，使得每一个额外的比特都可以用来检验某个信息比特子集中是否发生错误。一个简单的例子就是奇偶校验码。从 $n-1$ 个信息比特的分组出发，选取第 $n$ 个比特，使得整个分组的奇偶校验数为0(分组中1的个数为偶数)。这样，如果在传输过程中发生了奇数次错误，那么接收者将能够注意到奇偶性的变化，并察觉到错误。这是检错码(error-detecting code)的最简单的例子。该编码既不能察觉到出现偶数次错误，也不能提供任何有关纠正这些错误的信息。

我们可以推广奇偶校验的思想，允许存在多个奇偶校验位，也可以允许奇偶校验依赖于各种各样的信息比特子集。下面将描述的汉明码是奇偶校验码的一个例子。利用线性代数中的一些简单思想来描述它。

为说明汉明码的基本思想，考虑分组长度为7的二元码。所有的运算都是模2运算。考虑所

有长度为3的非0二元向量的集合，以它们为列向量构成一个矩阵：

$$H=\begin{bmatrix}0&0&0&1&1&1&1\\0&1&1&0&0&1&1\\1&0&1&0&1&0&1\end{bmatrix} \tag{7-117}$$

考虑 $H$ 的零空间(即与 $H$ 相乘得到000的向量)中长度为7的向量的集合。由线性空间理论，因为 $H$ 的秩为3，故期望 $H$ 的零空间的维数为4。这 $2^4$ 个码字如下

0000000 0100101 1000011 1100110
0001111 0101010 1001100 1101001
0010110 0110011 1010101 1110000
0011001 0111100 1011010 1111111

由于这个码字集是矩阵的零空间，所以从任意两个码字的和仍是一个码字的意义上看，这是
211 线性的。因此，码字集形成7维向量空间中的一个4维线性子空间。

观察这些码字，不难注意到除了全是0的码字外，任何码字中1的最小数目为3。该最小数称为码的**最小重量**(minimun weight)。可以看出，由于 $H$ 的所有列互不相同，没有两列的和可以为000，因此码的最小重量至少为3。基于任意两列的和必然为该矩阵中某一列的事实，我们可以推出最小距离恰好为3。

由于该码是线性的，任意两个码字的差仍是一个码字，因此，任意两个码字之间至少在3个位置上有所不同。两个码字不同的最小位置数称为该码的**最小距离**(minimum distance)。码的最小距离是用来表示码字之间相隔多远的一个度量，并且可以决定在信道的输出端码字之间差异的程度。对线性码来说，最小距离等于最小重量。我们的目的是设计出最小距离尽可能大的码。

上述码的最小距离是3。因此，如果码字 $\mathbf{c}$ 仅占一个位置损坏，那么产生的新字符串将与其他任何码字之间至少在两个位置上是不同的，它与 $\mathbf{c}$ 更加接近。但是，是否可以不通过穷举搜索就可以发现哪一个是距离最近的码字呢?

回答是肯定的，可以利用矩阵 $H$ 的结构译码。矩阵 $H$ 称作**奇偶校验矩阵**(parity check matrix)并具有如下性质：对任意码字 $\mathbf{c}$ 均有 $H\mathbf{c}=0$。设 $\mathbf{e}_i$ 是第 $i$ 个位置为1其余位置为0的向量。如果码字的第 $i$ 个位置损坏，则接收到的向量为 $\mathbf{r}=\mathbf{c}+\mathbf{e}_i$。如果将矩阵 $H$ 与这个接收到的向量相乘，则得到

$$H\mathbf{r}=H(\mathbf{c}+\mathbf{e}_i)=H\mathbf{c}+H\mathbf{e}_i=H\mathbf{e}_i \tag{7-118}$$

这正好是 $H$ 的第 $i$ 列向量。因此，通过计算 $H\mathbf{r}$，就可以发现接收向量的哪一个位置损坏了。还原该位置上的值就得到一个码字。这样就有了一个简单的程序用来纠正接收序列中的一个错误。我们已经构造出分组长度为7的16个码字组成的码簿，它能纠正至多一个错误。这个码就是**汉明码**(Hamming code)。

