第 3 章　离散信道及其信道容量

通信系统的基本功能是实现信息的传递，信道是信息传递的通道，是信号传输的媒质。一般而言，信源发出的消息，必须以适合于信道传输的信号形式经过信道的传输，才能被信宿接收。从信源的角度看，信源发出的每个符号承载的平均信息量由信源熵来定量描述；而从信宿的角度看，信宿收到的每个符号平均能提供多少信息量由平均互信息来定量描述。信宿接收的信息量不仅与信源有关，而且受到信道传输特性的影响。在信息论中，信道问题主要研究在什么条件下信道能够传送最大的信息量，即信道容量问题。

本章首先介绍离散信道的分类和数学模型，然后定量地研究信道传输的平均互信息及其重要性质，并重点讨论几种典型单符号离散信道的信道容量，进而研究一般单符号离散信道的信道容量计算方法，而后讨论多符号离散信道的信道容量，最后讨论信源和信道的匹配问题。

本章符号约定：离散信道的输入为 N 个符号构成的序列，用随机矢量表示为 $\boldsymbol{X}=(X_1X_2\cdots X_N)$，其中每个分量 $X_l(l=1,2,\cdots,N)$ 取值于同一集合 $\{a_1,a_2,\cdots,a_r\}$。对应地，离散信道的输出 $\boldsymbol{Y}=(Y_1Y_2\cdots Y_N)$，其中每个分量 $Y_l(l=1,2,\cdots,N)$ 取值于同一集合 $\{b_1,b_2,\cdots,b_s\}$。

3.1　离散信道的分类

离散信道的输入和输出都是时间离散、取值离散的随机序列，其统计特性可用信道转移概率来描述。离散信道包括单符号离散信道和多符号离散信道，其分类如图 3.1 所示。

- 离散信道
  - 单符号离散信道
    - 典型的单符号离散信道
      - 无噪信道
      - 对称离散信道
      - 准对称离散信道
    - 一般的单符号离散信道
    - 单符号离散信道的串联信道
  - 多符号离散信道
    - 离散无记忆信道
    - 一般的多符号离散信道
    - 单符号离散信道的N个独立并联信道

图 3.1　离散信道的分类

按照信道输出和输入之间的依赖关系，多符号离散信道可以划分为无记忆信道和有记忆信道。无记忆信道是指信道的输出只与信道该时刻的输入有关，而与其他时刻的输入和输出无关。有记忆信道是指信道某一时刻的输出不仅与该时刻的输入有关，还与前面时刻的输入和输出有关。

实际中常常会遇到两个或更多个信道组合在一起使用的情况。例如，有时消息会依次地通过几个信道串联发送，如微波中继接力通信、数据处理系统等，这种组合信道称为串联信道；待发送的消息比较多时，可能要使用两个或更多个信道并行发送，这种组合信道

称为并联信道。本章中串联信道和并联信道将分别归入单符号离散信道和多符号离散信道进行分析。

3.2 离散信道的数学模型

3.2.1 单符号离散信道的数学模型

如果信道的输入、输出都取值于离散符号集，且都用一个随机变量来表示，此时的信道称为**单符号离散信道**，如图 3.2 所示。它是最简单的离散信道，可用**信道的概率空间** $\{X,P(b_j\,|\,a_i),Y\}$ 来描述。

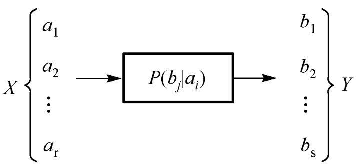

图 3.2 单符号离散信道

设单符号离散信道的输入随机变量为 X，其可能的取值 $x\in X=\{a_1,a_2,\cdots,a_r\}$，输出随机变量为 Y，其可能的取值 $y\in Y=\{b_1,b_2,\cdots,b_s\}$，其中 r 和 s 可以相等，也可以不相等。由于信道中存在噪声干扰，因此输入符号在传输中会产生错误，这种信道干扰对传输信号的影响可用条件概率 $P(y\,|\,x)=P(b_j\,|\,a_i)$ $(i=1,2,\cdots,r;j=1,2,\cdots,s)$ 来描述。这个条件概率集中体现了信道对输入符号 a_i $(i=1,2,\cdots,r)$ 的传递作用。不同的信道，就有不同的条件概率。因此，条件概率 $P(b_j\,|\,a_i)$ $(i=1,2,\cdots,r;j=1,2,\cdots,s)$ 称为**信道的传递概率**或**转移概率**。

按信道的输入输出符号的对应关系，把 $(r\times s)$ 个条件概率 $P_{ij}=P(b_j\,|\,a_i)$ 排列成一个 $(r\times s)$ 阶矩阵，即

$$\boldsymbol{P}=\begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_r \end{array}\begin{array}{c} \begin{array}{cccc} \quad b_1 \quad & \quad b_2 \quad & \cdots & \quad b_s \quad \end{array} \\ \begin{bmatrix} P(b_1\,|\,a_1) & P(b_2\,|\,a_1) & \cdots & P(b_s\,|\,a_1) \\ P(b_1\,|\,a_2) & P(b_2\,|\,a_2) & \cdots & P(b_s\,|\,a_2) \\ \vdots & \vdots & & \vdots \\ P(b_1\,|\,a_r) & P(b_2\,|\,a_r) & \cdots & P(b_s\,|\,a_r) \end{bmatrix} \end{array} \tag{3.2.1}$$

简记为

$$\boldsymbol{P}=\begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1s} \\ P_{21} & P_{22} & \cdots & P_{2s} \\ \vdots & \vdots & & \vdots \\ P_{r1} & P_{r2} & \cdots & P_{rs} \end{bmatrix} \tag{3.2.2}$$

矩阵 $\boldsymbol{P}$ 完整地描述了单符号离散信道的传递特性，所以把矩阵 $\boldsymbol{P}$ 称为单符号离散信道的信道传递概率矩阵，或者信道转移概率矩阵，简称**信道矩阵**。式（3.2.1）中的传递概率满足

