第3章 单符号离散信道

CHAPTER 3

信息论中有关信道问题的讨论，使用编码信道模型，即用信道输入/输出符号转移概率或转移概率密度函数表征信道特性。

如果信道输入/输出的是随机过程，则对应的是波形信道；如果信道的输入/输出是随机向量，而且向量中每个随机变量的取值是连续的，则对应的信道是连续信道，每个随机变量的取值是离散的，则对应的信道是离散信道。离散信道中，若输入/输出分别仅有一个随机变量，则这种信道称为单符号离散信道。

本章将对单符号离散信道的信息传输、信道容量计算等一系列基本问题展开谈论。

3.1 信道的数学模型

单符号离散信道模型如图 3.1 所示。

输入变量 X 有 r 种取值，即输入符号集 $X=\{a_1,a_2,\cdots,a_r\}$，输出变量 Y 有 s 种取值，即输出符号集 $Y=\{b_1,b_2,\cdots,b_s\}$。信道转移概率

$$P(Y=b_j \mid X=a_i)=p(b_j \mid a_i)=p_{ij}$$

共有 $r\times s$ 个取值，体现了信道的符号传递特性。

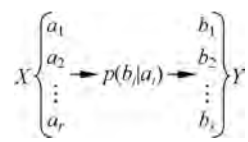

图 3.1 单符号离散信道模型

写成矩阵形式，形成一个 $r\times s$ 矩阵，矩阵行数代表信道输入符号个数，矩阵列数代表输出符号个数。

$$\boldsymbol{P}=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rs} \end{bmatrix}$$

而且满足 $0\leqslant p_{ij}\leqslant 1$，$i$ 是行的标号，j 是列的标号。

其中，$\sum_{j=1}^{s} p(b_j \mid a_i)=1(i=1,2,\cdots,r)$，表明信道矩阵行和是 1。信道特性也可以形象、直观地用信道转移图表示，如图 3.2 所示。

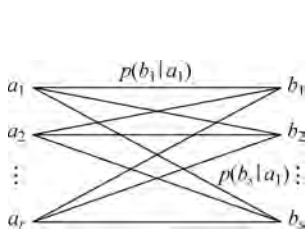

图 3.2 信道转移图

【例 3.1】 二进制对称信道(Binary Symmetric Channel，BSC)，输入/输出符号集分别为 $X=\{0,1\}$，$Y=\{0,1\}$，信道转移概率满足 $P_{Y|X}(0|0)=$

$P_{Y|X}(1|1)=1-p$，$P_{Y|X}(1|0)=P_{Y|X}(0|1)=p$，$p$ 为错误传输概率。写出信道的转移概率矩阵，并画出转移概率图。

解：转移概率矩阵为 $\boldsymbol{P}=\begin{bmatrix}1-p & p \\ p & 1-p\end{bmatrix}$，相应的信道转移概率矩阵如图 3.3 所示。

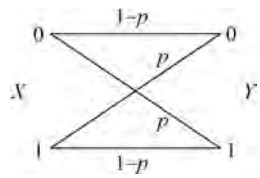

图 3.3　二进制对称信道转移概率图

【例 3.2】　二进制删除信道（Binary Erasure Channel，BEC），输入符号集 $X=\{0,1\}$，输出符号集 $Y=\{0,2,1\}$，转移概率如图 3.4 所示。写出信道的转移概率矩阵。

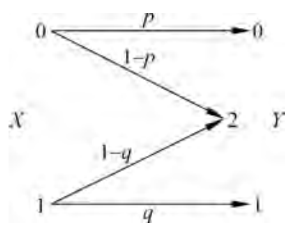

图 3.4　二进制删除信道转移概率图

解：信道的转移概率矩阵为

$$\boldsymbol{P}=\begin{bmatrix}p & 1-p & 0 \\ 0 & 1-q & q\end{bmatrix}$$

3.2　信道的交互信息量

在求解信道输入/输出单个符号对应互信息量前，先讨论信道输入/输出变量的统计特性。

为方便讨论，首先定义几种概率的名称。

1. 先验概率

信源 X 输出符号 a_i 的概率 $P(X=a_i)=p(a_i)$ 称为先验概率。

2. 正向转移概率

从信道输入符号 a_i 到信道输出符号 b_j 的条件概率

$$P(Y=b_j \mid X=a_i)=p(b_j \mid a_i)=p_{ij} \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.1}$$

称为正向转移概率。

3. 后验概率

从信道输出符号 b_j 到输出符号 a_i 的条件概率

$$P(X=a_i \mid Y=b_j)=p(a_i \mid b_j)=p_{ji} \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.2}$$

称为后验概率，又称为反向转移概率。

利用 $P(X|Y)=\dfrac{P(XY)}{P(Y)}$，则有

$$p(a_i \mid b_j)=\frac{p(a_ib_j)}{p(b_j)}=\frac{p(a_i)p(b_j \mid a_i)}{\sum_{i=1}^{r}p(a_i)p(b_j \mid a_i)} \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.3}$$

可以看到，已知先验概率和信道转移概率、后验概率即为确定。

4. 联合概率

信源符号 a_i 和信道输出符号 b_j 的联合概率

$$P(X=a_i;Y=b_j)=p(a_ib_j)=p(a_i)p(b_j \mid a_i) \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.4}$$

由信源概率和信道转移概率唯一确定。

类似信道转移概率矩阵，可以写出联合概率矩阵，也是 r 行 s 列矩阵，即

$$\boldsymbol{P}_{XY}(a_ib_j)=\begin{bmatrix} p(a_1b_1) & p(a_1b_2) & \cdots & p(a_1b_s) \\ p(a_2b_1) & p(a_2b_2) & \cdots & p(a_2b_s) \\ \vdots & \vdots & \ddots & \vdots \\ p(a_rb_1) & p(a_rb_2) & \cdots & p(a_rb_s) \end{bmatrix} \tag{3.5}$$

5. 信宿概率

信道输出符号概率 $P(Y=b_j)=p(b_j)$，$p(b_j)=\sum_{i=1}^{r}p(a_ib_j)=\sum_{i=1}^{r}p(a_i)p(b_j \mid a_i)$。已知信源符号概率和信道转移概率，即可求得信道输出符号概率。现在转入互信息量的讨论。

在通信系统中，信源发出某符号 a_i，由于受到噪声的随机干扰，在信道的输出端输出符号 a_i 的某种变型 b_j，按信息的定义，信宿收到 b_j 后，从 b_j 中获取关于 a_i 的信息量 $I(a_i;b_j)$，等于信宿收到 b_j 前、后，对符号 a_i 的不确定性的消除，即有

[信宿收到 b_j 后，从 b_j 中获取关于 a_i 的信息量] $I(a_i;b_j)$

=[收到 b_j 前，收信者对信源发出 a_i 的不确定性]－

[收到 b_j 后，收信者对信源发出 a_i 仍然存在的不确定性]

=收信者收到 b_j 前、后，对信源发出 a_i 的不确定性的消除　(3.6)

信宿收到 b_j 前，对信源发出符号 a_i 的先验不确定性

$$I(a_i)=\log\frac{1}{p(a_i)} \quad (i=1,2,\cdots,r) \tag{3.7}$$

信宿收到 b_j 后，对信源发出的符号 a_i 的后验不确定性

$$I(a_i \mid b_j)=\log\frac{1}{p(a_i \mid b_j)} \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.8}$$

则可得互信息为

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

$$=\log \frac{p(a_i \mid b_j)}{p(a_i)} \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.9}$$

信宿收到 b_j 后，从 b_j 中获取关于 a_i 的信息量 $I(a_i;b_j)$称为输入符号 a_i 和输出符号 b_j 之间的交互信息量，简称为互信息。它表示信道在把输入符号 a_i 传递为输出符号 b_j 的过程中，信道所传递的信息量。式(3.9)称为符号 a_i 和 b_j 之间的互信息函数。

现在就互信息量表达式所代表的物理含义做进一步说明。

(1) 当 $p(a_i|b_j)=1$ 时，有

$$I(a_i;b_j)=\log \frac{1}{p(a_i)}=I(a_i) \quad (i=1,2,\cdots,r) \tag{3.10}$$

$p(a_i|b_j)=1$ 表明收到 $b_j(j=1,2,\cdots,s)$后，即可确切无误地判断发端符号为 a_i，消除对 a_i 的全部不确定性，接收端获得关于 a_i 的全部信息量 $I(a_i)(i=1,2,\cdots,r)$。

(2) 当 $p(a_i)<p(a_i|b_j)<1$ 时，有

$$I(a_i;b_j)=\log \frac{p(a_i \mid b_j)}{p(a_i)}>0 \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.11}$$

这就意味着，收信者收到 b_j 后，判断信源发出 $a_i(i=1,2,\cdots,r)$的可能性，比对于收到 $b_j(j=1,2,\cdots,s)$前判断信源发出 $a_i(i=1,2,\cdots,r)$的可能性更大；也就是说收到 $b_j(j=1,2,\cdots,s)$后对信源发出 $a_i(i=1,2,\cdots,r)$的不确定性，比收到 $b_j(j=1,2,\cdots,s)$前有所减小，收信者从 $b_j(j=1,2,\cdots,s)$中就可获取关于 $a_i(i=1,2,\cdots,r)$的一定信息量。

(3) 当 $p(a_i|b_j)=p(a_i)$时，有

$$I(a_i;b_j)=\log \frac{p(a_i \mid b_j)}{p(a_i)}=\log 1=0 \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.12}$$

由 $p(a_ib_j)=p(a_i)p(b_j|a_i)=p(a_i)p(b_j)$得到，符号 a_i 与符号 b_j 统计独立。

说明，收信者收到 $b_j(j=1,2,\cdots,s)$前、后，判断信源发出 $a_i(i=1,2,\cdots,r)$的可能性大小没有任何变化，收信者在收到 $b_j(j=1,2,\cdots,s)$前、后，对判断信源发出 $a_i(i=1,2,\cdots,r)$的不确定性没有任何减小，收信者从 $b_j(j=1,2,\cdots,s)$中得不到关于 $a_i(i=1,2,\cdots,r)$的任何信息量。

(4) 当 $0<p(a_i|b_j)<p(a_i)$时，有

$$I(a_i;b_j)=\log \frac{p(a_i \mid b_j)}{p(a_i)}<0 \quad (i=1,2,\cdots,r;\ \cdots j=1,2,\cdots,s) \tag{3.13}$$

由于信道噪声的干扰，收到符号 $b_j(j=1,2,\cdots,s)$后，猜测符号 $a_i(i=1,2,\cdots,r)$的难度反而加大。

互信息量 $I(a_i;b_j)$的另外两种表达形式：

第一种表达形式为

$$\begin{aligned} I(a_i;b_j) &= \log \frac{p(a_i \mid b_j)}{p(a_i)} \\ &= \log \frac{p(a_ib_j)}{p(a_i)p(b_j)} \\ &= \log \frac{1}{p(a_i)p(b_j)}-\log \frac{1}{p(a_ib_j)} \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \end{aligned} \tag{3.14}$$

其含义是：通信前事件 a_i 与 b_j 统计独立，即 $p(a_ib_j)=p(a_i)p(b_j)$，$\log \frac{1}{p(a_i)p(b_j)}$

表达联合事件 a_ib_j 不确定性；通信后，事件 a_i 与事件 b_j 建立统计关联，$\log\dfrac{1}{p(a_ib_j)}$ 表达通信后，联合事件 a_ib_j 的确定性。二者差值同样表明，信宿收到 b_j 后，从 b_j 中获取关于 a_i 的信息量 $I(a_i;b_j)$ 等于通信前后不确定性的消除。

第二种表达形式为

$$
\begin{aligned}
I(a_i;b_j) &= \log\frac{p(a_ib_j)}{p(a_i)p(b_j)} \\
&= \log\frac{p(b_j\mid a_i)}{p(b_j)} \\
&= \log\frac{1}{p(b_j)} - \log\frac{1}{p(b_j\mid a_i)} \\
&= I(b_j;a_i) \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s)
\end{aligned}
\tag{3.15}
$$

其中，$a_i \to b_j$，构成正向信道；$b_j \to a_i$，构成反向信道。

说明在反向信道中，从 a_i 中获得关于 b_j 的信息量等于在正向信道中从 b_j 中获得关于 a_i 的信息量，即 $I(a_i;b_j)=I(b_j;a_i)$。

【例 3.3】 二进制删除信道(BEC)，其中输入符号集 $A=\{a_0,a_1\}=\{0,1\}$，输出符号集 $B=\{b_0,b_1,b_2\}=\{0,1,2\}$，$p(a_0)=p(a_1)=0.5$，写出信道的转移概率矩阵及互信息量 $I(a_i;b_j)$。

解：设信道的转移概率矩阵为

$$
\boldsymbol{P}=\begin{bmatrix} p & 1-p & 0 \\ 0 & 1-q & q \end{bmatrix}
$$

联合概率矩阵为

$$
\boldsymbol{P}(a_ib_j)=\begin{bmatrix} \frac{1}{2}p & \frac{1}{2}(1-p) & 0 \\ 0 & \frac{1}{2}(1-q) & \frac{1}{2}q \end{bmatrix}
$$

则信宿概率分布

$$
\boldsymbol{P}(b_j)=\begin{bmatrix} \frac{1}{2}p & 1-\frac{1}{2}(p+q) & \frac{1}{2}q \end{bmatrix}
$$

那么互信息量 $\boldsymbol{I}(a_i;b_j)$ 用矩阵形式简写为

$$
\begin{aligned}
\boldsymbol{I}(a_i;b_j) &= \begin{bmatrix} \log\dfrac{p(a_ib_j)}{p(a_i)p(b_j)} \end{bmatrix} \\
&= \begin{bmatrix} \log 2 & \log\dfrac{1-p}{1-\frac{1}{2}(p+q)} & \log 0 \\ \log 0 & \log\dfrac{1-q}{1-\frac{1}{2}(p+q)} & \log 2 \end{bmatrix}
\end{aligned}
$$

【例 3.4】 4 个等概率消息，编程的码字为 $M_1=000$，$M_2=011$，$M_3=101$，$M_4=110$，通过如图 3.5 所示二元对称无记忆信道($\varepsilon<0.5$)传输，求：

(1) 事件“接收到第一个数字为 0”与发送 M_1 之间的互信息；

(2) 当接收到第二个数字也为 0 时，关于 M_1 的附加信息；

(3) 当接收到第三个数字也为 0 时，求又增加了多少关于 M_1 的信息。

解：记“0”表示第一个接收数字为 0，“00”表示第一、二个接收数字都为 0，“000”表示前三个接收数字都为 0；$q(\cdot)$表示接收符号的概率；$p(y|x)$为信道的转移概率。

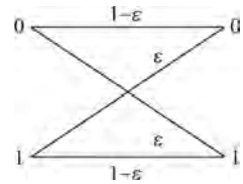

图 3.5　二元对称无记忆信道

(1) $q(\text{“0”}) = \sum_{i=1}^{4} p(M_i)p(0 \mid M_i) = \frac{1}{4}[2(1-\varepsilon)+2\varepsilon] = \frac{1}{2}$

互信息为

$$I(M_1;\text{“0”}) = \log\frac{p(\text{“0”} \mid M_1)}{q(\text{“0”})} = \log\frac{1-\varepsilon}{1/2} = \log[2(1-\varepsilon)]$$

(2) $q(\text{“00”}) = \sum_{i=1}^{4} p(M_i)p(00 \mid M_i)$

$$= \frac{1}{4}[p(0 \mid 0)p(0 \mid 0)+p(0 \mid 0)p(0 \mid 1)+p(0 \mid 1)p(0 \mid 0)+p(0 \mid 1)p(0 \mid 1)]$$

$$= \frac{1}{4}[(1-\varepsilon)^2 + 2\varepsilon(1-\varepsilon) + \varepsilon^2] = \frac{1}{4}$$

互信息为

$$I(M_1;\text{“00”}) = \log\frac{p(\text{“00”} \mid M_1)}{q(\text{“00”})} = \log\frac{(1-\varepsilon)^2}{1/4} = 2\log[2(1-\varepsilon)]$$