至此，我们还没有给出一个简单的编码程序；可以考虑16条消息的集合到码字集合的映射。但是，当仔细检查表中所有码字的前4位之后，将会观察到它们正好构成了4个比特的所有 $2^4$ 种组合。于是，可以将这4个比特看作是要发送消息的4个比特，而另3个比特由编码决定。对于
212 一般情形，将线性码进行修改，可以使得映射更加明显：让码字中的前 $k$ 个比特代表消息，而后面 $n-k$ 个比特留作奇偶校验位。这样得到的编码称作**系统码**(systematic code)。该码往往由它的分组长度 $n$，信息比特数 $k$ 以及最小距离 $d$ 三个参数来确定。例如，上述编码称作(7,4,3)汉明码，即 $n=7$，$k=4$ 和 $d=3$。

可以利用简单的文氏图(Venn Diagram)表示来解释汉明码的工作原理。考虑如下文氏图，它有三个圆和四个相交区域，如图7-10所示。为了发送信息序列1101，将序列中的4个信息比特

分别放在图中四个相交的区域中。然后在三个剩余的区域中各放置一个校验位使得每个圆中的校验为偶数(即每个圆中有偶数个1)。于是，校验位就变成如图7-11中所示。

现在不妨设其中的一个比特被改变了。例如，图7-12中有一个信息比特从1变成了0。此时，有两个圆违背了原先的校验约束(图中加黑部分)。因而，当我们知道了这两个约束违背，不难看出，导致产生约束违背的这个单一的比特错误只可能在两圆的相交部分发生(即改变的那个比特)。类似地，通过分析其他情形，也不难看出，这种码可以检测并纠正发生在接收到码字中的任何单个比特错误。

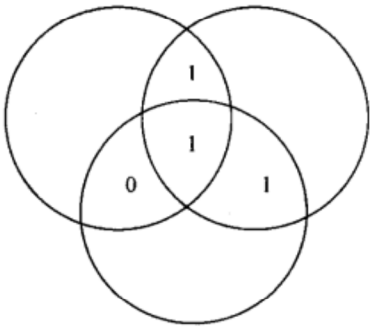


图7-10 信息比特的文氏图

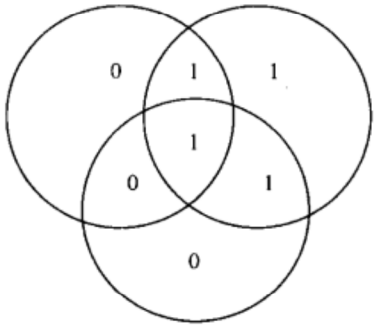


图7-11 每个圆的信息比特与带偶校验的校验位的文氏图

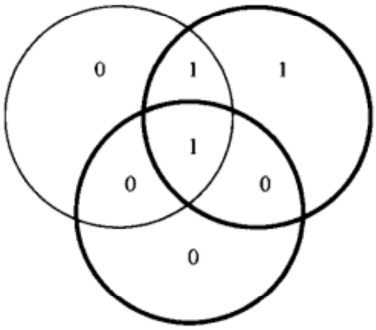


图7-12 一个信息比特改变后的文氏图

很容易推广这一程序来构造更大的矩阵 $H$。一般来说，如果使用矩阵 $H$ 中的 $l$ 行，那么所得编码的分组长度为 $n=2^l-1$，$k=2^l-l-1$，以及最小距离为3。所有这些码都称作汉明码，并可以纠正一个错误。 213