（1） $0\leqslant P(b_j\,|\,a_i)\leqslant 1 \qquad (i=1,2,\cdots,r;\ \ j=1,2,\cdots,s)$ （3.2.3）

（2） $\sum\limits_{j=1}^{s}P(b_j\,|\,a_i)=1 \qquad (i=1,2,\cdots,r)$ （3.2.4）

式（3.2.3）表明信道矩阵中每个元素均非负。当 $P(b_j\,|\,a_i)=0$ 时，表示在输入符号 a_i 的前提下，信道不可能输出 b_j；当 $P(b_j\,|\,a_i)=1$ 时，表示在输入符号 a_i 的前提下，信道输出 b_j 是一个确定事件。

式（3.2.4）表明信道矩阵中每一行之和必等于 1。这是因为在已知信道输入某符号 $a_i\ (i=1,2,\cdots,r)$ 的前提下，由于噪声的随机干扰，信道输出哪一种符号虽然是不确定的，但一定是信道输出符号集 $Y:\{b_1,b_2,\cdots,b_s\}$ 中的某一种符号，绝不可能是符号集 $Y:\{b_1,b_2,\cdots,b_s\}$ 以外的任何其他符号。

信道传递特性可用图 3.3 来描述。图中左右两侧的点集合分别表示输入符号集 $X:\{a_1,a_2,\cdots,a_r\}$ 和输出符号集 $Y:\{b_1,b_2,\cdots,b_s\}$，由 a_i 到 b_j 的连线旁的数值表示信道输入 a_i 到 b_j 的传递概率 $P(b_j\,|\,a_i)\ (i=1,2,\cdots,r;\ j=1,2,\cdots,s)$。从每一个输入符号 a_i 出发的所有连线旁标出的数值之和均等于 1。

下面介绍三种重要的信道：无噪信道、二元对称信道（BSC）和二元删除信道（BEC）。

1．无噪信道

信道特点：离散无噪信道的输入和输出符号之间存在一一对应的关系。如图 3.4 所示。

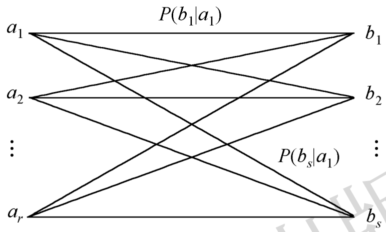

图 3.3　信道传递特性

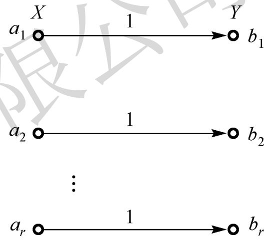

图 3.4　无噪信道

假设信道输入符号个数为 4，则信道转移概率为

$$P(b_j\,|\,a_i)=\begin{cases}0 & i\neq j\\ 1 & i=j\end{cases}\qquad (i,j=1,2,3,4)$$

对应的信道矩阵为

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

可见，无噪信道的信道转移概率矩阵为单位阵。

2．二元对称信道（BSC，Binary Symmetric Channel）

这是很重要的一种特殊信道，如图 3.5 所示。其输入输出符号集均取值于 $\{0,1\}$。此时 $r=s=2$，而且 $a_1=b_1=0, a_2=b_2=1$。转移概率为

$$P(b_1\,|\,a_1)=P(0\,|\,0)=1-p=\overline{p}$$
$$P(b_2\,|\,a_2)=P(1\,|\,1)=1-p=\overline{p}$$
$$P(b_1\,|\,a_2)=P(0\,|\,1)=p$$
$$P(b_2\,|\,a_1)=P(1\,|\,0)=p$$

通常概率 $1-p$ 用 $\overline{p}$ 表示。于是，可得 BSC 的信道转移概率矩阵为

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

3．二元删除信道（BEC，Binary Erasure Channel）

二元删除信道如图 3.6 所示，其中删除概率为ε。因为它的信道转移图类似于字母 M，所以二元删除信道也称为 M 信道。这时$r=2,s=3$。输入符号X取值于$\{0,1\}$，输出符号Y取值于$\{0,\mathrm{e},1\}$，其中 e 为删除符号。该信道以概率$1-\varepsilon$正确传输符号，以概率ε删除符号，信道矩阵为

$$\boldsymbol{P}=\begin{array}{c} \\ 0 \\ 1 \end{array}\begin{array}{c} \begin{array}{ccc} 0 & \mathrm{e} & 1 \end{array} \\ \begin{bmatrix} 1-\varepsilon & \varepsilon & 0 \\ 0 & \varepsilon & 1-\varepsilon \end{bmatrix} \end{array}$$

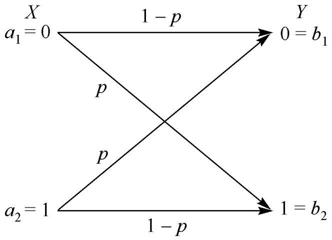

图 3.5　二元对称信道

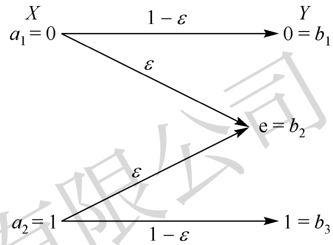

图 3.6　二元删除信道

在传输过程中，发送端发送的 0 或 1，接收端以概率$1-\varepsilon$正确接收为 0 或 1，同时以概率ε将其识别为“删除符号”，即接收端无法确定发送的是 0 还是 1。这种信道实际上是存在的，当信号波形传输中失真较大时，在接收端不是对接收信号硬性地判为 0 或 1，而是增加一个中间状态“e”（称为删除符号），接收端见“e”就删除。