附加信息为

$$\log[2(1-\varepsilon)]$$

(3) $q(\text{“000”}) = \sum_{i=1}^{4} p(M_i)p(000 \mid M_i) = \frac{1}{4}[(1-\varepsilon)^3 + 3\varepsilon^2(1-\varepsilon)] = \frac{1}{4}(1-\varepsilon)(4\varepsilon^2 - 2\varepsilon + 1)$

互信息为

$$I(M_1;\text{“000”}) = \log\frac{p(\text{“000”} \mid M_1)}{q(\text{“000”})} = \log\frac{(1-\varepsilon)^3}{(1-\varepsilon)(4\varepsilon^2 - 2\varepsilon + 1)/4}$$

$$= 2\log[2(1-\varepsilon)] - \log(4\varepsilon^2 - 2\varepsilon + 1)$$

又增加的信息为

$$-\log(4\varepsilon^2 - 2\varepsilon + 1)$$

3.3　条件互信息量

现在把目光转到级联信道的交互信息的讨论上。

如图 3.6 所示，信道Ⅰ与信道Ⅱ串接。信道Ⅰ的输入符号集 $X=\{a_1, a_2, \cdots, a_r\}$，输出符号集 $Y=\{b_1, b_2, \cdots, b_s\}$，信道Ⅱ的输入符号集 $Y=\{b_1, b_2, \cdots, b_s\}$，输出符号集 $Z=\{c_1, c_2, \cdots, c_s\}$

信道Ⅰ的传递概率

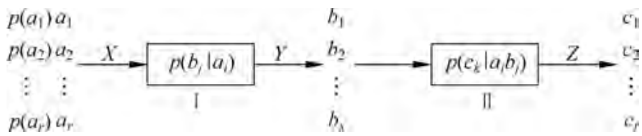

图 3.6 信道Ⅰ与信道Ⅱ串接

$$P(Y \mid X)=\{p(b_j \mid a_i)\} \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s)$$

信道Ⅱ的传递概率

$$P(Z \mid XY)=\{p(c_k \mid a_ib_j)\} \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t)$$

则 $p(a_ib_jc_k)=p(a_i)p(b_j|a_i)p(c_k|a_ib_j)(i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t)$

由 X、Y、Z 的联合概率 $p(a_ib_jc_k)$ 可以求得其他各种概率分布。

一维分布为

$$p(a_i)=\sum_{j=1}^{s}\sum_{k=1}^{t}p(a_ib_jc_k)$$

$$p(b_j)=\sum_{i=1}^{r}\sum_{k=1}^{t}p(a_ib_jc_k)$$

$$p(c_k)=\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_jc_k)$$

二维分布为

$$p(a_ib_j)=\sum_{k=1}^{t}p(a_ib_jc_k)$$

$$p(b_jc_k)=\sum_{i=1}^{r}p(a_ib_jc_k)$$

$$p(a_ic_k)=\sum_{j=1}^{s}p(a_ib_jc_k)$$

进一步求得条件概率分布为

$$p(a_ib_j \mid c_k)=\frac{p(a_ib_jc_k)}{p(c_k)}$$

$$p(a_ic_k \mid b_j)=\frac{p(a_ib_jc_k)}{p(b_j)}$$

$$p(b_jc_k \mid a_i)=\frac{p(a_ib_jc_k)}{p(a_i)}$$

$$p(a_i \mid b_jc_k)=\frac{p(a_ib_jc_k)}{p(b_jc_k)}$$

$$p(b_j \mid a_ib_j)=\frac{p(a_ib_jc_k)}{p(a_ic_k)}$$

$$p(c_k \mid a_ib_j)=\frac{p(a_ib_jc_k)}{p(a_ib_j)}$$

在串接信道输出某符号 c_k 条件下,从符号 b_j 中获取符号 a_i 的信息量定义为

$$I(a_i;b_j\mid c_k)=\log\frac{p(a_i\mid b_jc_k)}{p(a_i\mid c_k)}$$
$$=\log\frac{1}{p(a_i\mid c_k)}-\log\frac{1}{p(a_i\mid b_jc_k)}$$
$$=I(a_i\mid c_k)-I(a_i\mid b_jc_k)(i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t) \tag{3.16}$$

该式表明，c_k 已知条件下，b_j 出现前后对 a_i 的条件不确定性的消除。

用 $p(b_jc_k)$同时乘式(3.16)的右边，则有

$$I(a_i;b_j\mid c_k)=\log\frac{p(b_jc_k)p(a_i\mid b_jc_k)}{p(b_jc_k)p(a_i\mid c_k)}$$
$$=\log\frac{p(a_ib_jc_k)}{p(c_k)p(b_j\mid c_k)p(a_i\mid c_k)}$$
$$=\log\frac{p(a_ib_j\mid c_k)}{p(b_j\mid c_k)p(a_i\mid c_k)}$$
$$=\log\frac{1}{p(a_i\mid c_k)}+\log\frac{1}{p(b_j\mid c_k)}-\log\frac{1}{p(a_ib_j\mid c_k)}$$
$$=I(a_i\mid c_k)+I(b_j\mid c_k)-I(a_ib_j\mid c_k)$$
$$(i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t) \tag{3.17}$$

其中，信道Ⅰ通信前、后，输入/输出端同时出现 a_i 和 b_j 的条件不确定性的消除。

由式(3.17)做进一步变换，有

$$I(a_i;b_j\mid c_k)=\log\frac{p(a_ib_j\mid c_k)}{p(b_j\mid c_k)p(a_i\mid c_k)}$$
$$=\log\left[\frac{p(a_ib_j\mid c_k)}{p(a_i\mid c_k)}\cdot\frac{1}{p(b_j\mid c_k)}\right]$$
$$=\log\frac{p(b_j\mid a_ic_k)}{p(b_j\mid c_k)}=\log\frac{1}{p(b_j\mid c_k)}-\log\frac{1}{p(b_j\mid a_ic_k)}$$
$$=I(b_j\mid c_k)-I(b_j\mid a_ic_k)\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t)$$

所以有

$$I(a_i;b_j\mid c_k)=I(b_j;a_i\mid c_k)$$

在图3.6中，随机变量 c_k 与随机变量 X 和 Y 的联合符号(a_ib_j)之间的相关交互信息量为

$$I(a_ib_j;c_k)=\log\frac{p(c_k\mid a_ib_j)}{p(c_k)}=\log\frac{p(c_k\mid b_j)p(c_k\mid a_ib_j)}{p(c_k\mid b_j)p(c_k)}$$
$$=\log\frac{p(c_k\mid b_j)}{p(c_k)}+\log\frac{p(c_k\mid a_ib_j)}{p(c_k\mid b_j)}$$
$$=I(b_j;c_k)+I(a_i;c_k\mid b_j)\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t)$$

这表明，相关交互信息量 $I(a_ib_j;c_k)$等于 b_j 与 c_k 之间的交互信息量 $I(b_j;c_k)$，再加上 b_j 已知的条件下，a_i 与 c_k 之间的条件交互信息量 $I(a_i;c_k|b_j)$所得之和。同样地

$$I(a_ib_j;c_k)=\log\frac{p(c_k\mid a_ib_j)}{p(c_k)}=\log\frac{p(c_k\mid a_i)p(c_k\mid a_ib_j)}{p(c_k\mid a_i)p(c_k)}$$
$$=\log\frac{p(c_k\mid a_i)}{p(c_k)}+\log\frac{p(c_k\mid a_ib_j)}{p(c_k\mid a_i)}$$

$$=I(a_i;c_k)+I(b_j;c_k\mid a_i)\quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t)$$

这表明，相关交互信息量 $I(a_ib_j;c_k)$ 也等于 a_i 与 c_k 之间的交互信息量 $I(a_i;c_k)$，再加上 a_i 已知的条件下，b_j 与 c_k 之间的条件交互信息量 $I(b_j;c_k|a_i)$ 所得之和。

【例 3.5】 如表 3.1 所示列出了无失真信源编码的信源消息、消息的先验概率以及每一个消息所对应的码字。

表 3.1 信源消息、消息概率、码字

信源消息	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
消息概率	1/4	1/4	1/8	1/8	1/16	1/16	1/16	1/16
码字	000	001	010	011	100	101	110	111

试以消息 a_5 及相应码字 100 为例，分别说明码字 100 中每一个码符号对消息 a_5 提供的信息量。

解：根据相关交互信息量的理论可得

$$\begin{aligned}I(a_5;100)&=\log\frac{p(a_5\mid 100)}{p(a_5)}=\log\frac{p(a_5\mid 100)p(a_5\mid 1)p(a_5\mid 10)}{p(a_5)p(a_5\mid 1)p(a_5\mid 10)}\\&=\log\frac{p(a_5\mid 1)}{p(a_5)}+\log\frac{p(a_5\mid 10)}{p(a_5\mid 1)}+\log\frac{p(a_5\mid 100)}{p(a_5\mid 10)}\\&=I(a_5;1)+I(a_5;0\mid 1)+I(a_5;0\mid 10)\end{aligned}$$

以下分别计算其中各项条件互信息。

(1) $I(a_5;1)=\log\dfrac{p(a_5|1)}{p(a_5)}$

其中，$p(a_5|1)$ 表示收到码符号"1"后，判断信源发消息 a_5 的后验概率。因为收到码符号"1"后，再收到码符号序列"00"就构成码字 100，即消息 a_5 出现，所以后验概率 $p(a_5|1)=p(00|1)$，即有 $p(a_5|1)=p(00|1)=p(100)|p(1)$。

其中，码字 100 出现的概率 $p(100)$ 等于消息 a_5 出现的概率，即有 $p(100)=p(a_5)=1/16$。

从表 3.1 中可看出，8 个码字中有 4 个码字 100、101、110、111 的第一个码符号是"1"，所以有

$$\begin{aligned}p(1)&=p(100)+p(101)+p(110)+p(111)\\&=\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}=\frac{1}{4}\end{aligned}$$

即可得

$$p(a_5\mid 1)=\frac{p(100)}{p(1)}=\frac{\frac{1}{16}}{\frac{1}{4}}=\frac{1}{4}$$

这样就有

$$I(a_5;1)=\log\frac{p(a_5\mid 1)}{p(a_5)}=\log\frac{\frac{1}{4}}{\frac{1}{16}}=2$$

(2) $I(a_5;0|1)=\log\dfrac{p(a_5|10)}{p(a_5|1)}$

其中，$p(a_5|10)$表示收到码符号“10”后，判断信源发消息 a_5 的后验概率。因为收到码符号“10”后，再收到码符号序列“0”就构成码字 100，即消息 a_5 出现，所以后验概率 $p(a_5|10)=p(0|10)$，即有 $p(a_5|10)=p(0|10)=\dfrac{p(100)}{p(10)}$。

从表 3.1 中可看出，8 个码字中有 2 个码字 100、101 的前两个码符号序列是“10”，所以有

$$p(10)=p(100)+p(101)=\frac{1}{16}+\frac{1}{16}=\frac{1}{8}$$

即可得

$$p(a_5\mid 10)=\frac{p(100)}{p(10)}=\frac{\frac{1}{16}}{\frac{1}{8}}=\frac{1}{2}$$

这样就有

$$I(a_5;0\mid 1)=\log\frac{p(a_5\mid 10)}{p(a_5\mid 1)}=\log\frac{\frac{1}{2}}{\frac{1}{4}}=1$$

(3) $I(a_5;0|10)=\log\dfrac{p(a_5|100)}{p(a_5|10)}$

其中，$p(a_5|100)$表示收到码字 100 后，判断信源发消息 a_5 的后验概率。显然收到 100 后也就是收到了消息 a_5，所以有

$$p(a_5\mid 100)=1$$

这样就有

$$I(a_5;0\mid 10)=\log\frac{p(a_5\mid 100)}{p(a_5\mid 10)}=\log\frac{1}{p(a_5\mid 10)}=\log\frac{1}{\frac{1}{2}}=1$$

以上三项计算结果表明，在消息 a_5 相对应的码字 100 中：第一个码符号“1”提供关于消息 a_5 的信息量 $I(1;a_5)=I(a_5;1)=2$；在收到第一个码符号“1”的条件下，第二个码符号“0”提供关于 a_5 的条件交互信息量 $I(0;a_5|1)=I(a_5;0|1)=1$；在收到第一个码符号“1”和第二个码符号“0”组成的码符号序列“10”的条件下，第三个码符号“0”提供关于 a_5 的条件交互信息量 $I(0;a_5|10)=I(a_5;0|10)=1$。所以，从码字 100 中的三个码符号总共提供关于消息 a_5 的相关交互信息量

$$\begin{aligned}I(100;a_5)=I(a_5;100)&=I(a_5;1)+I(a_5;0\mid 1)+I(a_5;0\mid 10)\\&=2+1+1=4\end{aligned}$$

另一方面，消息 a_5 的自信息量

$$I(a_5)=\log\frac{1}{p(a_5)}=\log\frac{1}{\frac{1}{16}}=\log 16=4$$

这从信息测量的角度证实了由于消息 a_5 与相应的码字 100 是一一对应的确定关系，相关交互信息量 $I(a_5;100)$ 就是消息 a_5 的自信息量 $I(a_5)$。

3.4 平均交互信息量

$I(a_i;b_j)$ 表示单个事件间的交互信息量，要求信道两端平均一对符号传递信息量的多少，就要计算平均交互信息量，如图 3.7 和图 3.8 所示。

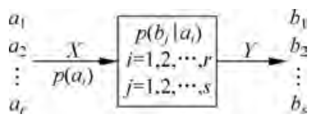

图 3.7 信息传输方向为 X 到 Y

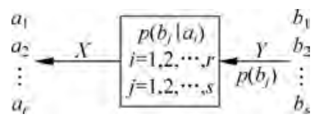

图 3.8 信息传输方向为 Y 到 X

首先给出两种特殊形式，即

$$\begin{aligned} I(X;b_j) &= \sum_{i=1}^{r} p(a_i \mid b_j) I(a_i;b_j) \\ &= \sum_{i=1}^{r} p(a_i \mid b_j) \log \frac{p(a_i \mid b_j)}{p(a_i)} \quad (j=1,2,\cdots,s) \end{aligned} \tag{3.18}$$

这表示信宿收到 b_j 后，获得有关信源的信息量，即

$$\begin{aligned} I(a_i;Y) &= \sum_{j=1}^{s} p(b_j \mid a_i) I(a_i;b_j) \\ &= \sum_{j=1}^{s} p(b_j \mid a_i) \log \frac{p(b_j \mid a_i)}{p(b_j)} \quad (i=1,2,\cdots,r) \end{aligned} \tag{3.19}$$

这表示信宿收到 a_i 后，获得有关信源的信息量。

由上述两个式子可知，X 与 Y 之间平均传递一个符号所传输的平均信息量 $I(X;Y)$ 应该是 $I(a_i;b_j)$ 在联合集 XY 中的统计均值，即

$$I(X;Y) = \sum_{i=1}^{r}\sum_{j=1}^{s} p(a_i b_j) I(a_i;b_j) = I(Y;X) \tag{3.20}$$

同样，$I(X;Y)$ 也有三种不同的表达形式。

(1) 用信源概率 $p(a_i)$ 和正向传输概率 $p(b_j|a_i)$ 表达，即

$$\begin{aligned} I(X;Y) &= \sum_{i=1}^{r}\sum_{j=1}^{s} p(a_i b_j) I(a_i;b_j) = \sum_{i=1}^{r}\sum_{j=1}^{s} p(a_i b_j) \log \frac{p(a_i \mid b_j)}{p(a_i)} \\ &= -\sum_{i=1}^{r}\sum_{j=1}^{s} p(a_i b_j) \log p(a_i) - \left[-\sum_{i=1}^{r}\sum_{j=1}^{s} p(a_i b_j) \log p(a_i \mid b_j)\right] \\ &= -\sum_{i=1}^{r} p(a_i) \log p(a_i) - \sum_{j=1}^{s} p(b_j)\left[-\sum_{i=1}^{r} p(a_i \mid b_j) \log p(a_i \mid b_j)\right] \\ &= H(X) - \sum_{j=1}^{s} p(b_j) H(X \mid Y=b_j) = H(X) - H(X \mid Y) \end{aligned} \tag{3.21}$$