汉明码是所有线性奇偶校验码中最简单的例子。通过汉明码说明了构造其他线性码的基本原则。但是，当分组长度较大时，分组中很可能会出现不止一个错误。在20世纪50年代早期，里德(Reed)和所罗门(Solomon)针对非二元信道，发明了一类多重纠错码。20世纪50年代后期，Bose，Ray-Chaudhuri [72]和 Hocquenghem [278]利用伽罗瓦(Galois)域论推广了汉明码的思想，从而构造出针对任意 $t$ 的 $t$ 纠错码(称作 BCH 码)。自那时起，许多作者开发出了许多其他的编码以及这些码的有效译码算法。随着集成电路技术的发展，现在已经可以在硬件中实施相当复杂的编码，并且 214
能够部分实现香农的信道容量定理中所预言的纠错能力。例如，所有 CD 播放器都配置有基于两个交织的(interleaved)(32,28,5)和(28,24,5)R-S码的纠错电路，可以纠正大约4000个脉冲错误。

上面描述的所有码都是分组码(block code)——将一组信息比特映射成一个信道码字，且不依赖于过去的信息比特。也可以设计出这样的码：每个输出组不仅依赖于当前的输入组，而且依赖于过去的一些输入组。这种码的一个高级结构化的形式称作卷积码(convolutional code)。卷积码理论在过去的40年里得到了相当大的发展。这里不再深入讨论，但是有兴趣的读者可以参考编码理论的教科书[69,356]。

在设计出的编码算法当中，经历了很多年，没有一种编码算法能够接近香农信道容量定理中所给出的界。对一个交叉概率为 $p$ 的二元对称信道，我们需要一种码，它能在长度为 $n$ 且占 $n(1-H(p))$ 个信息比特的分组中纠正多达 $np$ 个错误。例如，在长度为 $n$ 的分组中，前面提及的重复码可以纠正多达 $n/2$ 个错误，但是它的码率随着 $n$ 的增大而趋于0。在1972年以前，对于能够在长度为 $n$ 的分组中纠正 $na$ 个错误的编码，它们的码率都渐近于0。而到1972年，Justesen [301]设计出了一类码，具有正的渐近码率和正的渐近最小距离，并且都与分组长度成正比。

到了1993年，Berrou等人在文章[57]中提出了下列观点：将两个交织卷积码与一个并行协作的译码器组合起来能获得远比此前任何码更好的效果。每个译码器将自身对每个比特值的“意见”反馈给另一个译码器，并利用该译码器的意见来帮助确定自身的这个比特值。这种迭代过程不停地重复，直到两个译码器都对比特的取值达成共识为止。令人惊讶的是，这个迭代程序对于许多信道都能在接近于容量的码率下进行有效地译码。这也重新提升了学者们对Robert Gallager在其学位论文[231, 232]中引入的低密度奇偶性校验(low-density parity check, LDPC)码的研究兴趣。1997年，MacKay与Neal [368]证明了对于LDPC码，迭代的消息传输算法(类似于用来译解turbo码的算法)可以使码率以很高的概率达到信道容
215 量。至今，turbo码与LDPC码仍然是研究的热点，并且应用在无线通信和卫星通信信道中。

## 7.12 反馈容量

带反馈的信道如图7-13所示。假定所有接收到的字符立即以无噪声的方式传输回发送者，这样，发送者可以利用它们来决定下面将要发送哪一个字符。反馈会给我们带来好处吗？令人吃惊的是，回答为否定。现在来证明。我们把$(2^{nR}, n)$反馈码(feedback code)定义为一个映射序列$x_i(W, Y^{i-1})$和一个译码函数序列$g: \mathcal{Y}^n \to \{1,2,\cdots,2^{nR}\}$，其中$x_i$是仅与消息$W \in 2^{nR}$和先前接收到的值$Y_1, Y_2, \cdots, Y_{i-1}$的函数。于是，当$W$服从$\{1,2,\cdots,2^{nR}\}$上的均匀分布时，有