下面推导单符号离散信道的一些概率关系。

假设信道矩阵如式（3.2.1），信道输入符号概率空间为

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

信道的输出符号集$Y=\{b_1,b_2,\cdots,b_s\}$。

（1）输入输出随机变量的联合概率

$$P(a_ib_j)=P(a_i)P(b_j\,|\,a_i)=P(b_j)P(a_i\,|\,b_j) \tag{3.2.5}$$

其中，$P(b_j\,|\,a_i)$是信道传递概率，$P(a_i)$称为输入符号的先验概率，$P(a_i\,|\,b_j)$为输入符号的后验概率。

（2）输出符号的概率

$$P(b_j)=\sum_{i=1}^{r}P(a_i)P(b_j\,|\,a_i),\quad (j=1,2,\cdots,s) \tag{3.2.6}$$

也可写成矩阵形式，即

$$\begin{bmatrix} P(b_1) \\ P(b_2) \\ \vdots \\ P(b_s) \end{bmatrix}=\boldsymbol{P}^{\mathrm{T}}\begin{bmatrix} P(a_1) \\ P(a_2) \\ \vdots \\ P(a_r) \end{bmatrix} \tag{3.2.7}$$

式中，$\boldsymbol{P}^{\mathrm{T}}$ 为信道矩阵 $\boldsymbol{P}$ 的转置矩阵。

（3）后验概率

根据贝叶斯公式，由先验概率和信道转移概率，可得后验概率为

$$\begin{aligned} P(a_i \mid b_j) &= \frac{P(a_i b_j)}{P(b_j)}, \quad P(b_j) \neq 0 \\ &= \frac{P(a_i)P(b_j \mid a_i)}{\sum_{i=1}^{r} P(a_i)P(b_j \mid a_i)}, \quad (i=1,2,\cdots,r; \quad j=1,2,\cdots,s) \end{aligned} \tag{3.2.8}$$

3.2.2　多符号离散信道的数学模型

一般离散信道为多符号离散信道，即信道的输入和输出是时间离散、取值有限的随机矢量。单符号离散信道可以看作多符号离散信道的特例。

1．多符号离散信道

多符号离散信道的数学模型如图 3.7 所示，其中信道输入 $\boldsymbol{X}=(X_1X_2\cdots X_N)$，信道输出 $\boldsymbol{Y}=(Y_1Y_2\cdots Y_N)$。由于信道输入的随机矢量 $\boldsymbol{X}$ 和输出的随机矢量 $\boldsymbol{Y}$ 往往不是确定关系，信道特性通常采用数学符号 $\{\boldsymbol{X}, P(\boldsymbol{y} \mid \boldsymbol{x}), \boldsymbol{Y}\}$ 来表示。

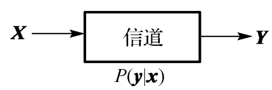

图 3.7　多符号离散信道的数学模型

设多符号离散信道的输入矢量 $\boldsymbol{X}=(X_1X_2\cdots X_N)$，每一时刻的随机变量 $X_l(l=1,2,\cdots,N)$ 均取自且取遍于信道的输入符号集 $X:\{a_1,a_2,\cdots,a_r\}$，则 $\boldsymbol{X}=(X_1X_2\cdots X_N)$ 共有 r^N 种不同的消息，某一具体消息 $\boldsymbol{x}$ 可表示为

$$\boldsymbol{x}=\boldsymbol{\alpha}_k=(x_1x_2\cdots x_N) \qquad x_l \in \{a_1,a_2,\cdots,a_r\} \tag{3.2.9}$$

式中，l 表示随机矢量中某分量的序号，$l=1,2,\cdots,N$；k 表示序列的序号，$k=1,2,\cdots,r^N$。

对应地，在信道的输出端为一个 N 维随机矢量 $\boldsymbol{Y}=(Y_1Y_2\cdots Y_N)$，共有 s^N 种不同的消息，其中某一具体消息 $\boldsymbol{y}$ 可表示为

$$\begin{gathered} \boldsymbol{y}=\boldsymbol{\beta}_h=(y_1y_2\cdots y_N) \qquad y_l \in \{b_1,b_2,\cdots,b_s\} \\ (l=1,\cdots,N) \qquad (h=1,\cdots,s^N) \end{gathered} \tag{3.2.10}$$

可见，与单符号离散信道相比，多符号离散信道的输入符号数由 r 个扩展为 r^N 个，输出符号数由 s 个扩展为 s^N 个。在输入消息 $\boldsymbol{\alpha}_k(k=1,\cdots,r^N)$ 的条件下，输出消息 $\boldsymbol{\beta}_h(h=1,\cdots,s^N)$ 的转移概率为

$$P(\boldsymbol{\beta}_h|\boldsymbol{\alpha}_k)=P(y_1y_2\cdots y_N|x_1x_2\cdots x_N) \ (k=1,\cdots,r^N; h=1,\cdots,s^N) \tag{3.2.11}$$

类似地，把这 $r^N \cdot s^N$ 个转移概率按输入、输出的对应关系，可以构成多符号离散信道的信道矩阵

$$\boldsymbol{P}= \begin{array}{c} \\ \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_{r^N} \end{array} \begin{array}{c} \begin{array}{cccc} \quad\boldsymbol{\beta}_1 \quad & \quad\boldsymbol{\beta}_2 \quad & \cdots & \quad\boldsymbol{\beta}_{s^N} \end{array} \\ \begin{bmatrix} P(\boldsymbol{\beta}_1|\boldsymbol{\alpha}_1) & P(\boldsymbol{\beta}_2|\boldsymbol{\alpha}_1) & \cdots & P(\boldsymbol{\beta}_{s^N}|\boldsymbol{\alpha}_1) \\ P(\boldsymbol{\beta}_1|\boldsymbol{\alpha}_2) & P(\boldsymbol{\beta}_2|\boldsymbol{\alpha}_2) & \cdots & P(\boldsymbol{\beta}_{s^N}|\boldsymbol{\alpha}_2) \\ \vdots & \vdots & & \vdots \\ P(\boldsymbol{\beta}_1|\boldsymbol{\alpha}_{r^N}) & P(\boldsymbol{\beta}_2|\boldsymbol{\alpha}_{r^N}) & \cdots & P(\boldsymbol{\beta}_{s^N}|\boldsymbol{\alpha}_{r^N}) \end{bmatrix} \end{array} \tag{3.2.12}$$