其中，

$$H(X \mid Y=b_j)=-\sum_{i=1}^{r}p(a_i \mid b_j)\log p(a_i \mid b_j) \quad (j=1,2,\cdots,s) \tag{3.22}$$

表示在随机变量 $Y=b_j(j=1,2,\cdots,s)$的前提下，对随机变量 X 仍然存在的平均不确定性。而条件熵

$$\begin{aligned} H(X \mid Y) &= \sum_{j=1}^{s}p(b_j)H(X \mid Y=b_j)=-\sum_{j=1}^{s}p(b_j)\sum_{i=1}^{r}p(a_i \mid b_j)\log p(a_i \mid b_j) \\ &= -\sum_{i=1}^{r}\sum_{j=1}^{s}p(b_j)p(a_i \mid b_j)\log p(a_i \mid b_j) \\ &= -\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)\log p(a_i \mid b_j) \end{aligned} \tag{3.23}$$

表示收到随机变量 Y 后，对随机变量 X 仍然存在的平均不确定性，通常称为疑义度。

式(3.21)表明，从收到 Y 中获取关于 X 的平均交互信息量 $I(X;Y)$，等于收到 Y 前对 X 的平均不确定性 $H(X)$，与收到 Y 后对 X 仍然存在的平均不确定性 $H(X|Y)$之差，即收到 Y 前、后，关于 X 的平均不确定性的消除。

(2) 用信宿概率 $p(b_j)$和反向传输概率 $p(a_i|b_j)$表达，即

$$\begin{aligned} I(X;Y) &= \sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)I(a_i;b_j)=\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)\log\frac{p(b_j \mid a_i)}{p(b_j)} \\ &= -\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)\log p(b_j)-\left[-\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)\log p(b_j \mid a_i)\right] \\ &= -\sum_{j=1}^{s}p(b_j)\log p(b_j)-\sum_{i=1}^{r}p(a_i)\left[-\sum_{j=1}^{s}p(b_j \mid a_i)\log p(b_j \mid a_i)\right] \\ &= H(Y)-\sum_{i=1}^{r}p(a_i)H(Y \mid X=a_i)=H(Y)-H(Y \mid X)=I(Y;X) \end{aligned} \tag{3.24}$$

其中，

$$H(Y \mid X=a_i)=-\sum_{j=1}^{s}p(b_j \mid a_i)\log p(b_j \mid a_i) \quad (i=1,2,\cdots,r) \tag{3.25}$$

表示在随机变量 $X=a_i(i=1,2,\cdots,r)$的前提下，随机变量 Y 仍然存在的平均不确定性。而条件熵为

$$\begin{aligned} H(Y \mid X) &= \sum_{i=1}^{r}p(a_i)H(Y \mid X=a_i)=-\sum_{i=1}^{r}p(a_i)\sum_{j=1}^{s}p(b_j \mid a_i)\log p(b_j \mid a_i) \\ &= -\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_i)p(b_j \mid a_i)\log p(b_j \mid a_i) \\ &= -\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)\log p(b_j \mid a_i) \end{aligned} \tag{3.26}$$

表示在反向信道中，收到随机变量 X 后，对随机变量 Y 仍然存在的平均不确定性。这个“反向疑义度”一般称为噪声熵。

式(3.24)表明，对于反向信道来说，从输出随机变量 X 中，获取关于 Y 的平均交互信息量 $I(Y;X)$，等于信宿收到 X 前，对 Y 的先验不确定性 $H(Y)$，与信宿收到 X 后，对 Y 仍然

存在的后验平均不确定性 $H(Y|X)$之差，即通信前、后，关于 Y 的平均不确定性的消除。

(3) 用信源概率 $p(a_i)$、信宿概率 $p(b_j)$和联合概率 $p(a_ib_j)$表达，即

$$
\begin{aligned}
I(X;Y) &= \sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)I(a_i;b_j) = \sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log\frac{p(a_ib_j)}{p(a_i)p(b_j)} \\
&= -\sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log p(a_i) - \sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log p(b_j) - \left[-\sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log p(a_ib_j)\right] \\
&= -\sum_{i=1}^{r} p(a_i)\log p(a_i) - \sum_{j=1}^{s} p(b_j)\log p(b_j) - \left[-\sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log p(a_ib_j)\right] \\
&= H(X) + H(Y) - H(XY) \qquad (3.27)
\end{aligned}
$$

其中，

$$
H(XY) = -\sum_{i=1}^{r}\sum_{j=1}^{s} p(a_ib_j)\log p(a_ib_j)
$$

表示通信后，信道两端同时出现 X 和 Y 的后验平均不确定性，通常称为共熵。

式(3.27)表明，信源 X 通过传递概率为 $P(Y|X)$的信道输出随机变量 Y，信道传递的平均交互信息量 $I(Y;X)$，等于通信前(随机变量 X 和 Y 统计独立)随机变量 X 和 Y 同时出现的平均不确定性$\{H(X)+H(Y)\}$，与通信后(随机变量 X 和 Y 由信道传递概率 $P(Y|X)$相联系)信道两端同时出现随机变量 X 和 Y 的平均不确定性 $H(XY)$之差，即通信前、后，随机变量 X 和 Y 同时出现的平均不确定性的消除。

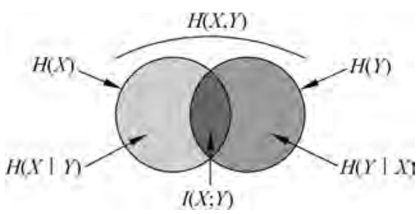

图 3.9 通信系统中各类熵关系图

上述讨论的通信系统中各类熵的关系可用维拉图形象表示，如图 3.9 所示。

【例 3.6】 设信源 X 的符号集 $X=\{a_1,a_2\}$，先验概率分布为 $p(a_1)=\omega(0<\omega<1)$；$p(a_2)=1-\omega$。信道的输入符号集 $X=\{a_1,a_2\}$，输出符号集 $Y=\{a_1,a_2\}$，传递概率 $P(Y|X)=\{p(a_j|a_i)=p_{ij}(i=1,2;\ j=1,2)\}$。现将信源 X 与图 3.10 信道相接。

(1) 试写出平均交互信息量 $I(X;Y)$的一般表达式；

(2) 若 $\omega=\frac{1}{2}$，$p_{11}=p_{22}=\bar{p}$；$p_{12}=p_{21}=p(0\leqslant\bar{p},p\leqslant1;\ \bar{p}+p=1)$。试计算如图 3.11 所示反向信道的 $I(X;Y)$。

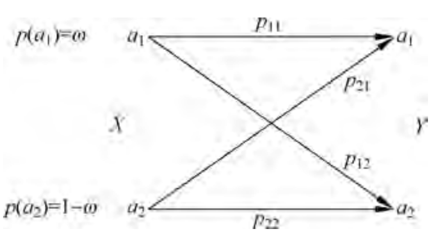

图 3.10 正向信道传递概率图

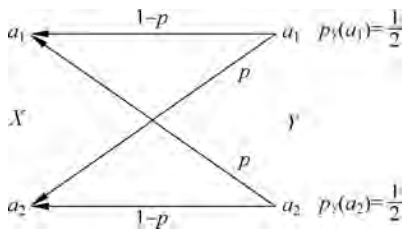

图 3.11 反向信道传递概率图

解：(1) 各联合概率为

$$
p(a_1a_1)=p(a_1)p(a_1\mid a_1)=\omega p_{11}
$$

$$p(a_1a_2)=p(a_1)p(a_2\mid a_1)=\omega p_{12}$$
$$p(a_2a_1)=p(a_2)p(a_1\mid a_2)=(1-\omega)p_{21}$$
$$p(a_2a_2)=p(a_2)p(a_2\mid a_2)=(1-\omega)p_{22}$$

随机变量 Y 的概率分布为

$$p_Y(a_1)=p(a_1a_1)+p(a_2a_1)=\omega p_{11}+(1-\omega)p_{21}$$
$$p_Y(a_2)=p(a_1a_2)+p(a_2a_2)=\omega p_{12}+(1-\omega)p_{22}$$

求随机变量 Y 的熵,则

$$\begin{aligned}H(Y)&=-\sum_{j=1}^{2}p_Y(a_j)\log p_Y(a_j)=-p_Y(a_1)\log p_Y(a_1)-p_Y(a_2)\log p_Y(a_2)\\&=-\{[\omega p_{11}+(1-\omega)p_{21}]\log[\omega p_{11}+(1-\omega)p_{21}]\}-\{[\omega p_{12}+(1-\omega)p_{22}]\cdot\\&\quad\log[\omega p_{12}+(1-\omega)p_{22}]\}\end{aligned}$$

由 $p(a_j|a_i)=p_{ij}(i=1,2;\ j=1,2)$求得条件熵,即

$$\begin{aligned}H(Y\mid X)&=-\sum_{i=1}^{2}\sum_{j=1}^{2}p(a_ia_j)\log p(a_j\mid a_i)\\&=-[p(a_1a_1)\log p(a_1\mid a_1)+p(a_1a_2)\log p(a_2\mid a_1)+p(a_2a_1)\log p(a_1\mid a_2)+\\&\quad p(a_2a_2)\log p(a_2\mid a_2)]\\&=-[\omega p_{11}\log p_{11}+\omega p_{12}\log p_{12}+(1-\omega)p_{21}\log p_{21}+(1-\omega)p_{22}\log p_{22}]\end{aligned}$$

求得平均交互信息量,则

$$\begin{aligned}I(X;Y)&=H(Y)-H(Y\mid X)\\&=-\{[\omega p_{11}+(1-\omega)p_{21}]\log[\omega p_{11}+(1-\omega)p_{21}]\}-\{[\omega p_{12}+(1-\omega)p_{22}]\cdot\\&\quad\log[\omega p_{12}+(1-\omega)p_{22}]\}+\{\omega p_{11}\log p_{11}+\omega p_{12}\log p_{12}+(1-\omega)p_{21}\\&\quad\log p_{21}+(1-\omega)p_{22}\log p_{22}\}\end{aligned}$$

这说明平均交互信息量是信源概率分布 ω 和信道传递概率 $p_{ij}(i=1,2;\ j=1,2)$的函数。

(2) 各联合概率为

$$p(a_1a_1)=\frac{1}{2}\bar{p}$$
$$p(a_1a_2)=\frac{1}{2}p$$
$$p(a_2a_1)=\frac{1}{2}p$$
$$p(a_2a_2)=\frac{1}{2}\bar{p}$$

随机变量 Y 的概率分布为

$$p_Y(a_1)=\frac{1}{2}\bar{p}+\frac{1}{2}p=\frac{1}{2}$$
$$p_Y(a_2)=\frac{1}{2}p+\frac{1}{2}\bar{p}=\frac{1}{2}$$

求随机变量 Y 的熵,则

$$H(Y)=-\left(\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\right)=1$$

由 $p(a_j|a_i)=p_{ij}(i=1,2;\ j=1,2)$，求得噪声熵为

$$H(Y\mid X)=-\left(\frac{1}{2}\bar{p}\log\bar{p}+\frac{1}{2}p\log p+\frac{1}{2}p\log p+\frac{1}{2}\bar{p}\log\bar{p}\right)$$
$$=-(\bar{p}\log\bar{p}+p\log p)=H(\bar{p},p)$$

求得平均交互信息量为

$$I(X;Y)=H(Y)-H(Y\mid X)=1-H(\bar{p},p)$$

3.5 平均交互信息量的性质

平均交互信息量 $I(X;Y)$除具有对称性以外，即 $I(X;Y)=I(Y;X)$，还具有以下基本性质。

1. 平均互信息的非负性

$$I(X;Y)\geqslant 0$$

当且仅当 X 和 Y 统计独立时，等式成立。

【证明】 利用 Jensen 不等式得

$$-I(X;b_j)=\sum_{i=1}^{r}p(a_i\mid b_j)\log\frac{p(a_i)}{p(a_i\mid b_j)}\leqslant\log\left[\sum_{i=1}^{r}p(a_i\mid b_j)\frac{p(a_i)}{p(a_i\mid b_j)}\right]$$
$$=\log\left[\sum_{i=1}^{r}p(a_i)\right]=\log 1=0$$

即有

$$I(X;b_j)\geqslant 0\quad(j=1,2,\cdots,s)$$
$$I(X;Y)=\sum_{j=1}^{s}p(b_j)I(X;b_j)\geqslant 0$$

当且仅当对一切 i,j 都有

$$p(a_ib_j)=p(a_i)p(b_j)\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s)$$

即当 X 和 Y 统计独立时，$I(X;Y)=0$。

2. 平均互信息的极值性

由上述性质，直接得到

$$I(X;Y)\leqslant H(X)$$
$$I(X;Y)\leqslant H(Y)$$

【证明】 由于 $\log\dfrac{1}{p(a_i|b_j)}\geqslant 0$，而 $H(X|Y)$是对 $\log\dfrac{1}{p(a_i|b_j)}$求统计平均，即

$$H(X\mid Y)=\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)\log\frac{1}{p(a_i\mid b_j)}$$

因此有

$$H(X\mid Y)\geqslant 0$$

同理

$$I(Y \mid X) \geqslant 0$$

所以

$$I(X;Y) = H(X) - H(X \mid Y) \leqslant H(X)$$
$$I(X;Y) = H(Y) - H(Y \mid X) \leqslant H(Y)$$

即

$$I(X;Y) \leqslant \min\{H(X), H(Y)\}$$

3. 平均互信息的凸函数性

定理 3.1 信道两端随机变量 X 和 Y 之间的平均互信息量 $I(X;Y)$，在信道转移概率 $p(b_j|a_i)$给定条件下，是输入随机变量 X 的概率分布 $p(X)=\{p(a_i), i=1,2,\cdots,r\}$的$\cap$型凸函数。

【证明】 当条件概率 $p(y|x)$是固定时，平均互信息 $I(X;Y)$只是 $p(x)$的函数，简写成 $I[p(x)]$。现选择输入信源 X 的两种已知的概率分布 $p_1(x)$和 $p_2(x)$。其对应的联合概率分布为 $p_1(xy)=p_1(x)p(y|x)$和 $p_2(xy)=p_2(x)p(y|x)$，因而平均互信息分别为 $I[p_1(x)]$和 $I[p_2(x)]$。再选择输入变量 X 的另一种概率分布 $P(x)$，令 $0<\theta<1, \theta+\bar{\theta}=1$，而 $p(x)=\theta p_1(x)+\bar{\theta}p_2(x)$，因而得其相应的平均互信息为 $I[p(x)]$。

根据平均互信息的定义得

$$\theta I[p_1(x)] + \bar{\theta} I[p_2(x)] - I[p(x)]$$
$$= \sum_{X,Y} \theta p_1(xy) \log \frac{p(y \mid x)}{p_1(y)} + \sum_{X,Y} \bar{\theta} p_2(xy) \log \frac{p(y \mid x)}{p_2(y)} - \sum_{X,Y} p(xy) \log \frac{p(y \mid x)}{p(y)}$$
$$= \sum_{X,Y} \theta p_1(xy) \log \frac{p(y \mid x)}{p_1(y)} + \sum_{X,Y} \bar{\theta} p_2(xy) \log \frac{p(y \mid x)}{p_2(y)} - \sum_{X,Y} [\theta p_1(xy) + \bar{\theta} p_2(xy)] \log \frac{p(y \mid x)}{p(y)}$$

根据概率关系，得

$$p(xy) = p(x)p(y \mid x) = \theta p_1(x) p(y \mid x) + \bar{\theta} p_2(x) P(y \mid x) = \theta p_1(xy) + \bar{\theta} p_2(xy)$$

得

$$\theta I[p_1(x)] + \bar{\theta} I[p_2(x)] - I[p(x)]$$
$$= \theta \sum_{X,Y} p_1(xy) \log \frac{p(y)}{p_1(y)} + \bar{\theta} \sum_{X,Y} p_2(xy) \log \frac{p(y)}{p_2(y)}$$