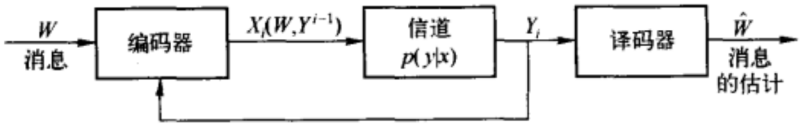


图7-13 带反馈的离散无记忆信道

$$P_e^{(n)} = \Pr\{g(Y^n) \neq W\} \tag{7-119}$$

**定义** 离散无记忆信道的带反馈容量$C_{FB}$(capacity with feedback)定义为反馈码可以达到的所有码率的上确界。

**定理** 7.12.1(反馈容量)

$$C_{FB} = C = \max_{p(x)} I(X;Y) \tag{7-120}$$

**证明：** 由于非反馈码是反馈码的特例，不带反馈能够达到的任何码率也可以通过带反馈的方式达到，因此

$$C_{FB} \geqslant C \tag{7-121}$$

证明相反的不等式稍微复杂一些。无法再直接使用证明不带反馈的编码逆定理中给出的方法。由于$X_i$依赖于过去接收到的字符，引理7.9.2不再成立，而且式(7-93)中的结论(即$Y_i$仅依赖
216 于$X_i$且条件独立于未来的$X$的结论)也不再成立。

但是，只要经过简单的修改，原来的方法依然起作用；取代$X^n$，我们使用下标$W$，则可以证明类似的系列不等式。设$W$服从$\{1,2,\cdots,2^{nR}\}$上的均匀分布，则$\Pr\{W \neq \hat{W}\} = P_e^{(n)}$，根据费诺不等式和数据处理不等式，我们有

$$nR = H(W) = H(W|\hat{W}) + I(W;\hat{W}) \tag{7-122}$$

$$\leqslant 1 + P_e^{(n)} nR + I(W;\hat{W}) \tag{7-123}$$

$$\leqslant 1 + P_e^{(n)} nR + I(W;Y^n) \tag{7-124}$$

下面我们可以估计 $I(W;Y^n)$的界如下：

$$I(W;Y^n) = H(Y^n) - H(Y^n \mid W) \tag{7-125}$$

$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid Y_1, Y_2, \cdots, Y_{i-1}, W) \tag{7-126}$$

$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid Y_1, Y_2, \cdots, Y_{i-1}, W, X_i) \tag{7-127}$$

$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid X_i) \tag{7-128}$$

这是由于 $X_i$ 是关于 $Y_1, Y_2, \cdots, Y_{i-1}$ 和 $W$ 的函数；以及在给定 $X_i$ 的条件下，$Y_i$ 独立于 $W$ 和 $Y$ 的过去样本。由离散无记忆信道容量的定义，我们可以得到

$$I(W;Y^n) = H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid X_i) \tag{7-129}$$

$$\leqslant \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i \mid X_i) \tag{7-130}$$

$$= \sum_{i=1}^{n} I(X_i;Y_i) \tag{7-131}$$

$$\leqslant nC \tag{7-132}$$

综合上述，可得

$$nR \leqslant P_e^{(n)} nR + 1 + nC \tag{7-133}$$

217

两边同时除以 $n$ 并令 $n \to \infty$，得到

$$R \leqslant C \tag{7-134}$$

于是，使用反馈并不能带给我们更高的码率，即

$$C_{FB} = C \tag{7-135}$$□

正如我们在二元擦除信道的例子中看到的那样，反馈在简化编码和译码方面可以起到很大的作用。然而，它并不能增加信道的容量。

## 7.13 信源信道分离定理

现在是将已经证明的两个主要结果结合在一起的时候了：数据压缩($R > H$：定理 5.4.2)和数据传输($R < C$：定理 7.7.1)。为了通过信道传输信源，条件 $H < C$ 是充分必要的吗？例如，考虑通过离散无记忆信道传输数字语音或音乐。设计一个码将语音样本序列直接映射成信道的输入信号，或者先将语音压缩成最有效的格式，然后使用适当的信道编码从该信道将它发送出去。由于数据压缩不依赖于信道，而信道编码又不依赖于信源分布，因此，如果使用两步骤方法，我们并不十分清楚会不会损失一些信息。

在这节中我们将证明：在有噪声信道中，两步骤方法与其他传输信息的方法一样有效。该结果有一些重要的实际应用。这意味着可以将通信系统的设计转化成信源编码与信道编码两个部分的组合。为数据最有效的表达设计信源码，也能够分离独立地设计适合于信道的信道码。这种组合的方法与将两个问题一起考虑所能设计出的任何方法一样有效。