且满足
$$\sum_{h=1}^{s^N} P(\boldsymbol{\beta}_h|\boldsymbol{\alpha}_k)=1 \qquad (k=1,2,\cdots,r^N) \tag{3.2.13}$$

2．离散无记忆信道的 N 次扩展信道

如果离散信道的输出只与信道该时刻的输入有关，而与其他时刻的输入无关，即多符号离散信道的转移概率等于 N 个时刻单符号离散信道的转移概率的乘积， 即

$$P(\boldsymbol{y}|\boldsymbol{x})=P(\boldsymbol{\beta}_h\,|\,\boldsymbol{\alpha}_k)=P(y_1y_2\cdots y_N|x_1x_2\cdots x_N)=\prod_{l=1}^{N}P(y_l|x_l) \tag{3.2.14}$$

则该多符号离散信道称为**离散无记忆信道**，简记为 DMC。类似于离散无记忆信源的 N 次扩展信源的定义，当离散无记忆信道的输入矢量和输出矢量均为 N 维矢量时，称此时的离散无记忆信道为离散无记忆的 N 次扩展信道。

由式（3.2.14）可见，离散无记忆信道的转移概率 $P(\boldsymbol{\beta}_h|\boldsymbol{\alpha}_k)$ $(k=1,\cdots,r^N;\,h=1,\cdots,s^N)$ 可由单符号离散信道的转移概率 $P(y_l=b_j\,|\,x_l=a_i)(i=1,2,\cdots,r;\,j=1,2,\cdots,s)$ 直接求得。无疑，这会给离散无记忆的 N 次扩展信道的分析带来方便。

【例 3.1】 求图 3.5 所示的二元无记忆离散对称信道的二次扩展信道的信道矩阵。

解： 因为二元对称信道的输入和输出变量 X 和 Y 的取值都是 0 和 1，因此，二次扩展信道的输入符号集为 $\boldsymbol{X}=\{00,01,10,11\}$，共有 $2^2=4$ 个符号。输出符号集为 $\boldsymbol{Y}=\{00,01,10,11\}$，共 4 个符号。根据无记忆信道的特性，求得二次扩展信道的转移概率为

$$P(\boldsymbol{\beta}_1|\boldsymbol{\alpha}_1)=P(00|00)=P(0|0)P(0|0)=\bar{p}^2$$

$$P(\boldsymbol{\beta}_2|\boldsymbol{\alpha}_1)=P(01|00)=P(0|0)P(1|0)=\bar{p}p$$

$$P(\boldsymbol{\beta}_3|\boldsymbol{\alpha}_1)=P(10|00)=P(1|0)P(0|0)=p\bar{p}$$

$$P(\boldsymbol{\beta}_4|\boldsymbol{\alpha}_1)=P(11|00)=P(1|0)P(1|0)=p^2$$

同理，可求得其他转移概率 $P(\beta_h|\alpha_k)$，因此，二次扩展信道的信道矩阵为

$$\boldsymbol{P}_{1,2}=\begin{array}{c} \\ \boldsymbol{\alpha}_1\\ \boldsymbol{\alpha}_2\\ \boldsymbol{\alpha}_3\\ \boldsymbol{\alpha}_4 \end{array}\begin{array}{c} \begin{array}{cccc}\boldsymbol{\beta}_1 & \boldsymbol{\beta}_2 & \boldsymbol{\beta}_3 & \boldsymbol{\beta}_4\end{array}\\ \begin{bmatrix} \bar{p}^2 & \bar{p}p & p\bar{p} & p^2\\ \bar{p}p & \bar{p}^2 & p^2 & p\bar{p}\\ p\bar{p} & p^2 & \bar{p}^2 & \bar{p}p\\ p^2 & p\bar{p} & \bar{p}p & \bar{p}^2 \end{bmatrix}\end{array} \tag{3.2.15}$$

该二次扩展信道可用图 3.8 表示。

实际信道通常为有记忆信道，即信道某一时刻的输出不仅与该时刻的输入有关，还与前面时刻的输入和输出有关。处理这类信道的常见方法之一是把记忆较强的 N 个符号作为一个矢量符号来处理，而假定各个矢量符号之间是无记忆的，这种处理方法会引入误差，通常 N 越大，误差越小。另一种处理方法是将 $P(\boldsymbol{y}|\boldsymbol{x})$ 看作马尔可夫链，即把信道某时刻的输入和输出序列看作信道的状态，那么信道的统计特性可用 $P(y_ls_l\,|\,x_ls_{l-1})$ 来描述，其中

$P(y_l s_l | x_l s_{l-1})$ 表示已知现时刻的输入符号和前一时刻信道所处状态的条件下，信道的输出符号和所处状态的联合条件概率。

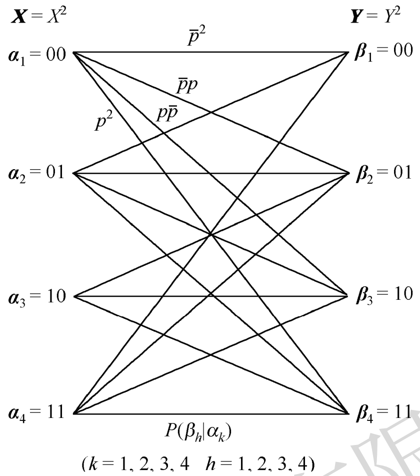

图 3.8　二元对称信道的二次扩展信道

一般情况下，有记忆信道的研究很复杂，本章将侧重研究离散无记忆信道。如前所述，对于离散平稳无记忆信道，可以归结为对单符号离散信道的研究。

3.3　离散随机变量的互信息和平均互信息

绪论中简单介绍的互信息是对消息符号之间相互提供信息量进行度量，本节将引入平均互信息，并深入讨论互信息和平均互信息的性质。

3.3.1　互信息的定义

定义 3.1　对于两个离散随机事件集 X 和 Y，事件 b_j 的出现给出关于事件 a_i 的信息量，定义为事件 a_i 和事件 b_j 的**互信息**，用 $I(a_i;b_j)$ 表示。

$$I(a_i;b_j)=I(a_i)-I(a_i|b_j)=\log\frac{P(a_i|b_j)}{P(a_i)}\quad (i=1,2,\cdots,r;j=1,2,\cdots,s) \tag{3.3.1}$$