因为 $\log x$ 是 x 的$\cap$型函数，所以对上式中第一项，根据 Jensen 不等式得

$$\sum_{X,Y} p_1(xy) \log \frac{p(y)}{p_1(y)} \leqslant \log \sum_{X,Y} p_1(xy) \log \frac{p(y)}{p_1(y)}$$
$$= \log \sum_{Y} \frac{p(y)}{p_1(y)} \sum_{X} p_1(xy) = \log \sum_{Y} \frac{p(y)}{p_1(y)} p_1(y) = \log \sum_{Y} p_1(y) = 0$$

同理，有

$$\sum_{X,Y} p_2(xy) \log \frac{p(y)}{p_2(y)} \leqslant 0$$

又因为 θ 和 $\bar{\theta}$ 都小于 1 且大于 0，即得 $\theta I[p_1(x)]+\bar{\theta} I[p_2(x)]-I[p(x)]\leqslant 0$，因而

$$I[\theta p_1(x) + \bar{\theta} p_2(x)] \geqslant \theta I[p_1(x)] + \bar{\theta} I[p_2(x)]$$

因此根据凸函数的定义知，$I(X;Y)$是概率分布 $p(x)$的$\cap$型凸函数。

定理 3.2 信道两端随机变量 X 和 Y 之间的平均交互信息量 $I(X;Y)$，在信源概率分布 $p(a_i)$ 给定的条件下，是信道转移概率 $p(Y|X)$：$\{p(b_j|a_i), i=1,2,\cdots,r;\ j=1,2,\cdots,s\}$ 的 $\cup$ 型凸函数。

【证明】 当概率分布 $p(x)$ 固定时，平均互信息 $I(X;Y)$ 只是条件概率 $p(y|x)$ 的函数，简写成 $I[p(y|x)]$。选择两种条件概率分别为 $p_1(y|x)$ 和 $p_2(y|x)$。相对应的平均互信息分别为 $I[p_1(y|x)]$ 和 $I[p_2(y|x)]$，再选择第三种条件概率满足 $p(y|x)=\theta p_1(y|x)+\bar{\theta} p_2(y|x)$。设相应的平均互信息为 $I[p(y|x)]$，其中 $0<\theta<1, \theta+\bar{\theta}=1$。因而求得

$$I[p(y \mid x)]-\theta I[p_1(y \mid x)]-\bar{\theta} I[p_2(y \mid x)]$$

$$=\sum_{X,Y}[\theta p_1(xy)+\bar{\theta} p_1(xy)]\log\frac{p(x \mid y)}{p(x)}-\sum_{X,Y}\theta p_1(xy)\log\frac{p_1(x \mid y)}{p(x)}-\sum_{X,Y}\bar{\theta} p_2(xy)$$

$$\log\frac{p_2(x \mid y)}{p(x)}=\theta\sum_{X,Y}p_1(xy)\log\frac{p(x \mid y)}{p_1(x \mid y)}+\sum_{X,Y}\bar{\theta} p_2(xy)\log\frac{p(x \mid y)}{p_2(x \mid y)}$$

运用 Jensen 不等式，上式中第一项为

$$\sum_{X,Y}p_1(xy)\log\frac{p(x \mid y)}{p_1(x \mid y)}\leqslant\theta\log\left[\sum_{X,Y}p_1(xy)\frac{p(x \mid y)}{p_1(x \mid y)}\right]=\theta\log\left[\sum_{X,Y}p_1(y)p(x \mid y)\right]$$

$$=\theta\log\left[\sum_{Y}p_1(y)\sum_{X}p(x \mid y)\right]=\theta\log\sum_{Y}p_1(y)=\theta\log 1=0$$

同理

$$\sum_{X,Y}\theta p_2(xy)\log\frac{p(x \mid y)}{p_2(x \mid y)}\leqslant 0$$

所以

$$I[p(y \mid x)]-\theta I[p_1(y \mid x)]-\bar{\theta} I[p_2(y \mid x)]\leqslant 0$$

即

$$I[\theta p_1(y \mid x)+\theta p_2(y \mid x)]\leqslant\theta I[p_1(y \mid x)]+\bar{\theta} I[p_2(y \mid x)]$$

根据凸函数的定义得：平均互信息 $I(X;Y)$ 是条件概率 $p(Y|X)$ 的 $\cup$ 型凸函数。

3.6 信道容量及其一般算法

3.6.1 信道容量的定义

信道的信息传输率定义为信道中平均每个符号所传送的信息量，即平均互信息。

$$R=I(X;Y)=H(X)-H(X \mid Y)$$

信道的信息传输速率定义为信道平均每秒传输的信息量。若传输一个符号平均需要 t s，则信道的信息传输速率表示为

$$R_t=\frac{1}{t}I(X;Y)=\frac{1}{T_s}[H(X)-H(X \mid Y)]$$

给定某个信道，平均交互信息量 $I(X;Y)$ 是信源概率分布 $p(x)$ 的 $\cap$ 型凸函数，存在极大值，这个极大值就定义为信道容量。

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

信道的最大信息传输速率是信道容量的另一种表述形式为

$$C_t = R_{t\max} = \frac{1}{t}\max\{I(X;Y)\}$$

3.6.2 信道容量的一般算法

平均互信息量 $I(X;Y)$ 是输入信源概率分布 $p(x)$ 的∩型凸函数，所以极大值是一定存在的。而 $I(X;Y)$ 是 r 个输入信号变量 $\{p(a_1),p(a_2),\cdots,p(a_r)\}$ 的多元函数，并且任何信源概率分布都必须遵循约束条件

$$\sum_{i=1}^{r} p(a_i) = 1$$

所以求信道容量 C 就是在约束条件式的约束下，求 $I(X;Y)$ 的最大值问题，并导出取最大值时的条件 $p(a_i)(i=1,2,\cdots,r)$。

此类问题可以通过拉格朗日乘子法来计算。为此，作辅助函数

$$F[p(a_1),p(a_2),\cdots,p(a_r)] = I(X;Y) - \lambda\sum_X p(a_i)$$

其中，λ 为拉格朗日乘子。

$$\frac{\partial F}{\partial p(a_i)} = \frac{\partial\left[I(X;Y) - \lambda\sum_X p(a_i)\right]}{\partial p(a_i)} = 0 \tag{3.28}$$

时求得 $I(X;Y)$ 的值即为信道容量。

由于

$$\begin{aligned} I(X;Y) &= \sum_{i=1}^{r}\sum_{j=1}^{s} p(a_i)p(b_j \mid a_i)\log\frac{p(b_j \mid a_i)}{p(b_j)} \\ &= \sum_{i=1}^{r}\sum_{j=1}^{s} p(a_i)p(b_j \mid a_i)[\log p(b_j \mid a_i) - \log p(b_j)] \end{aligned}$$

而 $p(b_j) = \sum_{i=1}^{r} p(a_i)p(b_j/a_i)$

所以

$$\frac{\partial F}{\partial p(a_i)}\log p(b_j) = \left[\frac{\partial \ln p(b_j)}{\partial p(a_i)}\right]\log e = \frac{p(b_j \mid a_i)}{p(b_j)}\log e$$

对式(3.28)整理得

$$\begin{aligned} \frac{\partial F}{\partial p(a_i)} &= \frac{\partial\left[I(X;Y) - \lambda\sum_X p(a_i)\right]}{\partial p(a_i)} \\ &= \sum_{j=1}^{s} p(b_j \mid a_i)\log\frac{p(b_j \mid a_i)}{p(b_j)} - \sum_{k=1}^{r}\sum_{j=1}^{s} p(a_k)p(b_j \mid a_k)\frac{p(b_j \mid a_i)}{p(b_j)}\log e - \lambda \end{aligned} \tag{3.29}$$

而

$$\begin{aligned} \sum_{k=1}^{r}\sum_{j=1}^{s} p(a_k)p(b_j \mid a_k)\frac{p(b_j \mid a_i)}{p(b_j)} &= \sum_{k=1}^{r}\sum_{j=1}^{s} p(a_k b_j)\frac{p(b_j \mid a_i)}{p(b_j)} \\ &= \sum_{j=1}^{s} p(b_j)\frac{p(b_j \mid a_i)}{p(b_j)} = \sum_{j=1}^{s} p(b_j \mid a_i) = 1 \end{aligned}$$

因此,式(3.29)可以简化为

$$\frac{\partial F}{\partial p(a_i)}=\sum_{j=1}^{s}p(b_j\mid a_i)\log\frac{p(b_j\mid a_i)}{p(b_j)}-\log e-\lambda$$

令$\frac{\partial F}{\partial p(a_i)}=0$,得

$$\sum_{j=1}^{s}p(b_j\mid a_i)\log\frac{p(b_j\mid a_i)}{p(b_j)}=\lambda+\log e\quad(i=1,2,\cdots,r)\tag{3.30}$$

式(3.30)两边分别乘以 $p(a_i)$,并求和得

$$\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_i)p(b_j\mid a_i)\log\frac{p(b_j\mid a_i)}{p(b_j)}=\lambda+\log e\tag{3.31}$$

式(3.31)左边即为平均互信息的极大值 C,即

$$C=\lambda+\log e\tag{3.32}$$

结合式(3.32),把式(3.30)中前 r 个方程改写成

$$\sum_{j=1}^{s}p(b_j\mid a_i)\log p(b_j\mid a_i)-\sum_{j=1}^{s}p(b_j\mid a_i)\log p(b_j)=C\quad(i=1,2,\cdots,r)$$

移项后得

$$\sum_{j=1}^{s}p(b_j\mid a_i)\log p(b_j\mid a_i)=\sum_{j=1}^{s}p(b_j\mid a_i)[\log p(b_j)+C]\quad(i=1,2,\cdots,r)$$

令

$$\beta_j=C+\log p(b_j)$$

得

$$\sum_{j=1}^{s}p(b_j\mid a_i)\log p(b_j\mid a_i)=\sum_{j=1}^{s}p(b_j\mid a_i)\beta_j\quad(i=1,2,\cdots,r)$$

这是含有 s 个未知参数 β_j,有 r 个方程的非齐次线性方程组。

如果设 $r=s$,信道传递矩阵 $\boldsymbol{P}$ 是非奇异矩阵,则此方程组有解,并且可以求出 β_j 的数值,然后根据 $\sum_{j=1}^{s}p(b_j)=1$ 的附加条件求得信道容量

$$C=\log\sum_j 2^{\beta_j}$$

由这个 C 值就可以解得对应的输出概率分布为

$$p(b_j)=e^{\beta_j-C}$$

再根据 $p(b_j)=\sum_{i=1}^{r}p(a_i)p(b_j\mid a_i)(j=1,2,\cdots,s)$,即可解出最佳输入分布 $p(a_i)$。

观察式(3.30)可以发现,该式左边正好是输出端接收到符号 Y 后,获得的关于输入符号 x_i 的信息量,结合式(3.32)可知

$$I(x_i;Y)=\sum_{j=1}^{s}p(b_j\mid a_i)\log\frac{p(b_j\mid a_i)}{p(b_j)}=C$$

由此可以导出以下定理。

定理 3.3 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是输入概率分布 $\{p_i\}$ 满足

$$\begin{cases} I(x_i;Y)=C & (p_i\neq 0) \\ I(x_i;Y)\leqslant C & (p_i=0) \end{cases}$$

这时的 $I(X;Y)$ 就是信道容量 C。

如果求解达到信道容量时，最佳概率分布中，某些 $p(a_i)<0$，则这些解无效。

它表明所求极大值 C 出现的区域不满足概率条件，这时最大值必须在边界上，即某些 a_i 的概率 $p(a_i)=0$。因此，必须设某些信源符号 a_i 的概率 $p(a_i)=0$，然后重新进行运算。当 $r<s$ 时，求解非齐次线性方程就比较困难，即使求出解，也无法保证求得的信源符号概率都大于或等于零。因此，必须反复进行运算，这就使运算变得非常复杂。

【例 3.7】 设离散信道的输入符号集为 $\{a_1,a_2,a_3,a_4\}$，输出符号集为 $\{b_1,b_2,b_3,b_4\}$，其信道转移概率矩阵为 $\boldsymbol{P}(y|x)=\begin{bmatrix} 0.25 & 0.5 & 0 & 0.25 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0.125 & 0.125 & 0.25 & 0.5 \end{bmatrix}$，求其信道容量及最佳输入分布。

解：列得

$$\begin{cases} 0.25\beta_1+0.5\beta_2+0.25\beta_4=0.25\log 0.25+0.5\log 0.5+0.25\log 0.25 \\ 0.5\beta_1+0.5\beta_3=0.5\log 0.5+0.5\log 0.5 \\ 0.5\beta_2+0.5\beta_3=0.5\log 0.5+0.5\log 0.5 \\ 0.125\beta_1+0.125\beta_2+0.25\beta_3+0.5\beta_4=-(0.125\times 3+0.125\times 3+0.25\times 2+0.5\times 1) \end{cases}$$

解得

$$\beta_1=\beta_2=-\frac{7}{6},\quad \beta_3=-\frac{5}{6},\quad \beta_4=-\frac{5}{2}$$

则信道的信道容量

$$C=\log\sum_j 2^{\beta_j}=0.7039$$

由 $p(b_j)=\mathrm{e}^{(\beta_j-C)}$ 得

$$p(b_1)=p(b_2)=0.2735,\quad p(b_3)=0.3445,\quad p(b_4)=0.1085$$

由于 $p(b_j)=\sum_{i=1}^{r}p(a_i)p(b_j\mid a_i)(j=1,2,\cdots,s)$ 可写成为

$$[p(a_1),p(a_2),p(a_3),p(a_4)]\boldsymbol{P}(y\mid x)=[p(b_1),p(b_2),p(b_3),p(b_4)]$$

因此

$$[p(a_1),p(a_2),p(a_3),p(a_4)]$$

$$=[0.2735\quad 0.2735\quad 0.3445\quad 0.1085]\begin{bmatrix} 0.25 & 0.5 & 0 & 0.25 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0.5 & 0 \\ 0.125 & 0.125 & 0.25 & 0.5 \end{bmatrix}^{-1}$$

$$=[0.2698\quad 0.3915\quad 0.2566\quad 0.0821]$$

因此,最佳分布为

$$p(a_1)=0.2698,\quad p(a_2)=0.3915,\quad p(a_3)=0.2566,\quad p(a_4)=0.0821$$

现在验证以上结果

$$\begin{aligned}I(x=a_1;Y)&=\sum_{j=1}^{4}p(b_j\mid a_1)\log\frac{p(b_j\mid a_1)}{p(b_j)}\\&=\frac{1}{4}\log\frac{1/4}{0.2735}+\frac{1}{2}\log\frac{1/2}{0.2735}+\frac{1}{4}\log\frac{1/4}{0.1085}\\&=0.7039\end{aligned}$$

同理可以计算

$$I(x=a_2;Y)=\sum_{j=1}^{4}p(b_j\mid a_2)\log\frac{p(b_j\mid a_2)}{p(b_j)}=0.7039$$

$$I(x=a_3;Y)=\sum_{j=1}^{4}p(b_j\mid a_3)\log\frac{p(b_j\mid a_3)}{p(b_j)}=0.7039$$

$$I(x=a_4;Y)=\sum_{j=1}^{4}p(b_j\mid a_4)\log\frac{p(b_j\mid a_4)}{p(b_j)}=0.7039$$

显然

$$\begin{cases}I(x_i;Y)=0.7039 & (p_i\neq 0)\\ I(x_i;Y)\leqslant C & (p_i=0)\end{cases}$$