数据的通常表示是使用二元字母表。最现代的通信系统是数字化的，并且为了能在通常的信道上传输，数据简化为二进制表示。这使复杂度大大减小。像 ATM 和因特网这样的网络系统允许语音、视频和数字数据共用相同的通信信道。

218

两步骤处理与任何一步骤处理都一样有效。虽然这一结论看上去是那么显然，但有必要提

醒读者，这未必总是正确的。例如，在某些多用户信道中，这种分解是不可行的。我们也将考虑两个简单的情形，这时定理看上去会有误导性。简单的例子是通过擦除信道发送英文文本。首先找出文本最有效的二进制表示，然后通过信道发送它。这时，发生的错误将很难译码。如果直接发送这个英文文本，虽然会损失大约一半的字母，但仍然可以知道文本的含义。类似地，人类的耳朵有一些非同寻常的能力，如果噪声是白色的，可以在非常高的噪声水平下分辨出语音。在这种情况下，直接通过有噪声信道发送未被压缩的语音会比发送压缩的语音更加合适。明显地，信源中的冗余适应于信道。

现在对上述问题做个严格的定义。假设有一个信源 $V$，从字母表$\mathcal{V}$中生成字符。对于由 $V$ 生成的随机过程，除了要求其取值于有限字母表且满足 AEP 之外，不做任何假设。这种过程的例子包括独立同分布的随机变量序列和平稳不可约马尔可夫链的状态序列。任何平稳遍历信源均满足 AEP，这将在 6.8 节中证明。

现在想通过信道发送字符序列 $V^n = V_1, V_2, \cdots, V_n$，并且保证接收者可以重构序列。为了达到这个目的，将序列映射成码字 $X^n(V^n)$，通过信道发送这个码字。接收者观察接收到的序列 $Y^n$后，给出发送序列 $V^n$ 的估计$\hat{V}^n$。如果 $V^n \neq \hat{V}^n$，则接收者犯了错误。我们定义误差概率为

$$\Pr(V^n \neq \hat{V}^n) = \sum_{y^n} \sum_{v^n} p(v^n) p(y^n \mid x^n(v^n)) I(g(y^n) \neq v^n) \tag{7-136}$$

其中 $I$ 为示性函数，$g(y^n)$是译码函数。这个系统如图 7-14 所示。

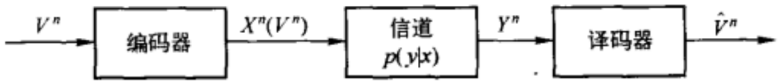


图 7-14 联合信源信道编码

219 下面给出联合信源信道编码定理：

**定理** 7.13.1(信源信道编码定理) 如果 $V_1, V_2, \cdots, V_n$ 为有限字母表上满足 AEP 和 $H(\mathcal{V}) < C$ 的随机过程，则存在一个信源信道编码使得误差概率 $\Pr(\hat{V}^n \neq V^n) \to 0$。反之，对任意平稳随机过程，如果 $H(\mathcal{V}) > C$，那么误差概率远离 0，从而不可能以任意低的误差概率通过信道发送这个过程。

**证明：** 可达性。证明前半部分的精髓就是此前描述的两步骤编码。由于已经假定随机过程满足 AEP，所以必然存在一个元素个数$\leqslant 2^{n(H(\mathcal{V})+\varepsilon)}$的典型集 $A_\varepsilon^{(n)}$，它拥有概率的绝大部分。仅对属于这个典型集的信源序列进行编码；其余所有序列将产生一个错误。它对误差概率的贡献不会超过 $\varepsilon$。