由式（3.3.1）可见，互信息 $I(a_i;b_j)$ 是已知事件 b_j 后所消除的关于事件 a_i 的不确定性，即关于事件 a_i 不确定性的减少量。

如果信源发出的符号消息为 a_i，经过信道传输后信宿收到的符号消息为 b_j。设 $P(a_i)$ 为先验概率，$P(a_i|b_j)$ 为后验概率，由式（3.3.1）有

$$互信息=\log\frac{后验概率}{先验概率}=\log\frac{1}{P(a_i)}-\log\frac{1}{P(a_i|b_j)} \tag{3.3.2}$$

那么互信息的物理含义可以理解为：信宿收到 b_j 后获得的关于信源输出 a_i 的信息量就是先验不确定性 $I(a_i)$ 减去后验不确定性 $I(a_i|b_j)$，实际上就是信道传递的信息量。互信息的引入，使信息的传递得到了定量的表示，是信息论发展的一个重要里程碑。

3.3.2 互信息的性质

（1）互信息具有对称性，即

$$I(a_i;b_j)=I(b_j;a_i) \tag{3.3.3}$$

证明：由式（3.3.1）有

$$\begin{aligned}I(a_i;b_j)&=\log\frac{P(a_i|b_j)}{P(a_i)}=\log\frac{P(a_i|b_j)P(b_j)}{P(a_i)P(b_j)}=\log\frac{P(a_ib_j)/P(a_i)}{P(b_j)}\\&=\log\frac{P(b_j|a_i)}{P(b_j)}=I(b_j;a_i)\qquad(i=1,2,\cdots,r;j=1,2,\cdots,s)\end{aligned}$$

[证毕]

互信息的对称性说明：从 b_j 得到的关于 a_i 的信息量 $I(a_i;b_j)$ 与从 a_i 得到的关于 b_j 的信息量 $I(b_j;a_i)$ 是一样的。

（2）当相互独立时，互信息为 0。

如果 a_i 和 b_j 相互独立，则 $P(a_ib_j)=P(a_i)P(b_j)$。此时互信息为

$$I(a_i;b_j)=\log\frac{P(a_ib_j)}{P(a_i)P(b_j)}=\log1=0\qquad(i=1,2,\cdots,r;j=1,2,\cdots,s) \tag{3.3.4}$$

这表明如果 a_i 与 b_j 之间相互独立时，从 b_j 得不到关于 a_i 的任何信息，反之亦然。

（3）互信息不可能大于自信息。

由于 $P(a_i|b_j)\leqslant1$，根据式（3.3.1），有

$$I(a_i;b_j)\leqslant\log\frac{1}{P(a_i)}=I(a_i) \tag{3.3.5}$$

同理，由于 $P(b_j|a_i)\leqslant1$，有

$$I(b_j;a_i)\leqslant\log\frac{1}{P(b_j)}=I(b_j)$$

可见，某一事件的自信息是任何其他事件所能提供的关于该事件的最大信息量。

（4）互信息可为正值或负值。

由式（3.3.1）可知，当 $P(a_i|b_j)>P(a_i)$ 时，互信息为正值。反之，当 $P(a_i|b_j)<P(a_i)$ 时，互信息为负值。当后验概率与先验概率相等时，互信息为 0，这就是两个随机事件相互独立的情况。

互信息 $I(a_i;b_j)$ 为正值，意味着事件 b_j 的出现有利于肯定事件 a_i 的发生；反之，则是不利的。

在实际工作和生活中，如果不能够直接得到某事件的信息，往往通过其他事件来获得该事件的信息。下面以实际生活中的一个简单例子来理解互信息的概念。

【例 3.2】 假设 a_1 表示“下雨”，a_2 表示“无雨”；b_1 表示“空中有乌云”，b_2 表示“没有乌云”，得到的调查结果为 $P(a_1)=0.1$，$P(a_1|b_1)=0.8$。计算：

（1）事件“下雨”的自信息；
（2）在“空中有乌云”条件下，事件“下雨”的自信息；
（3）“下雨”和“空中有乌云”的互信息；
（4）“无雨”和“空中有乌云”的互信息。

解：（1）$I(a_1)=-\log 0.1=\log 10=3.32$（bit）

（2）$I(a_1|b_1)=-\log 0.8=0.32$（bit）

（3）$I(a_1;b_1)=I(a_1)-I(a_1|b_1)=3$（bit）

（4）$P(a_2)=0.9$，$P(a_2|b_1)=0.2$

$$I(a_2;b_1)=I(a_2)-I(a_2|b_1)=-\log 0.9-(-\log 0.2)=\log\frac{0.2}{0.9}=-2.17\text{（bit）}$$

因为$I(a_1;b_1)>0$，根据互信息的物理含义可知，“空中有乌云”有助于肯定“下雨”。由于互信息的对称性，“下雨”有助于肯定“空中有乌云”。

3.3.3　平均互信息的定义

$I(a_i;b_j)$表征事件a_i和事件b_j之间的互信息，是一个随机变量。为了从总体上度量随机变量X和Y之间相互提供信息量的多少，下面介绍平均互信息的概念。

定义 3.2　集合X和集合Y之间的**平均互信息**（又称为**互熵**）定义为

$$I(X;Y)=\sum_{i=1}^{r}\sum_{j=1}^{s}P(a_ib_j)I(a_i;b_j)=\sum_{i=1}^{r}\sum_{j=1}^{s}P(a_ib_j)\log\frac{P(a_i|b_j)}{P(a_i)} \tag{3.3.6}$$