而每个符号贡献的互信息也正好是前文求解出的信道容量,证实了该求解过程是正确的。

3.7 几种特殊结构的信道容量计算

3.7.1 无噪无损信道

无噪无损信道如图 3.12 所示。

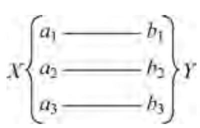

图 3.12 无噪无损信道

输入/输出符号一一对应,信道转移概率矩阵为单位矩阵,即

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

此时平均互信息满足

$$I(X;Y)=H(X)=H(Y)$$

噪声熵 $H(Y|X)$和损失熵 $H(X|Y)$都为 0。

此时输入符号的概率分布为等概分布。所以无噪无损信道的信道容量为

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

3.7.2 有噪无损信道

此时噪声熵 $H(Y|X)\neq 0$，损失熵为0，对应信道的一个输入符号 a_i 有多个输出符号 b_j 与之对应，如图3.13所示。

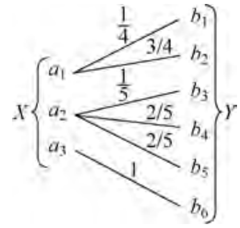

图3.13 有噪无损信道

信道转移概率矩阵

$$P=\begin{bmatrix}\frac{1}{4} & \frac{3}{4} & 0 & 0 & 0 & 0\\ 0 & 0 & \frac{1}{5} & \frac{2}{5} & \frac{2}{5} & 0\\ 0 & 0 & 0 & 0 & 0 & 1\end{bmatrix}$$

信道反向转移概率 $p(a_i|b_j)=1$，$I(X;Y)=H(X)<H(Y)$，信道容量

$$C=\max_{p(x)}\{H(X)\}=\log r$$

输入符号的概率分布为等概分布时达到信道容量。

3.7.3 无噪有损信道

属于多个输入对应一个输出的信道如图3.14所示。

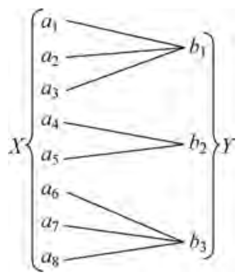

图3.14 无噪有损信道

顾名思义，这种信道噪声熵 $H(Y|X)=0$，损失熵 $H(X|Y)\neq 0$，即

$$I(X;Y)=H(Y)<H(X)$$

信道容量为

$$C=\max_{p(x)}\{H(Y)\}=\log s$$

一定存在一种输入符号分布，使得输出符号 $p(b_j)=\frac{1}{s}$，达到等概分布。

3.7.4 对称离散信道的信道容量

若信道转移概率矩阵每行元素构成相同，称为输入对称；若转移矩阵中，每列元素构成相同，称为输出对称。若输入和输出都对称，此时信道称为对称信道。例如

$$P_1=\begin{bmatrix}\frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}\\ \frac{1}{3} & \frac{1}{6} & \frac{1}{6} & \frac{1}{3}\end{bmatrix},\quad P_2=\begin{bmatrix}\frac{1}{2} & \frac{1}{3} & \frac{1}{6}\\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2}\\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3}\end{bmatrix}$$

都是对称信道矩阵。

若输入符号和输出符号个数相同，都等于 r，那么信道矩阵为

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

其中，$0\leqslant p,\bar{p}\leqslant 1$，且 $p+\bar{p}=1$，则此信道称为强对称信道或均匀信道。这类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。它是对称离散信道的一类特例。二元对称信道就是 $r=2$ 的均匀信道。对于均匀信道，其信道矩阵中各列之和也等于 1(一般信道的信道矩阵中各列之和不一定等于 1)。

对于对称信道，噪声熵 $H(Y|X)$ 为

$$\begin{aligned} H(Y\mid X) &= -\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_i)p(b_j\mid a_i)\log p(b_j\mid a_i) \\ &= \sum_{i=1}^{r}p(a_i)\left[-\sum_{j=1}^{s}p(b_j\mid a_i)\log p(b_j\mid a_i)\right] \\ &= \sum_{i=1}^{r}p(a_i)H(p_1',p_2',\cdots,p_s') = H(p_1',p_2',\cdots,p_s') \end{aligned}$$

可见，离散对称信道的噪声熵就是矩阵某一行元素所对应的熵 $H(p_1',p_2',\cdots,p_s')$。信道容量为

$$\begin{aligned} C &= \max_{p(x)}\{I(X;Y)\} = \max_{p(x)}\{H(Y)-H(Y\mid X)\} \\ &= \max_{p(x)}\{H(Y)\} - H(p_1',p_2',\cdots,p_s') \\ &= \log s - H(p_1',p_2',\cdots,p_s') \\ &= \log s - H(\boldsymbol{P}\text{ 的行向量}) \end{aligned} \tag{3.33}$$

当信道输入符号等概，即 $p(a_i)=\dfrac{1}{r}$ 时，信道输出符号概率

$$\begin{aligned} p(b_j) &= \sum_{i=1}^{r}p(a_i)p(b_j\mid a_i) \\ &= \frac{1}{r}\sum_{i=1}^{r}p(b_j\mid a_i) = \text{常数} \quad (j=1,2,\cdots,s) \end{aligned}$$

其中，$\sum_{i=1}^{r}p(b_j\mid a_i)$ 是对称信道转移概率中的列元素之和，是一个固定值。所以当信道输入符号等概时，输出符号也等概，达到式(3.33)所对应的信道容量，类似可以得到强对称信道的信道容量，即

$$\begin{aligned} C &= \log r - H\left(\bar{p},\frac{p}{r-1},\frac{p}{r-1},\cdots,\frac{p}{r-1}\right) \\ &= \log r + \bar{p}\log\bar{p} + \underbrace{\frac{p}{r-1}\log\frac{p}{r-1}+\cdots+\frac{p}{r-1}\log\frac{p}{r-1}}_{\text{共 } r-1 \text{ 项}} \end{aligned}$$

$$=\log r+\bar{p}\log\bar{p}+p\log\frac{p}{r-1}=\log r-p\log(r-1)-H(p)$$

3.7.5 准对称信道的信道容量

若信道转移概率矩阵中,每行都是同一行元素的不同排列,每列元素构成不同,但该信道矩阵按列可以划分为互不相交的子矩阵,每个子矩阵都是对称矩阵,则该信道称为准对称信道。例如,信道矩阵 $\boldsymbol{P}=\begin{bmatrix}0.7 & 0.1 & 0.2\\0.2 & 0.1 & 0.7\end{bmatrix}$ 可以划分成两个对称的子矩阵 $\boldsymbol{P}_1=\begin{bmatrix}0.7 & 0.2\\0.2 & 0.7\end{bmatrix}$ 和 $\boldsymbol{P}_2=\begin{bmatrix}0.1\\0.1\end{bmatrix}$,因此它是对称信道。

定理 3.4 对于准对称离散信道,当输入等概率时达到信道容量,其信道容量为

$$C=H(Y)-H(p'_1,p'_2,\cdots,p'_s)=\log r-\sum_{k=1}^{n}N_k\log M_k-H(p'_1,p'_2,\cdots,p'_s)$$

其中,$H(Y)$为输入等概时信道输出的熵,n 为准对称离散信道矩阵按列可以划分成互不相交的子集,N_k 是第 k 个子矩阵中的行元素之和,M_k 是第 k 个子矩阵中的列元素之和,$H(p'_1,p'_2,\cdots,p'_s)$就是信道矩阵 $\boldsymbol{P}$ 中行元素集合$\{p'_1,p'_2,\cdots,p'_s\}$的 s 个元素构成的熵函数。

【例 3.8】 设某信道的转移矩阵为

$$\boldsymbol{P}=\begin{bmatrix}0.5 & 0.3 & 0.2\\0.3 & 0.5 & 0.2\end{bmatrix}$$

计算其信道容量。

分析:该信道为一个准对称信道,计算信道容量即为输入等概时的平均互信息量。

解:将 $\boldsymbol{P}=\begin{bmatrix}0.5 & 0.3 & 0.2\\0.3 & 0.5 & 0.2\end{bmatrix}$ 划分为两个对称子矩阵,即

$$\boldsymbol{P}_1=\begin{bmatrix}0.5 & 0.3\\0.3 & 0.5\end{bmatrix},\quad \boldsymbol{P}_2=\begin{bmatrix}0.2\\0.2\end{bmatrix}$$

则

$$r=2,\quad N_1=0.5+0.3=0.8,\quad M_1=0.5+0.3=0.8$$
$$N_2=0.2,\quad M_2=0.2+0.2=0.4,\quad n=2$$

所以该准对称离散信道的信道容量为

$$\begin{aligned}C&=\log r-\sum_{k=1}^{n}N_k\log M_k-H(p'_1,p'_2,\cdots,p'_s)\\&=\log 2-(0.8\log 0.8+0.2\log 0.4)-H(0.5,0.3,0.2)\\&=0.036(\text{bit/符号})\end{aligned}$$

需要指出的是,由于本例的信道为准对称信道,信道容量也可以通过计算输入等概时的平均互信息得到。

3.8 信道容量的迭代计算

对于一般离散信道来说,信道容量的计算比较复杂。迭代计算是一种常用的近似方法。

设单符号离散信道的输入符号集 $X=\{a_1,a_2,\cdots,a_r\}$，输出符号集 $Y=\{b_1,b_2,\cdots,b_s\}$，信道的传递概率 $P(Y|X)$：$\{p(b_j|a_i)(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\}$。若输入符号集 $X=\{a_1,a_2,\cdots,a_r\}$的概率分布 $P(X)$：$\{p(a_i)(i=1,2,\cdots,r)\}$，则信道的平均交互信息量为

$$I(X;Y)=H(X)-H(X\mid Y)=-\sum_{i=1}^{r}p(a_i)\ln p(a_i)+\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_i)p(b_j\mid a_i)\ln p(a_i\mid b_j) \tag{3.34}$$

是输入信源 $X=\{a_1,a_2,\cdots,a_r\}$的概率分布 $P(X)$：$\{p(a_i)(i=1,2,\cdots,r)\}$和后验概率 $P(X|Y)$：$\{p(a_i|b_j)(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\}$的函数

$$I(X;Y)=I[p(a_i),p(a_i\mid b_j)] \tag{3.35}$$

而先验概率 $p(a_i)$和后验概率 $p(a_i|b_j)$不是两个独立的变量，它们之间按照关系式

$$p(a_i\mid b_j)=\frac{p(a_i)p(b_j\mid a_i)}{\sum_{i=1}^{r}p(a_i)p(b_j\mid a_i)}\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.36}$$

发生相应的变化。

为了导出近似算法，暂时把后验概率 $p(a_i|b_j)$当作自变量，而把本来要随之发生相应变化的应急变量 $p(a_i)$近似看作固定不变量。由于 $p(b_j|a_i)$也固定不变，则式(3.34)所示的平均交互信息量 $I(X;Y)$就可以看作是后验概率 $p(a_i|b_j)$的函数，即

$$I(X;Y)=I[p(a_i\mid b_j)] \tag{3.37}$$

由于“底”大于1的对数是∩型凸函数，所以 $I[p(a_i|b_j)]$具有上凸性。这样，就可以在条件

$$\sum_{i=1}^{r}p(a_i\mid b_j)=1\quad(j=1,2,\cdots,s) \tag{3.38}$$

的约束下，对变量 $p(a_i|b_j)$求∩型凸函数 $I[p(a_i|b_j)]$的条件极大值，以及达到极大值的 $p(a_i|b_j)$ $(i=1,2,\cdots,r;\ j=1,2,\cdots,s)$。

因此，作辅助函数

$$F[p(a_i\mid b_j),\lambda]=I[p(a_i\mid b_j)]+\lambda_j\left[\sum_{i=1}^{r}p(a_i\mid b_j)-1\right]\quad(j=1,2,\cdots,s)$$

取函数 $F[p(a_i|b_j),\lambda]$对 $p(a_i|b_j)$的偏导数，并置为0，得到稳定点方程

$$\frac{\partial}{\partial p(a_i\mid b_j)}F[p(a_i\mid b_j),\lambda]=0\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.39}$$

把式(3.34)代入式(3.39)，有

$$\frac{p(a_i)p(b_j\mid a_i)}{p(a_i\mid b_j)}+\lambda_j=0\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.40}$$

即有

$$p(a_i\mid b_j)=-\frac{p(a_i)p(b_j\mid a_i)}{\lambda_j}\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.41}$$

由约束条件式(3.38)，得

$$\lambda_j=-\sum_{i=1}^{r}p(a_i)p(b_j\mid a_i)\quad(j=1,2,\cdots,s) \tag{3.42}$$

则可以得到

$$p^*(a_i \mid b_j) = \frac{p(a_i)p(b_j \mid a_i)}{\sum_{i=1}^{r} p(a_i)p(b_j \mid a_i)} \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.43}$$

这表明,当采用"把后验概率 $P(X|Y)$:$\{p(a_i|b_j)(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\}$看作变量,信源 $X=\{a_1,a_2,\cdots,a_r\}$的概率分布 $P(X)$:$\{p(a_i)(i=1,2,\cdots,r)\}$看作固定不变的量"这种近似处理的方法时,使平均交互信息量 $I[p(a_i|b_j)]$达到最大值,即信道容量 C 的后验概率 $p^*(a_i|b_j)$就是信源 $X=\{a_1,a_2,\cdots,a_r\}$的概率分布 $P(X)$:$\{p(a_i)(i=1,2,\cdots,r)\}$时,给定信道 $P(Y|X)$:$\{p(b_j|a_i)(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\}$的一般意义下的后验概率 $p(a_i|b_j)$。这是因为对给定信道 $P(Y|X)$:$\{p(b_j|a_i)(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\}$,当输入信源 $X=\{a_1,a_2,\cdots,a_r\}$的概率分布 $P(X)$:$\{p(a_i)(i=1,2,\cdots,r)\}$固定不变时,其信道的平均交互信息量 $I(X;Y)=I[p(a_i),p(b_j|a_i)]$只有一个确定的值,其最大值也只可能就是这唯一的确定值。达到这唯一确定值的后验概率 $p^*(a_i|b_j)$当然只可能就是由式(3.36)所规定的一般意义下的后验概率 $p(a_i|b_j)$。所以,由式(3.43)可知,当变动后验概率 $p(a_i|b_j)$,而固定信源 $X=\{a_1,a_2,\cdots,a_r\}$的概率分布 $P(X)$:$\{p(a_i)(i=1,2,\cdots,r)\}$时,信道容量

$$C = \max_{p(a_i|b_j)} \{I[p(a_i),p(a_i \mid b_j)]\} = I[p(a_i),p^*(a_i \mid b_j)] \tag{3.44}$$

另一方面,同样可把信源 $X=\{a_1,a_2,\cdots,a_r\}$的概率分布 $P(X)$:$\{p(a_i)(i=1,2,\cdots,r)\}$当作自变量,把本来要随之发生相应变化的应变量 $p(a_i|b_j)$近似地看作固定不变的量。在采用这种近似处理时,由于 $p(b_j|a_i)$和 $p(a_i|b_j)$都是固定不变,则式(3.34)所示的平均交互信息量 $I(X;Y)$就可以看作先验概率 $p(a_i)$的函数

$$I(X;Y) = I[p(a_i)]$$

由于 $I[p(a_i)]$是 $p(a_i)$的$\cap$型凸函数,所以在条件

$$\sum_{i=1}^{r} p(a_i) = 1 \tag{3.45}$$

的约束下,对变量 $p(a_i)$求函数 $I[p(a_i)]$的条件极大值,以及达到极大值的 $p^*(a_i)(i=1,2,\cdots,r)$。

因此,作辅助函数

$$F[p(a_i),\lambda] = I[p(a_i)] + \lambda\left[\sum_{i=1}^{r} p(a_i) - 1\right] \tag{3.46}$$

取函数 $F[p(a_i),\lambda]$对 $p(a_i)$的偏导,并置为0,得到稳定点方程

$$\frac{\partial}{\partial p(a_i)} F[p(a_i),\lambda] = 0 \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s) \tag{3.47}$$