给 $A_\varepsilon^{(n)}$ 中的所有序列加上下标。由于至多有 $2^{n(H+\varepsilon)}$个这样的序列，$n(H+\varepsilon)$比特足以给出它们的下标了。如果

$$H(\mathcal{V}) + \varepsilon = R < C \tag{7-137}$$

我们能以小于 $\varepsilon$ 的误差概率将需要的下标发送给接收者。接收者可以通过穷举典型集 $A_\varepsilon^{(n)}$，选择与被估计下标相应的序列，从而重构出 $V^n$。这个序列将以很高的概率与传输序列相一致。具体来说，对充分大的 $n$，我们有

$$P(V^n \neq \hat{V}^n) \leqslant P(V^n \notin A_\varepsilon^{(n)}) + P(g(Y^n) \neq V^n \mid V^n \in A_\varepsilon^{(n)}) \tag{7-138}$$

$$\leqslant \varepsilon + \varepsilon = 2\varepsilon \tag{7-139}$$

因此，如果

$$H(\mathcal{V})<C \tag{7-140}$$

那么对充分大的 $n$，我们能够以低的误差概率重构出序列。

*逆定理*。我们希望证明，对于任意的信源信道码序列

$$X^n(V^n):\mathcal{V}^n\to\mathcal{X}^n \tag{7-141}$$

$$g_n(Y^n):\mathcal{Y}^n\to\mathcal{V}^n \tag{7-142}$$

220

$\Pr(\hat{V}^n\neq V^n)\to 0$ 蕴含结论 $H(\mathcal{V})\leqslant C$。$X^n(\cdot)$是数据序列 $V^n$ 的任意（也许是随机的）码字分配，$g_n(\cdot)$是任何译码函数（对输出序列 $Y^n$ 的估计分配 $\hat{V}^n$）。根据费诺不等式，必有

$$H(V^n\mid\hat{V}^n)\leqslant 1+\Pr(\hat{V}^n\neq V^n)\log|\mathcal{V}^n|=1+\Pr(\hat{V}^n\neq V^n)n\log|\mathcal{V}| \tag{7-143}$$

因此，对于这个码，

$$H(\mathcal{V})\overset{(a)}{\leqslant}\frac{H(V_1,V_2,\cdots,V_n)}{n} \tag{7-144}$$

$$=\frac{H(V^n)}{n} \tag{7-145}$$

$$=\frac{1}{n}H(V^n\mid\hat{V}^n)+\frac{1}{n}I(V^n;\hat{V}^n) \tag{7-146}$$

$$\overset{(b)}{\leqslant}\frac{1}{n}(1+\Pr(\hat{V}^n\neq V^n)n\log|\mathcal{V}|)+\frac{1}{n}I(V^n;\hat{V}^n) \tag{7-147}$$

$$\overset{(c)}{\leqslant}\frac{1}{n}(1+\Pr(\hat{V}^n\neq V^n)n\log|\mathcal{V}|)+\frac{1}{n}I(X^n;Y^n) \tag{7-148}$$

$$\overset{(d)}{\leqslant}\frac{1}{n}+\Pr(\hat{V}^n\neq V^n)\log|\mathcal{V}|+C \tag{7-149}$$

其中(a)由平稳过程熵率的定义推出，(b)由费诺不等式得到，(c)由数据处理不等式（由于 $V^n\to X^n\to Y^n\to\hat{V}^n$ 构成马尔可夫链）得到，(d)由信道的无记忆性得出。令 $n\to\infty$，我们有 $\Pr(\hat{V}^n\neq V^n)\to 0$，因此

$$H(\mathcal{V})\leqslant C \tag{7-150}$$□

于是，我们能够通过信道传输平稳遍历信源当且仅当它的熵率小于信道容量。联合信源信道分离定理促使我们将信源编码问题从信道编码问题中独立出来考虑。信源编码器试图找到信源的最有效表示，而信道编码器编码消息要具备能够对抗信道中产生的噪声和错误的能力。分离定理表明，分离编码器（如图 7-15）与联合编码器（如图 7-14）能够达到相同的码率。