可见，平均互信息$I(X;Y)$是互信息$I(a_i;b_j)$在X和Y的联合概率空间中的统计平均值，它是从整体上表示一个随机变量Y所给出的关于另一个随机变量X的平均信息量。

由式（3.3.6）可得

$$\begin{aligned}I(X;Y)&=-\sum_{i=1}^{r}\sum_{j=1}^{s}P(a_ib_j)\log P(a_i)-\{-\sum_{i=1}^{r}\sum_{j=1}^{s}P(a_ib_j)\log P(a_i|b_j)\}\\&=-\sum_{i=1}^{r}P(a_i)\log P(a_i)\sum_{j=1}^{s}P(b_j|a_i)-\sum_{j=1}^{s}P(b_j)\{-\sum_{i=1}^{r}P(a_i|b_j)\log P(a_i|b_j)\}\\&=-\sum_{i=1}^{r}P(a_i)\log P(a_i)-\sum_{j=1}^{s}P(b_j)H(X|Y=b_j)\\&=H(X)-H(X|Y)\end{aligned} \tag{3.3.7}$$

式（3.3.7）表明，从Y中获取关于X的平均互信息$I(X;Y)$，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的平均不确定性的减少量。平均不确定性的消除量就是接收端所获得的平均信息量。

因为$H(X|Y)=H(XY)-H(Y)$，由式（3.3.7）容易得到

$$I(X;Y)=H(X)+H(Y)-H(XY)=H(Y)-H(Y|X) \tag{3.3.8}$$

当X表示信道输入、Y表示信道输出时，条件熵$H(X|Y)$表示信宿在收到Y后仍然对信源X存在的不确定性，通常称$H(X|Y)$为**信道疑义度**，也称为**损失熵**。条件熵$H(Y|X)$表示在已知输入变量X的条件下，对随机变量Y尚存在的不确定性，通常称条件熵$H(Y|X)$为

噪声熵或散布度。

【**例 3.3**】 二进制通信系统等概使用符号 0 和 1，由于存在失真，传输时会产生误码，无记忆信道模型如图 3.9 所示。试计算噪声熵、信道疑义度和平均互信息。

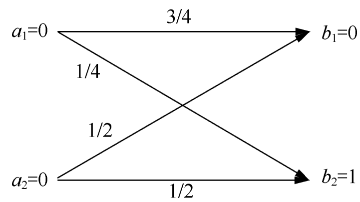

图 3.9　无记忆信道模型

解：信源熵为

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

噪声熵为

$$H(Y|X)=\frac{1}{2}H\left(\frac{1}{4},\frac{3}{4}\right)+\frac{1}{2}H\left(\frac{1}{2},\frac{1}{2}\right)=0.9056 \text{（比特/符号）}$$

信道疑义度为

$$H(X|Y)=H(XY)-H(Y)=H(X)+H(Y|X)-H\left(\frac{3}{8},\frac{5}{8}\right)=0.9512 \text{（比特/符号）}$$

平均互信息为

$$I(X;Y)=H(X)-H(X|Y)=0.0488 \text{（比特/符号）}$$

3.3.4　平均互信息的性质

（1）平均互信息具有对称性，即

$$I(X;Y)=I(Y;X) \tag{3.3.9}$$

证明：由于 $P(a_ib_j)=P(b_ja_i)$，得

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

[证毕]

$I(X;Y)$ 表示从 Y 中提取的关于 X 的信息量，而 $I(Y;X)$ 表示从 X 中提取的关于 Y 的信息量，它们是相等的。

（2）平均互信息具有非负性，即

$$I(X;Y)\geqslant 0 \tag{3.3.10}$$

当且仅当 X 和 Y 统计独立时，等式成立。该性质可利用詹森不等式证明得到，请参见附录 A。

如果离散信道输入符号为 X，输出符号为 Y，则该性质说明：通过一个信道获得的平

均信息量不会是负值，一般总能获得一些信息量。也就是说，观察一个信道的输出，从平均的角度来看总能消除一些不确定性，接收到一定的信息。

特别地，在信道输入 X 和输出 Y 是统计独立时，就不可能从一个随机变量获得关于另一个随机变量的信息，所以

$$I(X;Y)=I(Y;X)=0 \tag{3.3.11}$$

如果通过输入的随机变量 X 不能得到任何关于输出 Y 的信息，则此时的信道称为**无用信道**。此时信道的转移概率 $P(b_j\,|\,a_i)=P(b_j)$ 与 a_i 无关，即信道矩阵的各行是一样的，此时

$$H(Y\,|\,X)=H(Y)\,,\quad I(X;Y)=I(Y;X)=0$$

（3）平均互信息具有极值性，即

$$I(X;Y)\leqslant H(X) \tag{3.3.12}$$

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

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

因此有

$$H(X\,|\,Y)\geqslant 0 \tag{3.3.13}$$

所以

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

[证毕]

该性质说明：接收者通过信道获得的信息量不可能超过信源本身固有的信息量。只有当 $H(X\,|\,Y)=0$，即信道中传输信息无损失时，接收到 Y 后获得关于 X 的信息量才等于符号集 X 中平均每个符号所含有的信息量。可见，在一般情况下，平均互信息必在 0 和 $H(X)$ 值之间。

特别地，当两个随机变量 X 和 Y 一一对应时，从一个变量就可以充分获得关于另一个变量的信息，即

$$I(X;Y)=I(Y;X)=H(X)=H(Y) \tag{3.3.14}$$

（4）平均互信息 $I(X;Y)$ 具有凸函数性

因为

$$I(X;Y)=\sum_{i=1}^{r}\sum_{j=1}^{s}P(a_i)P(b_j\,|\,a_i)\log\frac{P(b_j\,|\,a_i)}{P(b_j)}$$