把式(3.34)代入式(3.47),有

$$\begin{aligned}
&\frac{\partial}{\partial p(a_i)} F[p(a_i),\lambda] \\
=&\frac{\partial}{\partial p(a_i)}\left\{-\sum_{i=1}^{r} p(a_i)\ln p(a_i) + \sum_{i=1}^{r}\sum_{j=1}^{s} p(a_i)p(b_j \mid a_i)\ln p(a_i \mid b_j) + \lambda\left[\sum_{i=1}^{r} p(a_i) - 1\right]\right\}
\end{aligned}$$

$$=-[\ln p(a_i)+1]+\sum_{j=1}^{s}p(b_j\mid a_i)\ln p(a_i\mid b_j)+\lambda$$

$$=-\ln p(a_i)+\sum_{j=1}^{s}p(b_j\mid a_i)\ln p(a_i\mid b_j)+\lambda-1$$

$$=0\quad(i=1,2,\cdots,r)\tag{3.48}$$

即有

$$\ln p(a_i)=\sum_{j=1}^{s}p(b_j\mid a_i)\ln p(a_i\mid b_j)+\lambda-1$$

即

$$p(a_i)=\exp\left\{\sum_{j=1}^{s}p(b_j\mid a_i)\ln p(a_i\mid b_j)+\lambda-1\right\}\quad(i=1,2,\cdots,r)\tag{3.49}$$

由约束条件式(3.45),得

$$1=\sum_{i=1}^{r}\exp\left\{\sum_{j=1}^{s}p(b_j\mid a_i)\ln p(a_i\mid b_j)+\lambda-1\right\}$$

$$=\sum_{i=1}^{r}\exp\left\{\sum_{j=1}^{s}p(b_j\mid a_i)\ln p(a_i\mid b_j)\right\}\cdot e^{\lambda-1}\tag{3.50}$$

即得

$$e^{\lambda-1}=\frac{1}{\sum_{i=1}^{r}\exp\left\{\sum_{j=1}^{s}p(b_j\mid a_i)\ln p(a_i\mid b_j)\right\}}$$

则可以得到

$$p^*(a_i)=\frac{\exp\left\{\sum_{j=1}^{s}p(b_j\mid a_i)\ln p(a_i\mid b_j)\right\}}{\sum_{i=1}^{r}\exp\sum_{j=1}^{s}p(b_j\mid a_i)\ln p(a_i\mid b_j)}\quad(i=1,2,\cdots,r)\tag{3.51}$$

若令

$$E_i=\exp\left\{\sum_{j=1}^{s}p(b_j\mid a_i)\ln p(a_i\mid b_j)\right\}\quad(i=1,2,\cdots,r)$$

由式(3.51),得

$$p^*(a_i)=\frac{E_i}{\sum_{i=1}^{r}E_i}\quad(i=1,2,\cdots,r)\tag{3.52}$$

则信道容量

$$C=\max_{p(a_i)}\{I[p(a_i),p(a_i\mid b_j)]\}=I[p^*(a_i),p(a_i\mid b_j)]\tag{3.53}$$

综上所述,式(3.44)是在信源 X:$\{a_1,a_2,\cdots,a_r\}$的概率分布 $P(X)$:$\{p(a_i)(i=1,2,\cdots,r)\}$固定不变,变动后验概率 $P(X|Y)$:$\{p(a_i|b_j)(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\}$的假设前提下,传递概率为 $P(Y|X)$:$\{p(b_j|a_i)(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\}$的给定信道的信道容量 C 的近似迭代公式。式(3.53)是在后验概率 $P(X|Y)$:$\{p(a_i|b_j)(i=1,

$2,\cdots,r;\ j=1,2,\cdots,s)\}$固定不变,变动信源 $X:\{a_1,a_2,\cdots,a_r\}$的概率分布 $P(X):\{p(a_i)(i=1,2,\cdots,r)\}$的假设前提下,传递概率为 $P(Y|X):\{p(b_j|a_i)(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\}$的给定信道的信道容量 C 的近似迭代公式。实际上,由式(3.36)可知,对于传递概率固定为 $p(b_j|a_i)$的给定信道,在变动后验概率 $p(a_i|b_j)$时,先验概率 $p(a_i)$不可能固定不变;在变动先验概率 $p(a_i)$时,后验概率 $p(a_i|b_j)$不可能固定不变。迭代计算法就是用分别单独变动 $p(a_i|b_j)$和 $p(a_i)$的方法,逼近 $p(a_i|b_j)$和 $p(a_i)$同时变动的实际情况,求得信道容量 C 的近似值。

先假定一组 $p(a_i)(i=1,2,\cdots,r)$作为起始值,并记为 $p(a_i)^{(1)}(i=1,2,\cdots,r)$。把 $p(a_i)^{(1)}(i=1,2,\cdots,r)$作为固定值,变动 $p(a_i|b_j)(i=1,2,\cdots,r;\ j=1,2,\cdots,s)$。由式(3.43)求得后验概率

$$p^*(a_i\mid b_j)=\frac{p(a_i)^{(1)}p(b_j\mid a_i)}{\sum_{i=1}^{r}p(a_i)^{(1)}p(b_j\mid a_i)}=p(a_i\mid b_j)^{(1)}\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s)\tag{3.54}$$

由式(3.44)求得信道容量

$$C=\max_{p(a_i|b_j)}\{I[p(a_i)^{(1)},p(a_i\mid b_j)]\}=I[p(a_i)^{(1)},p(a_i\mid b_j)^{(1)}]=C(1,1)$$

再把由式(3.54)所得的 $p(a_i|b_j)^{(1)}$作为固定值,变动先验概率 $p(a_i)$,由式(3.52)求得使平均交互信息量 $I[p(a_i)]$达到最大值的输入信源 $X=\{a_1,a_2,\cdots,a_r\}$的概率分布

$$\begin{aligned}p^*(a_i)&=\frac{\exp\left\{\sum_{j=1}^{s}p(b_j\mid a_i)\log p(a_i\mid b_j)^{(1)}\right\}}{\sum_{i=1}^{r}\exp\sum_{j=1}^{s}p(b_j\mid a_i)\log p(a_i\mid b_j)^{(1)}}\\&=\frac{E_i^{(1)}}{\sum_{i=1}^{r}E_i^{(1)}}=p(a_i)^{(2)}\quad(i=1,2,\cdots,r)\end{aligned}$$

由式(3.53)求得信道容量

$$C=\max_{p(a_i)}\{I[p(a_i),p(a_i\mid b_j)^{(1)}]\}=I[p(a_i)^{(2)},p(a_i\mid b_j)^{(1)}]=C(2,1)$$

以此类推,一般可有

$$C(n,n)=\max_{p(a_i|b_j)}\{I[p(a_i)^{(n)},p(a_i\mid b_j)]\}=I[p(a_i)^{(n)},p(a_i\mid b_j)^{(n)}]$$

$$C(n+1,n)=\max_{p(a_i)}\{I[p(a_i),p(a_i\mid b_j)^{(n)}]\}=I[p(a_i)^{(n+1)},p(a_i\mid b_j)^{(n)}]$$

在实际计算中,逐段比较 $p(a_i)^{(n)}$和 $p(a_i)^{(n+1)}$,$p(a_i|b_j)^{(n)}$和 $p(a_i|b_j)^{(n+1)}$,$C(n,n)$和 $C(n+1,n)$的值。当 n 次迭代和 $n+1$ 次迭代的计算值的差小到可以允许的程度时,就可认为达到了所求信道容量 C。

3.9 平均交互信息量的不增性

在实际通信系统中,常需对信道传输的数据作适当处理,若把数据处理装置也看作是一

个信道，由两个信道串接，就组成了一个串接信道。

设信道Ⅰ和信道Ⅱ串接，组成图3.15所示串接信道。信道Ⅰ的输入符号集 $X=\{a_1, a_2, \cdots, a_r\}$，输出符号集 $Y=\{b_1, b_2, \cdots, b_s\}$，信道Ⅱ的输出符号集 $Z=\{c_1, c_2, \cdots, c_t\}$。又设，信道Ⅰ的传递概率 $P(Y|X)$：$\{p(b_j|a_i)(i=1,2,\cdots,s;\ j=1,2,\cdots,s)\}$，信道Ⅱ的传递概率 $P(Z|XY)$：$\{p(c_k|a_ib_j)(i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t)\}$。且假定 $p(a_ib_jc_k)>0(i=1,2,\cdots,r;\ j=1,2,\cdots s;\ k=1,2,\cdots,t)$。

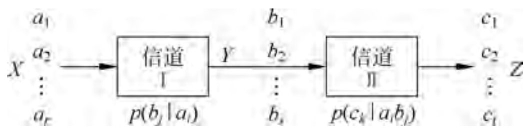

图3.15 串接信道

由一个信道组成的通信系统只有输入/输出两个随机变量 X 和 Y。由两个信道串接组成的串接信道中，就有 X、Y、Z 三个随机变量，那么由三个随机变量构成的随机变量序列 (XYZ)，在信息传输上又有什么新的特点和规律呢？

定理3.5 设由两个信道串接构成随机变量序列 (XYZ)，则

$$I(XY;Z)\geqslant I(Y;Z) \tag{3.55}$$

当且仅当

$$P(Z\mid XY)=P(Z\mid Y) \tag{3.56}$$

即 (XYZ) 是马尔可夫链时，才有

$$I(XY;Z)=I(Y;Z) \tag{3.57}$$

【证明】 根据平均交互信息量的定义，有

$$I(XY;Z)=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}p(a_ib_jc_k)\log\frac{p(c_k\mid a_ib_j)}{p(c_k)} \tag{3.58}$$

而

$$I(Y;Z)=\sum_{j=1}^{s}\sum_{k=1}^{t}p(b_jc_k)\log\frac{p(c_k\mid b_j)}{p(c_k)}=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}p(a_ib_jc_k)\log\frac{p(c_k\mid b_j)}{p(c_k)} \tag{3.59}$$

由于 $\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}p(a_ib_jc_k)=1$，以及"底"大于1的对数是 ∩ 型凸函数，可有

$$\begin{aligned}I(Y;Z)-I(XY;Z)&=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}p(a_ib_jc_k)\log\frac{p(c_k\mid b_j)}{p(c_k\mid a_ib_j)}\\&\leqslant\log\left[\sum_{i}\sum_{j}\sum_{k}p(a_ib_jc_k)\log\frac{p(c_k\mid b_j)}{p(c_k\mid a_ib_j)}\right]\\&=\log\left[\sum_{i}\sum_{j}\sum_{k}p(a_ib_j)p(c_k\mid b_j)\right]=\log1=0\end{aligned} \tag{3.60}$$

即证得

$$I(XY;Z)\geqslant I(Y;Z)$$

由式(3.56)有

$$\frac{p(c_k\mid b_j)}{p(c_k\mid a_ib_j)}=1\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t) \tag{3.61}$$

则由式(3.60)有

$$I(Y;Z)-I(XY;Z)=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}p(a_ib_jc_k)\log\frac{p(c_k\mid b_j)}{p(c_k\mid a_ib_j)}$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}p(a_ib_jc_k)\log 1=0 \tag{3.62}$$

即证得

$$I(XY;Z)=I(Y;Z)$$

定理 3.6 设由两个信道串接构成随机变量序列(XYZ),则

$$I(XY;Z)\geqslant I(X;Z) \tag{3.63}$$

当且仅当

$$P(Z\mid YX)=P(Z\mid X) \tag{3.64}$$

即(YXZ)是马尔可夫链时,才有

$$I(XY;Z)=I(X;Z) \tag{3.65}$$

【证明】 根据平均交互信息量的定义,有

$$I(XY;Z)=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}p(a_ib_jc_k)\log\frac{p(c_k\mid a_ib_j)}{p(c_k)}$$

而

$$I(X;Z)=\sum_{i=1}^{r}\sum_{k=1}^{t}p(a_ic_k)\log\frac{p(c_k\mid a_i)}{p(c_k)}=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}p(a_ib_jc_k)\log\frac{p(c_k\mid a_i)}{p(c_k)} \tag{3.66}$$

由于$\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}p(a_ib_jc_k)=1$,以及“底”大于1的对数是$\cap$型凸函数,可有

$$I(X;Z)-I(XY;Z)=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}p(a_ib_jc_k)\log\frac{p(c_k\mid a_j)}{p(c_k\mid a_ib_j)}$$

$$\leqslant\log\left[\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}p(a_ib_jc_k)\log\frac{p(c_k\mid a_i)}{p(c_k\mid a_ib_j)}\right]$$

$$=\log\left[\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}p(a_ib_jc_k)\log p(c_k\mid a_i)\right]=\log 1=0 \tag{3.67}$$

即证得

$$I(XY;Z)\geqslant I(X;Z)$$

由式(3.64)有

$$\frac{p(c_k\mid a_i)}{p(c_k\mid a_ib_j)}=1\quad(i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t) \tag{3.68}$$

则由式(3.67)有

$$I(X;Z)-I(XY;Z)=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}p(a_ib_jc_k)\log\frac{p(c_k\mid a_i)}{p(c_k\mid a_ib_j)}$$

$$=\sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}p(a_ib_jc_k)\log 1=0 \tag{3.69}$$

即证得

$$I(XY;Z)=I(X;Z)$$

定理 3.7 设由两个信道串接构成的随机变量序列(XYZ)。当

$$P(Z \mid YX)=P(Z \mid Y) \tag{3.70}$$

即当(XYZ)是马尔可夫链时,有

$$I(X;Z) \leqslant I(Y;Z) \tag{3.71}$$

当且仅当

$$P(Z \mid XY)=P(Z \mid X) \tag{3.72}$$

即(YXZ)亦是马尔可夫链时,有

$$I(X;Z)=I(Y;Z) \tag{3.73}$$

【证明】 由定理 3.5 可知当随机变量序列(XYZ)是马尔可夫链时,有

$$I(XY;Z)=I(Y;Z) \tag{3.74}$$

由定理 3.6 可知,在随机变量序列(YXZ)不是马尔可夫链的情况下,有

$$I(XY;Z) \geqslant I(X;Z) \tag{3.75}$$

由式(3.74)和式(3.75)可得,在(XYZ)是马尔可夫链,而(YXZ)不是马尔可夫链的情况下,有

$$I(X;Z) \leqslant I(Y;Z)$$

由定理 3.6 可知,当(YXZ)是马尔可夫链时,有

$$I(XY;Z)=I(X;Z) \tag{3.76}$$

则由式(3.74)和式(3.75)证得,当(XYZ)和(YXZ)都是马尔可夫链时,即

$$P(Z \mid XY)=P(Z \mid Y)$$
$$P(Z \mid YX)=P(Z \mid X)$$

有

$$I(X;Z)=I(Y;Z)$$

证毕。

对于图 3.15 所示串接信道,在工程上一般把随机变量序列(XYZ)看作马尔可夫链,整个串接信道的传递概率 $p(c_k|a_i)(i=1,2,\cdots,r;\ k=1,2,\cdots,t)$为

$$p(c_k \mid a_i)=\sum_{j=1}^{s} p(b_j \mid a_i)p(c_k \mid b_j) \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t) \tag{3.77}$$

这就是说,整个串接信道的信道矩阵 $\boldsymbol{P}(s\times t)$,等于信道Ⅰ的信道矩阵 $\boldsymbol{P}_{\mathrm{I}}(r\times s)$与信道Ⅱ的信道矩阵 $\boldsymbol{P}_{\mathrm{II}}(s\times t)$相乘,即

$$\boldsymbol{P}=\boldsymbol{P}_{\mathrm{I}} \cdot \boldsymbol{P}_{\mathrm{II}} \tag{3.78}$$

若要求随机变量序列(XYZ)和(YXZ)都是马尔可夫链,则要有

$$\begin{cases} P(Z \mid XY)=P(Z \mid Y) \\ P(Z \mid YX)=P(Z \mid X) \end{cases}$$

即

$$\begin{cases} p(c_k \mid a_i b_j)=\dfrac{c_k}{b_j} \\ p(c_k \mid b_j a_i)=\dfrac{c_k}{a_i} \end{cases} \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t) \tag{3.79}$$

即

$$p(c_k \mid a_i)=p(c_k \mid b_j) \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t) \tag{3.80}$$

这就是说,整个串接信道的信道矩阵 $\boldsymbol{P}$ 与信道Ⅱ的信道矩阵 $\boldsymbol{P}_{\mathrm{II}}$ 要完全一样,即

$$\boldsymbol{P}=\boldsymbol{P}_{\mathrm{I}} \cdot \boldsymbol{P}_{\mathrm{II}}=\boldsymbol{P}_{\mathrm{II}} \tag{3.81}$$