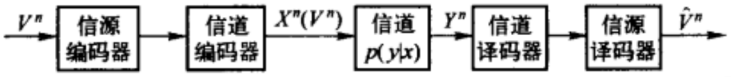


图 7-15　分离信源信道编码

由此结论，我们已经将信息论中的两个基本定理（数据压缩与数据传输定理）联系在了一起。接下来用几句话概括这两个结果的证明过程。数据压缩定理来源于 AEP，表明全部信源序列存在一个拥有了绝大部分概率的“小型”的子集（大小为 $2^{nH}$），根据这个子集使用 $H$ 比特/字符并以很小的误差概率来表示这个信源。数据传输定理基于联合的 AEP；它依据的事实是：对于大的分组长度，信道的输出序列非常有可能与输入码字是联合典型的，而任何其他码字是联合典型的概率约为 $2^{-nI}$。因而，我们可以使用大约 $2^{nI}$ 个码字而保持可忽略的误差概率。信源信道分离定理说明，我们可以独立地设计信源码和信道码，然后结合两者的结果以达到最优的效果。

221

## 要点

**信息容量** 可区分的输入信号数量的对数值由下面等式给出

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

**例子**

- 二元对称信道：$C=1-H(p)$。
- 二元擦除信道：$C=1-\alpha$。
- 对称信道：$C=\log|\mathcal{Y}|-H$(转移矩阵的行)。

**C 的性质**

1. $0\leqslant C\leqslant \min\{\log|\mathcal{X}|,\log|\mathcal{Y}|\}$。
2. $I(X;Y)$是关于 $p(x)$的连续凹函数。

**联合典型性** 服从分布 $p(x,y)$的联合典型序列$\{(x^n,y^n)\}$的集合 $A_\epsilon^{(n)}$ 为

$$A_\epsilon^{(n)}=\{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n:\tag{7-151}$$

222

$$\left|-\frac{1}{n}\log p(x^n)-H(X)\right|<\epsilon\tag{7-152}$$

$$\left|-\frac{1}{n}\log p(y^n)-H(Y)\right|<\epsilon\tag{7-153}$$

$$\left|-\frac{1}{n}\log p(x^n,y^n)-H(X,Y)\right|<\epsilon\Big\}\tag{7-154}$$

其中 $p(x^n,y^n)=\prod_{i=1}^{n}p(x_i,y_i)$。

**联合 AEP**：设$(X^n,Y^n)$为 i.i.d. 服从分布 $p(x^n,y^n)=\prod_{i=1}^{n}p(x_i,y_i)$ 且长度为 $n$ 的序列，则

1. $\Pr((X^n,Y^n)\in A_\epsilon^{(n)})\to 1, n\to\infty$。
2. $|A_\epsilon^{(n)}|\leqslant 2^{n(H(X,Y)+\epsilon)}$。
3. 如果$(\tilde{X}^n,\tilde{Y}^n)\sim p(x^n)p(y^n)$，则 $\Pr((\tilde{X}^n,\tilde{Y}^n)\in A_\epsilon^{(n)})\leqslant 2^{-n(I(X;Y)-3\epsilon)}$。

**信道编码定理** 所有小于信道容量 $C$ 的码率都是可达的，而所有大于信道容量的码率是不可达的；也就是说，对任意的码率 $R<C$，存在误差概率满足 $\lambda^{(n)}\to 0$ 的一个$(2^{nR},n)$码序列。反之，如果码率 $R>C$，那么 $\lambda^{(n)}$ 将远离 0。

**反馈容量** 对于离散无记忆信道，反馈并不能增加信道容量，即 $C_{FB}=C$。

**信源信道定理** 如果随机过程的熵率 $H>C$，则该过程不能通过离散无记忆信道被可靠地传输。相反，如果随机过程满足 AEP，且 $H<C$，则信源可以被可靠地传输。

## 习题

7.1 输出的预处理。如果一个统计学家面对具有转移概率为 $p(y|x)$且信道容量 $C=\max_{p(x)}I(X,Y)$的通信信道，他会对输出做出很有帮助的预处理：$\tilde{Y}=g(Y)$，并且断定这样做能够严格地改进容量。

(a) 请证明他错了。

223

(b) 在什么条件下他不会严格地减小容量?