而且

$$P(b_j)=\sum_{i=1}^{r}P(a_i)P(b_j\,|\,a_i)\quad（对\ j=1,2,\cdots,s 都成立）$$

所以，平均互信息 $I(X;Y)$ 是条件概率 $P(b_j\,|\,a_i)(i=1,2,\cdots,r;\,j=1,2,\cdots,s)$ 和概率分布 $P(a_i)(i=1,2,\cdots,r)$ 的函数。即

$$I(X;Y)=I[P(a_i),P(b_j\,|\,a_i)]=I[P(x),P(y\,|\,x)] \tag{3.3.15}$$

式（3.3.15）第二个等式右边是简明写法。通常在通信系统中，$P(x)$ 表示信源 X 的先验概率。

$P(y|x)$表示信道传递概率，即信宿收到的信息量由信源的先验概率和信道的传输特性决定，对不同信源和不同信道得到的平均互信息是不同的。

定理 3.1 在$P(y|x)$给定的条件下，平均互信息$I(X;Y)$是概率分布$P(x)$的$\cap$形凸函数（或称上凸函数）。即

$$I[\theta P_1(x)+\bar{\theta}P_2(x)]\geqslant \theta I[P_1(x)]+\bar{\theta}I[P_2(x)] \tag{3.3.16}$$

定理 3.1 表明：当固定某信道时，选择不同概率分布的信源与信道连接，在信道输出端接收到每个符号后获得的信息量是信源概率分布$P(x)$的$\cap$形凸函数。因此对于每一个固定信道，一定存在某种概率分布的信源，使输出端获得的平均信息量最大。可见信源与信道连接时，信道传输信息的能力不一定能得到充分利用。对于固定的信道，平均互信息的最大值是一定的，但是在传输信息时信道能否提供其最大传输能力，则取决于输入端的概率分布。当信道提供其最大传输能力时，相应的输入概率分布称为最佳输入分布，此时的信源称为**匹配信源**。

定理 3.2 在概率分布$P(x)$给定的条件下，平均互信息$I(X;Y)$是条件概率$P(y|x)$的$\cup$形凸函数（或称下凸函数）。即

$$I[\theta P_1(y|x)+\bar{\theta}P_2(y|x)]\leqslant \theta I[P_1(y|x)]+\bar{\theta}I[P_2(y|x)] \tag{3.3.17}$$

定理 3.2 表明：当信源固定后，选择不同信道来传输同一信源符号时，在信道的输出端获得关于信源的信息量是不同的。信道输出端获得关于信源的信息量是信道转移概率$P(y|x)$的$\cup$形凸函数。也就是说，存在一个信道使某一特定信源经过此信道传输时信道的平均互信息达到极小值。

定理 3.1 和定理 3.2 可利用詹森不等式证明得到，请参见附录 A。定理 3.1 和定理 3.2 分别是信道容量和信息率失真函数的理论基础。

【例 3.4】 设二元对称信道的输入概率空间$\begin{bmatrix} X \\ P(x) \end{bmatrix}=\begin{bmatrix} 0 & 1 \\ \omega & \bar{\omega}=1-\omega \end{bmatrix}$，而信道特性为图 3.5 所示的二元对称信道，求平均互信息。

解： 信道输出Y的概率分布为

$$P(y=0)=\omega\bar{p}+(1-\omega)p=\omega\bar{p}+\bar{\omega}p$$
$$P(y=1)=1-P(y=0)$$

可得

$$H(Y)=H(\omega\bar{p}+\bar{\omega}p)$$

又因为

$$H(Y|X)=H(p,\bar{p})=H(p)$$

所以，平均互信息为

$$I(X;Y)=H(Y)-H(Y|X)=H(\omega\bar{p}+\bar{\omega}p)-H(p) \tag{3.3.18}$$

可见，当信道固定（即固定p）时，可得$I(X;Y)$是ω的$\cap$形凸函数，其曲线如图 3.10 所示。从图中可知，当二元对称信道的信道矩阵固定后，输入变量X的概率分布不同，在接收端平均每个符号获得的信息量就不同。只有当输入变量X是等概分布时，即$P(x=0)=P(x=1)=1/2$时，平均互信息才能取得最大值。

当固定信源的概率分布为ω时，即得$I(X;Y)$是p的$\cup$形凸函数，如图 3.11 所示。从图

中可知，当 $p=0$ 时，平均互信息为最大，等于信源熵，此时的信道为无噪信道；当 $p=1/2$ 时，平均互信息等于零，也就是说，信源的信息全部损失在信道中，此时的信道为无用信道。

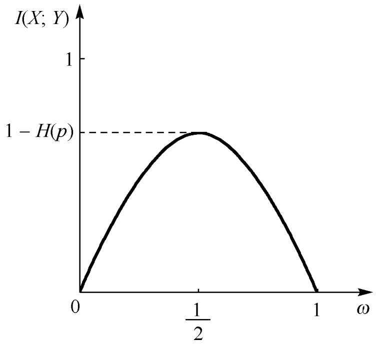

图 3.10　固定二元对称信道的平均互信息

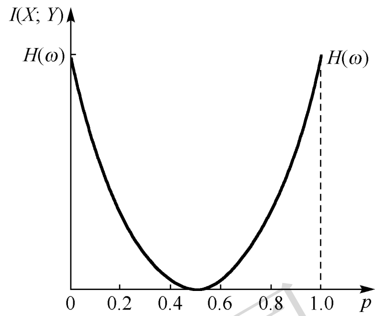

图 3.11　固定二元信源的平均互信息

3.3.5　平均互信息与各类熵之间的关系

如前所述，容易得到平均互信息和各类熵之间的关系。现总结如下：

（1）$I(X;Y)=H(X)-H(X|Y)=H(Y)-H(Y|X)=H(X)+H(Y)-H(XY)$

（2）$H(XY)=H(X)+H(Y|X)=H(Y)+H(X|Y)$

（3）$H(X)\geqslant H(X|Y)$；$H(Y)\geqslant H(Y|X)$

（4）当 X 和 Y 相互独立时，有：$I(X;Y)=0$；$H(X)=H(X|Y)$；$H(Y)=H(Y|X)$；$H(XY)=H(X)+H(Y)$。

用图 3.12 所示的维拉图可得到这些关系的清晰表示。图中，左边的圆代表随机变量 X 的熵，右边的圆代表随机变量 Y 的熵，两个圆的重叠部分是平均互信息 $I(X;Y)$。每个圆减去平均互信息后剩余的部分代表两个条件熵。联合熵 $H(XY)$ 是联合空间 XY 的熵，所以联合熵是两个圆之和再减去重叠部分。特殊地，当 X 和 Y 相互独立时，两圆不重叠，此时 $I(X;Y)=0$。