显然,信道Ⅰ的矩阵 $\boldsymbol{P}_{\mathrm{I}}$ 是 s 阶单位矩阵,是式(3.81)能得到满足的一种情况,这时有

$$
\begin{array}{cc}
 & \begin{array}{cccc} b_1 & b_2 & \cdots & b_s \end{array} \\
\begin{array}{c} a_1 \\ a_2 \\ \vdots \\ a_s \end{array} & \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix}
\end{array}
\cdot
\begin{array}{cc}
 & \begin{array}{cccc} c_1 & c_2 & \cdots & c_t \end{array} \\
\begin{array}{c} b_1 \\ b_2 \\ \vdots \\ b_s \end{array} & \begin{bmatrix} p(c_1 \mid b_1) & p(c_2 \mid b_1) & \cdots & p(c_t \mid b_1) \\ p(c_1 \mid b_2) & p(c_2 \mid b_2) & \cdots & p(c_t \mid b_2) \\ \vdots & \vdots & \ddots & \vdots \\ p(c_1 \mid b_s) & p(c_2 \mid b_s) & \cdots & p(c_t \mid b_s) \end{bmatrix}
\end{array}
\tag{3.82}
$$

$$
=\begin{array}{cc}
 & \begin{array}{cccc} c_1 & c_2 & \cdots & c_t \end{array} \\
\begin{array}{c} a_1 \\ a_2 \\ \vdots \\ a_r \end{array} & \begin{bmatrix} p(c_1 \mid b_1) & p(c_2 \mid b_1) & \cdots & p(c_t \mid b_1) \\ p(c_1 \mid b_2) & p(c_2 \mid b_2) & \cdots & p(c_t \mid b_2) \\ \vdots & \vdots & \ddots & \vdots \\ p(c_1 \mid b_s) & p(c_2 \mid b_s) & \cdots & p(c_t \mid b_s) \end{bmatrix}
\end{array}
$$

式(3.82)表明,当信道Ⅰ是一一对应的确定关系的一般无噪信道时,随机变量序列(XYZ)与随机变量序列(XYZ)一样,亦是马尔可夫链。

由以上分析可以得到这样一个结论:在一般情况下(XYZ)是马尔可夫链,输出随机变量 Z 通过信道Ⅱ一个信道,获取关于随机变量 Y 的信息量 $I(Y;Z)$,总是比输出随机变量 Z 通过信道Ⅱ、信道Ⅰ两个信道,获取关于随机变量 X 的信息量 $I(X;Z)$ 大。这就是说,随机变量 Y 通过信道Ⅰ到随机变量 X,总是要丢失一部分信息量(图3.16)。只有当随机变量序列(YXZ)亦是马尔可夫链(如信道Ⅰ是一一对应确定关系的一般无噪信道)时,随机变量 Y 到随机变量 X 才不会丢失信息量,因而从 Z 中获取关于 Y 的信息量 $I(Y;Z)$,才与从 Z 中获取关于 X 的信息量 $I(X;Z)$ 相等。

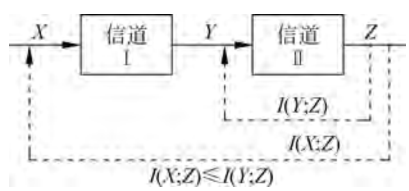

图3.16 关注信道Ⅱ

定理3.8 设由两个信道串接构成随机变量序列(XYZ),当

$$P(Z \mid XY)=P(Z \mid Y) \tag{3.83}$$

即当(XYZ)是马尔可夫链时,有

$$I(X;Y) \geqslant I(X;Z) \tag{3.84}$$

当且仅当

$$P(X \mid YZ)=P(X \mid Z) \tag{3.85}$$

即(YZX)亦是马尔可夫链时,才有

$$I(X;Y)=I(X;Z) \tag{3.86}$$

【证明】 由概率一般运算规则有

$$P(Z \mid XY) = \frac{P(XYZ)}{P(XY)} = \frac{P(YZ)P(X \mid YZ)}{P(Y)P(X \mid Y)}$$
$$= P(Z \mid Y) \cdot \frac{P(X \mid YZ)}{P(X \mid Y)} \tag{3.87}$$

由式(3.83),有

$$P(X \mid ZY) = P(X \mid Y) \tag{3.88}$$

这表明,当随机变量序列(XYZ)是马尔可夫链时,随机变量序列(ZYX)亦是马尔可夫链,根据定理3.5,有

$$I(ZY; X) = I(Y; X) \tag{3.89}$$

在一般情况下,随机变量序列(YZX)不是马尔可夫链,根据定理3.6,有

$$I(ZY; X) \geqslant I(Z; X) \tag{3.90}$$

由式(3.89)和式(3.90),有

$$I(Y; X) \geqslant I(Z; X) \tag{3.91}$$

根据平均交互信息量的交互性,可证得

$$I(X; Y) \geqslant I(X; Z)$$

当随机变量序列(YZX)亦是马尔可夫链时,根据定理3.6,有

$$I(ZY; X) = K(Z; X) \tag{3.92}$$

由式(3.89)有

$$I(Y; X) = I(Z; X)$$

根据平均交互信息量的交互性,证得

$$I(X; Y) = I(X; Z)$$

这样定理得到了证明。

对于图3.15所示串接信道,在一般情况下,随机变量序列(XYZ)都可看作是马尔可夫链,即随机变量序列(ZYX)亦可看作马尔可夫链。如要求随机变量序列(YZX)同时也是马尔可夫链,则有

$$P(X \mid ZY) = P(X \mid Y)$$
$$P(X \mid YZ) = P(X \mid Z)$$

即

$$\begin{cases} p(a_i \mid c_k b_j) = p(a_i \mid b_j) \\ p(a_i \mid c_k b_j) = p(a_i \mid c_k) \end{cases} \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t) \tag{3.93}$$

即

$$p(a_i \mid b_j) = p(a_i \mid c_k) \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t) \tag{3.94}$$

而

$$p(a_i \mid b_j) = \frac{p(a_i)p(b_j \mid a_i)}{\sum_{i=1}^{r} p(a_i)p(b_j \mid a_i)} \tag{3.95}$$

$$p(a_i \mid c_k)=\frac{p(a_i)p(c_k \mid a_i)}{\sum_{i=1}^{r} p(a_i)p(c_k \mid a_i)} \tag{3.96}$$

由式(3.94)、式(3.95)和式(3.96)可知,则

$$p(b_j \mid a_i)=p(c_k \mid a_i) \quad (i=1,2,\cdots,r;\ j=1,2,\cdots,s;\ k=1,2,\cdots,t) \tag{3.97}$$

这说明,如果要求随机变量序列(ZYX)是马尔可夫链的条件下,随机变量序列(YZX)亦是马尔可夫链,则必须要求图3.16串接信道的信道矩阵$\boldsymbol{P}$与信道Ⅰ的信道矩阵$\boldsymbol{P}_{\mathrm{I}}$完全相同,即

$$\boldsymbol{P}=\boldsymbol{P}_{\mathrm{I}}\cdot\boldsymbol{P}_{\mathrm{II}}=\boldsymbol{P}_{\mathrm{I}} \tag{3.98}$$

显然,信道Ⅱ的信道矩阵$\boldsymbol{P}_{\mathrm{II}}$是$s$阶单位矩阵,是式(3.98)能得到满足的一种情况,这时有

$$\begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_r \end{array}\begin{array}{c} \begin{array}{cccc} b_1 & b_2 & \cdots & b_s \end{array} \\ \begin{bmatrix} p(b_1 \mid a_1) & p(b_2 \mid a_1) & \cdots & p(b_s \mid a_1) \\ p(b_1 \mid a_2) & p(b_2 \mid a_2) & \cdots & p(b_s \mid a_2) \\ \vdots & \vdots & \ddots & \vdots \\ p(b_1 \mid a_r) & p(b_2 \mid a_r) & \cdots & p(b_s \mid a_r) \end{bmatrix}\end{array} \cdot \begin{array}{c} \\ b_1 \\ b_2 \\ \vdots \\ b_s \end{array}\begin{array}{c} \begin{array}{cccc} c_1 & c_2 & \cdots & c_s \end{array} \\ \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}\end{array}$$

$$=\begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_r \end{array}\begin{array}{c} \begin{array}{cccc} c_1 & c_2 & \cdots & c_s \end{array} \\ \begin{bmatrix} p(b_1 \mid a_1) & p(b_2 \mid a_1) & \cdots & p(b_s \mid a_1) \\ p(b_1 \mid a_2) & p(b_2 \mid a_2) & \cdots & p(b_s \mid a_2) \\ \vdots & \vdots & \ddots & \vdots \\ p(b_1 \mid a_r) & p(b_2 \mid a_r) & \cdots & p(b_s \mid a_r) \end{bmatrix}\end{array} \tag{3.99}$$

这表明,当信道Ⅱ是一一对应的确定关系的一般无噪信道时,随机变量序列(YZX)与随机变量序列(ZYX)一样,亦是马尔可夫链。

由以上分析可得到这样一个结论:对于图3.16所示串接信道,在一般情况下,由于随机变量序列(XYZ)可看作马尔可夫链,所以随机变量序列(ZYX)亦可看作马尔可夫链。随机变量X通过信道Ⅰ获取关于随机变量Y的信息量$I(X;Y)$,总比随机变量X通过信道Ⅰ和信道Ⅱ两个信道,获取关于随机变量Z的信息量$I(X;Z)$大。这就是说,随机变量Y通过信道Ⅱ到随机变量Z,总要丢失一部分信息量(图3.17)。只有当随机变量序列(YZX)亦是马尔可夫链(如信道Ⅱ是一一对应确定关系的一般无噪信道时),随机变量Y到随机变量Z才不会丢失信息量。因而从随机变量X中获取关于随机变量Y的信息量$I(X;Y)$,才与从随机变量X中获取关于随机变量Z的信息量$I(X;Z)$相等。

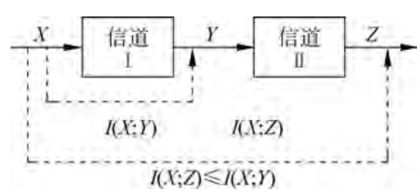

图3.17 关注信道Ⅰ

【例3.9】 设信道Ⅰ和信道Ⅱ相接,组成图3.18所示串接信道。若随机变量序列(XYZ)是马尔可夫链,试证明$I(X;Z)=I(X;Y)$。

证明: 因为随机变量序列(XYZ)是马尔可夫链,由图3.18可知,串接信道的信道矩阵

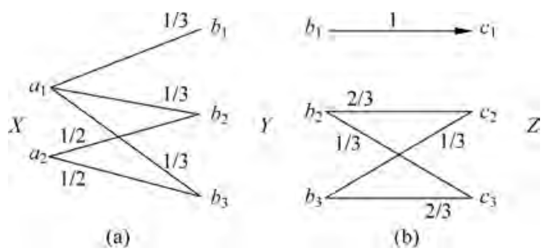

图 3.18 (XYZ)构成马尔可夫链

$$\boldsymbol{P}=\begin{array}{c} \\ a_1 \\ a_2 \end{array}\!\!\begin{array}{c}\begin{array}{ccc} b_1 & b_2 & b_3 \end{array}\\ \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \end{bmatrix}\end{array} \cdot \begin{array}{c} \\ b_1 \\ b_2 \\ b_3 \end{array}\!\!\begin{array}{c}\begin{array}{ccc} c_1 & c_2 & c_3 \end{array}\\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2/3 & 1/3 \\ 0 & 1/3 & 2/3 \end{bmatrix}\end{array} = \begin{array}{c} \\ a_1 \\ a_2 \end{array}\!\!\begin{array}{c}\begin{array}{ccc} c_1 & c_2 & c_3 \end{array}\\ \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \end{bmatrix}\end{array}$$

由于串接信道的信道矩阵 $\boldsymbol{P}$ 与信道Ⅰ的信道矩阵 $\boldsymbol{P}_{\mathrm{I}}$ 完全一致,所以有

$$p(b_j \mid a_i)=p(c_k \mid a_i) \quad (i=1,2;\ j=1,2,3;\ k=1,2,3)$$

根据式(3.93)~式(3.97),随机变量序列(YZX)亦是马尔可夫链。根据定理 3.8,可证得

$$I(X;Y)=I(X;Z)$$

这一特例说明,在(XYZ)是马尔可夫链的条件下,信道Ⅱ是一一对应确定关系的一般无噪信道,式(3.84)是式(3.85)成立的必要条件,但不是充分条件。

推论 设有两个信道串接构成随机变量序列(XYZ)。若随机变量序列(XYZ)是马尔可夫链,则有

$$I(X;Z)\leqslant\begin{cases} I(Y;Z) \\ I(X;Y) \end{cases} \tag{3.100}$$

【证明】 因为随机变量序列(XYZ)是马尔可夫链,则由定理 3.7 有

$$I(X;Z)\leqslant I(Y;Z)$$

又因为当(XYZ)是马尔可夫链时,随机变量序列(ZYX)亦是马尔可夫链,由定理 3.8,有 $I(X;Z)\leqslant I(X;Y)$。

即证得,当随机变量序列(XYZ)是马尔可夫链时,有

$$I(X;Z)\leqslant\begin{cases} I(Y;Z) \\ I(X;Y) \end{cases}$$

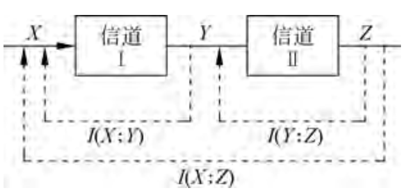

图 3.19 平均交互信息量之间的关系

这样,推论就得到了证明。平均交互信息量 $I(X;Z)$与 $I(Y;Z)$和 $I(X;Y)$之间的关系如图 3.19 所示。

推论指出,若以 Z 作为观察点,把信道Ⅰ看作一个数据处理装置(图 3.20),随机变量序列(XYZ)是马尔可夫链。因随机变量 Y 经过处理后,变成随机变量 X,则从随机变量 Z 中获取关于随机变量 X 的平

均交互信息量 $I(X;Z)$，不会超过数据处理前从随机变量 Z 中获取关于随机变量 Y 的平均交互信息量 $I(Y;Z)$，最多两者相等。这就是说，把随机变量 Y 变成随机变量 X 的数据处理过程，总是要丢失一部分信息。只有当随机变量序列 (YXZ) 亦是马尔可夫链时（数据处理是一一对应确定关系时），数据处理过程才不会丢失信息。

推论又指出，若以 X 作为观察点，把信道Ⅱ看作是一个数据处理装置（图 3.21），则随机变量序列 (ZYX) 是马尔可夫链。因为随机变量 Y 经过数据处理后，变成随机变量 Z，从随机变量 X 中获取关于随机变量 Z 的平均交互信息量 $I(X;Z)$，不会超过数据处理前从随机变量 X 中获取关于随机变量 Y 的平均交互信息量 $I(X;Y)$，最多两者相等。这就是说，把随机变量 Y 变成随机变量 Z 的数据处理过程，总是要丢失一部分信息。只有当随机变量序列 (YZX) 亦是马尔可夫链时（数据处理是一一对应的确定关系时），数据处理过程才不会丢失信息。

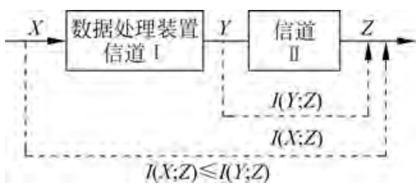

图 3.20　信道Ⅰ

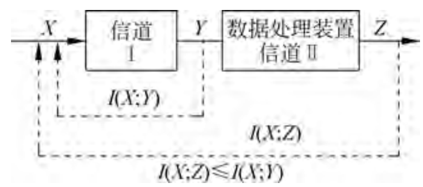

图 3.21　信道Ⅱ

综合推论所指出的两方面的结论，可以得到一个总的结论：任何无源数据处理过程都要丢失一部分信息量，最多不丢失信息量，但一定不会增加信息量。由定理 3.5～定理 3.8 及其推论所阐明的平均交互信息量的不增性，一般通称为数据处理定理。

数据处理定理是信息传输的重要规律。把数据处理定理与前面讨论的平均交互信息量的极值性结合起来，就可导出信息传输和处理的一个完整的概念，如图 3.22(a)和图 3.22(b)所示。对于信源 X 来说，经信息传输或信息处理，最终从随机变量 Q 中获取关于信源 X 的平均交互信息量 $I(X;Q)$，绝对不会超过信源 X 本身含有的平均信息量 $H(X)$，最多等于信源 X 本身含有的平均信息量 $H(X)$。信息所经过的传输信道或数据处理装置越多，丢失的信息量就可能越多。即

$$H(X) \geqslant I(X;Y) \geqslant I(X;Z) \geqslant I(X;W) \geqslant \cdots \geqslant I(X;Q) \tag{3.101}$$

对于随机变量 Q 来说，经信息传输或信息处理，最终从随机变量 X 中获取关于随机变量 Q 的平均交互信息量 $I(Q;X)$，绝不会超过随机变量 Q 本身含有的平均信息量 $H(Q)$，最多等于 $H(Q)$。信息所经过的传输信道或数据处理装置越多，丢失的信息量就可能越多。即

$$H(Q) \geqslant I(W;Q) \geqslant I(Z;Q) \geqslant I(Y;Q) \geqslant \cdots \geqslant I(X;Q) \tag{3.102}$$

在传输或处理过程中，一旦在某一环节（信道或处理装置）丢失一部分信息，以后的系统不管怎样传输或处理，只要不接触到丢失信息环节的输入端（如多次测量），就不能再恢复已丢失的信息。

【例 3.10】　设有三个二进制对称信道(BSC)串接组成图 3.23 所示串接信道。随机变量序列 $(XYZW)$ 是马尔可夫链。若信源 $X=\{0,1\}$ 是等概信源，试求平均交互信息量 $I(X;Y)$、$I(X;Z)$、$I(X;W)$，并比较它们的大小。

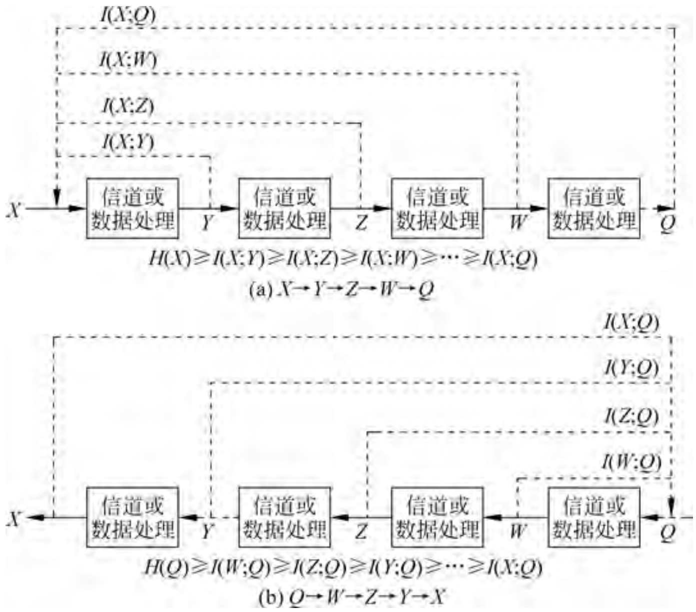

图 3.22 信息不增性示意图

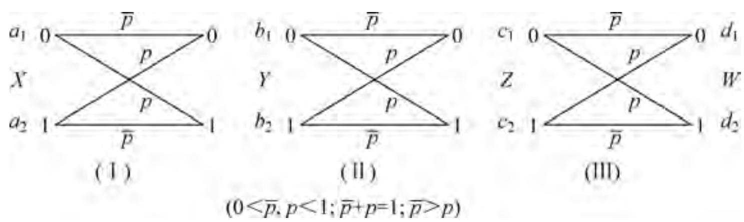

图 3.23 三个信道串联

解：(1) 因为信道Ⅰ的矩阵为

$$\begin{array}{cc} & \begin{array}{cc} 0 & 1 \end{array} \\ \boldsymbol{P}_{\mathrm{I}} = \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix} \end{array}$$

信源 X 的概率分布为

$$P\{X=0\}=p(0)=\frac{1}{2};\quad P\{X=1\}=p(1)=\frac{1}{2}$$

所以，随机变量 Y 的概率分布为

$$\begin{aligned} P\{Y=0\} &= P\{X=0\}\cdot P\{Y=0\mid X=0\}+P\{X=1\}\cdot P\{Y=0\mid X=1\} \\ &= p(0)\bar{p}+p(1)p=\frac{1}{2}(\bar{p}+p)=\frac{1}{2} \end{aligned}$$

$$\begin{aligned} P\{Y=1\} &= P\{X=0\}\cdot P\{Y=1\mid X=0\}+P\{X=1\}\cdot P\{Y=1\mid X=1\} \\ &= p(0)p+p(1)\bar{p}=\frac{1}{2}(\bar{p}+p)=\frac{1}{2} \end{aligned}$$

则随机变量 Y 的熵

$$H(Y)=H\left(\frac{1}{2},\frac{1}{2}\right)=1\ (\text{bit/符号})$$

由信道 Ⅰ 的矩阵 $\boldsymbol{P}_{\mathrm{I}}$，有

$$\begin{aligned}
H(Y\mid X)&=-\sum_{i=1}^{2}\sum_{j=1}^{2}p(a_i)p(b_j\mid a_i)\log p(b_j\mid a_i)\\
&=-\{[p(0)\bar{p}\log\bar{p}+p(0)p\log p]+[p(1)p\log p+p(1)\bar{p}\log\bar{p}]\}\\
&=p(0)[-(\bar{p}\log\bar{p}+p\log p)]+p(1)[-(\bar{p}\log\bar{p}+p\log p)]\\
&=[p(0)+p(1)][-(\bar{p}\log\bar{p}+p\log p)]\\
&=-(\bar{p}\log\bar{p}+p\log p)=H(\bar{p},p)=H(p)
\end{aligned}$$

所以，信道 Ⅰ 的平均交互信息量

$$\begin{aligned}
I(X;Y)&=H(Y)-H(Y\mid X)\\
&=1-H(p)\\
&=1-H\left[\frac{1}{2}-\frac{1}{2}(1-2p)\right]\ (\text{bit/符号})
\end{aligned}\tag{3.103}$$

(2) 因为随机变量序列($XYZW$)是马尔可夫链，所以信道Ⅰ、信道Ⅱ的串接信道($X-Z$)的信道矩阵 $\boldsymbol{P}_{X-Z}$，等于信道 Ⅰ 的矩阵 $\boldsymbol{P}_{\mathrm{I}}$ 与信道 Ⅱ 的矩阵 $\boldsymbol{P}_{\mathrm{II}}$ 的连乘，即有

$$\boldsymbol{P}_{X-Z}=\begin{array}{c} \\ 0\\ 1\end{array}\!\!\begin{array}{c}\begin{array}{cc}0 & 1\end{array}\\ \begin{bmatrix}\bar{p} & p\\ p & \bar{p}\end{bmatrix}\end{array}\cdot\begin{array}{c} \\ 0\\ 1\end{array}\!\!\begin{array}{c}\begin{array}{cc}0 & 1\end{array}\\ \begin{bmatrix}\bar{p} & p\\ p & \bar{p}\end{bmatrix}\end{array}=\begin{array}{c} \\ 0\\ 1\end{array}\!\!\begin{array}{c}\begin{array}{cc}0 & \qquad 1\end{array}\\ \begin{bmatrix}\bar{p}^2+p^2 & 2\bar{p}p\\ 2\bar{p}p & \bar{p}^2+p^2\end{bmatrix}\end{array}$$

随机变量 Z 的概率分布为

$$\begin{aligned}
P\{Z=0\}&=P\{X=0\}\cdot P\{Z=0\mid X=0\}+P\{X=1\}\cdot P\{Z=0\mid X=1\}\\
&=p(0)(\bar{p}^2+p^2)+p(1)(2\bar{p}p)=\frac{1}{2}[(\bar{p}^2+p^2)+2\bar{p}p]=\frac{1}{2}
\end{aligned}$$

$$P\{Z=1\}=1-P\{Z=0\}=\frac{1}{2}$$

则得随机变量 Z 的熵

$$H(Z)=H\left(\frac{1}{2},\frac{1}{2}\right)=1(\text{bit/符号})$$

由串接信道($X-Z$)的信道矩阵 $\boldsymbol{P}_{X-Z}$，有

$$\begin{aligned}
H(Z\mid X)&=-\sum_{i=1}^{2}\sum_{k=1}^{2}p(a_i)p(c_k\mid a_i)\log p(c_k\mid a_i)\\
&=-\{[p(0)(\bar{p}^2+p^2)\log(\bar{p}^2+p^2)+p(0)(2\bar{p}p)\log(2\bar{p}p)]+\\
&\qquad[p(1)(2\bar{p}p)\log(2\bar{p}p)+p(1)(\bar{p}^2+p^2)\log(\bar{p}^2+p^2)]\}\\
&=p(0)\{-[(\bar{p}^2+p^2)\log(\bar{p}^2+p^2)+(2\bar{p}p)\log(2\bar{p}p)]\}+\\
&\qquad p(1)\{-[(2\bar{p}p)\log(2\bar{p}p)+(\bar{p}^2+p^2)\log(\bar{p}^2+p^2)]\}\\
&=[p(0)+p(1)]\{-[(\bar{p}^2+p^2)\log(\bar{p}^2+p^2)+(2\bar{p}p)\log(2\bar{p}p)]\}
\end{aligned}$$

$$=H(2\bar{p}p)=H[\frac{1}{2}-\frac{1}{2}(1-2p)^2]$$

则

$$I(X;Z)=H(Z)-H(Z\mid X)=1-H(2\bar{p}p)$$
$$=1-H\left[\frac{1}{2}-\frac{1}{2}(1-2p)^2\right] \tag{3.104}$$

同理可得

$$I(X;W)=1-H\left[\frac{1}{2}-\frac{1}{2}(1-2p)^3\right] \tag{3.105}$$

则可有

$$I(X;Y)\geqslant I(X;Z)\geqslant I(X;W) \tag{3.106}$$

(3) 一般地，设有 N(大于1的正整数)个二进制对称信道，串接成如图3.24所示的串接信道。当信源 X 是等概分布时，同理可证。

$$I(X;Y_N)=1-H\left[\frac{1}{2}-\frac{1}{2}(1-2p)^N\right]\text{(bit/符号)}\quad (N=1,2,3,\cdots) \tag{3.107}$$

则有

$$I(X;Y_{N-1})\geqslant I(X;Y_N)\quad (N=2,3,4,\cdots) \tag{3.108}$$

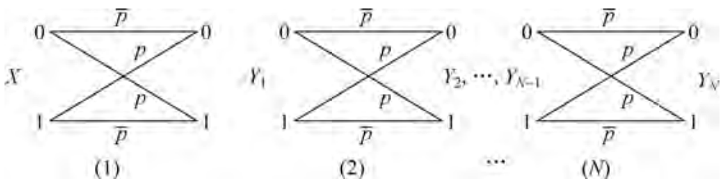

图 3.24 串接信道

式(3.108)是由 N 个二进制对称信道组成的串接信道，当信源 $X=\{0,1\}$ 为等概信源时，数据处理定理的具体表现。

实践：信道容量的迭代算法

【已知】 信源符号个数 r，信宿符号个数 s，信道转移概率矩阵 $\boldsymbol{P}=(p_{ij})^{r\times s}$。

【算法】

(1) 初始化信源分布 $p_i=\frac{1}{r}$，循环变量 $k=1$，门限为 Δ，$C^{(0)}=-\infty$。

(2) $\phi_{ij}^{(k)}=\dfrac{p_i^{(k)}p_{ij}}{\sum\limits_{i=1}^{r}p_i^{(k)}p_{ij}}$。

(3) $p_i^{(k+1)}=\dfrac{\exp\left[\sum\limits_{j=1}^{s}p_{ji}\ln\phi_{ij}^{(k)}\right]}{\sum\limits_{i=1}^{r}\exp\left[\sum\limits_{j=1}^{s}p_{ij}\ln\phi_{ij}^{(k)}\right]}$。

(4) $C^{(k+1)}=\log\left\{\sum_{i=1}^{r}\exp\left[\sum_{j=1}^{s}p_{ij}\ln\phi_{ij}^{(k)}\right]\right\}$。

(5) 若$\dfrac{|C^{(k+1)}-C^{(k)}|}{C^{(k+1)}}>\Delta$,则 $k=k+1$,转第(2)步。

(6) 输出 $\bar{P}^{*}=p_i^{(k+1)}$ 和 $C^{(k+1)}$,终止。

【要求】

(1) 输入:任意的一个信道转移概率矩阵。信源符号个数、信宿符号个数和每个具体的转移概率在运行时从键盘输入。

(2) 输出:最佳信源分布 $\bar{P}^{*}$,信道容量 C。

本章要点

1. 离散信道模型

$$X\left\{\begin{matrix}a_1\\a_2\\\vdots\\a_r\end{matrix}\right.\rightarrow p(b_j\mid a_i)\rightarrow\left.\begin{matrix}b_1\\b_2\\\vdots\\b_s\end{matrix}\right\}Y$$

2. 信道的交互信息量

$$\begin{aligned}I(a_i;b_j)&=\log\frac{p(a_i\mid b_j)}{p(a_i)}\\&=\log\frac{p(a_ib_j)}{p(a_i)p(b_j)}\\&=\log\frac{p(b_i\mid a_j)}{p(b_i)}\end{aligned}$$

3. 条件交互信息量

$$\begin{aligned}I(a_i;b_j\mid c_k)&=\log\frac{p(a_i\mid b_jc_k)}{p(a_i\mid c_k)}\\&=\log\frac{1}{p(a_i\mid c_k)}-\log\frac{1}{p(a_i\mid b_jc_k)}\\&=I(a_i\mid c_k)-I(a_i\mid b_jc_k)\end{aligned}$$

4. 平均交互信息量

$$I(X;Y)=\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)I(a_i;b_j)=I(Y;X)$$

5. 平均互信息的性质

(1) 非负性:$I(X;Y)\geqslant 0$

(2) 极值性:$I(X;Y)\leqslant H(X)$

$I(X;Y)\leqslant H(Y)$

(3) 凸函数性

平均互信息量 $I(X;Y)$,在信道转移概率 $p(b_j|a_i)$给定条件下,是输入随机变量 X 的概率分布 $P(X)$:$\{p(a_i),i=1,2,\cdots,r\}$的$\cap$型凸函数。

平均交互信息量 $I(X;Y)$,在信源概率分布 $P(a_i)$给定的条件下,是信道转移概率 $P(Y|X)$:$\{p(b_j|a_i),i=1,2,\cdots,r;\ j=1,2,\cdots,s\}$的$\cup$型凸函数。

6. 信道容量

$$C=R_{\max}=\max_{p(x)}\{I(X;Y)\}(\text{bit/符号})$$

$$C_t=R_{t\max}=\frac{1}{t}\max\{I(X;Y)\}(\text{bit/秒})$$

7. 离散对称信道容量

$$\begin{aligned}C&=\max_{p(x)}\{I(X;Y)\}=\max_{p(x)}\{H(Y)-H(Y\mid X)\}\\&=\max_{p(x)}\{H(Y)\}-H(p'_1,p'_2,\cdots,p'_s)\\&=\log s-H(p'_1,p'_2,\cdots,p'_s)=\log s-H(\boldsymbol{P}\text{ 的行向量})\end{aligned}$$

8. 准对称离散信道容量

$$C=H(Y)-H(p'_1,p'_2,\cdots,p'_s)=\log r-\sum_{k=1}^{n}N_k\log M_k-H(p'_1,p'_2,\cdots,p'_s)$$

9. 数据处理定理

信息所经过的传输信道或数据处理装置越多,丢失的信息量就可能越多。即

$$H(X)\geqslant I(X;Y)\geqslant I(X;Z)\geqslant I(X;W)\geqslant\cdots\geqslant I(X;Q)$$

第 3 章习题