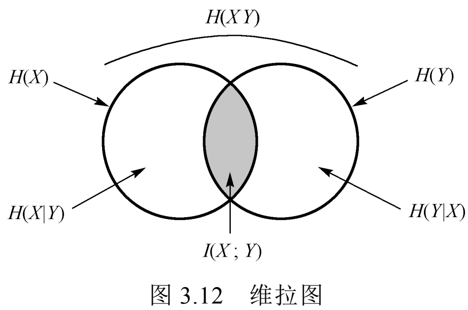

图 3.12　维拉图

【例 3.5】　已知两个独立的随机变量 X、Y 的分布律如下。

$$\begin{bmatrix} X \\ P(x) \end{bmatrix}=\begin{bmatrix} a_1 & a_2 \\ 0.5 & 0.5 \end{bmatrix},\quad \begin{bmatrix} Y \\ P(y) \end{bmatrix}=\begin{bmatrix} b_1 & b_2 & b_3 \\ 0.25 & 0.25 & 0.5 \end{bmatrix}$$

计算 $H(X), H(Y), H(XY), H(X|Y), H(Y|X), I(X;Y)$。

解：

$H(X)=H(0.5,0.5)=1$（比特/符号）

$H(Y)=H(0.25,0.25,0.5)=1.5$（比特/符号）

$H(X|Y)=H(X)=1$（比特/符号）

$H(Y|X)=H(Y)=1.5$（比特/符号）

$H(XY)=H(X)+H(Y)=2.5$（比特/符号对）

$I(X;\ Y)=0$

3.3.6 平均联合互信息和平均条件互信息

类似互信息的定义，联合互信息和条件互信息的定义如下。

定义 3.3 设联合集 XYZ，随机事件 $a_i\in X$，$b_j\in Y$，$c_k\in Z$。**联合互信息** $I(a_i;b_jc_k)$ 定义为

$$I(a_i;b_jc_k)=\log\frac{P(a_i|b_jc_k)}{P(a_i)} \tag{3.3.19}$$

$I(a_i;b_jc_k)$ 的物理含义为：已知 b_jc_k 后对事件 a_i 的不确定性的减少量，即已知 b_jc_k 后获得的关于 a_i 的信息量。

容易理解：

$$I(a_i;b_jc_k)=I(a_i)-I(a_i|b_jc_k) \tag{3.3.20}$$

定义 3.4 设联合集 XYZ，随机事件 $a_i\in X$，$b_j\in Y$，$c_k\in Z$。在给定 c_k 条件下，a_i 与 b_j 之间的**条件互信息**定义为

$$I(a_i;b_j|c_k)=\log\frac{P(a_i|b_jc_k)}{P(a_i|c_k)} \tag{3.3.21}$$

$I(a_i;b_j|c_k)$ 的物理含义为：在给定 c_k 条件下，b_j 发生后获得关于 a_i 的信息量。需要注意的是，c_k 不仅是 a_i 的条件，还是 b_j 的条件。

容易理解：

$$I(a_i;b_j|c_k)=I(a_i|c_k)-I(a_i|b_jc_k) \tag{3.3.22}$$

由式（3.3.19）和式（3.3.21），可得互信息、联合互信息和条件互信息的关系为

$$\begin{aligned}I(a_i;b_jc_k)&=\log\frac{P(a_i|b_jc_k)}{P(a_i)}=\log\frac{P(a_i|b_jc_k)}{P(a_i|c_k)}\frac{P(a_i|c_k)}{P(a_i)}\\&=I(a_i;b_j|c_k)+I(a_i;c_k)\end{aligned} \tag{3.3.23}$$

式（3.3.23）表明，b_jc_k 联合给出关于 a_i 的互信息等于 c_k 给出关于 a_i 的互信息与在 c_k 已知条件下 b_j 给出关于 a_i 的互信息之和。同理可得

$$I(a_i;b_jc_k)=I(a_i;b_j)+I(a_i;c_k|b_j)$$

类似平均互信息的定义，平均联合互信息和平均条件互信息的定义如下。

定义 3.5 设联合集 XYZ，平均联合互信息定义为联合互信息 $I(a_i;b_jc_k)$ 在概率空间 XYZ 中的统计平均值。

$$I(X;YZ)=E[I(a_i;b_jc_k)]=\sum_{a_i\in X}\sum_{b_j\in Y}\sum_{c_k\in Z}P(a_ib_jc_k)\log\frac{P(a_i\mid b_jc_k)}{P(a_i)} \tag{3.3.24}$$

定义 3.6　设联合集 XYZ，在已知 Z 条件下，X 与 Y 之间的平均条件互信息定义为条件互信息 $I(a_i;b_j\mid c_k)$ 在概率空间 XYZ 中的统计平均值。

$$I(X;Y\mid Z)=E[I(a_i;b_j\mid c_k)]=\sum_{a_i\in X}\sum_{b_j\in Y}\sum_{c_k\in Z}P(a_ib_jc_k)\log\frac{P(a_i\mid b_jc_k)}{P(a_i\mid c_k)} \tag{3.3.25}$$

可以证明，平均联合互信息和平均条件互信息具有以下关系

$$\begin{aligned}I(X;Y\mid Z)&=H(X\mid Z)-H(X\mid YZ)\\&=H(X\mid Z)+H(Y\mid Z)-H(XY\mid Z)\\&=H(Y\mid Z)-H(Y\mid XZ)=I(Y;X\mid Z)\end{aligned} \tag{3.3.26}$$

$$I(X;YZ)=H(X)-H(X\mid YZ) \tag{3.3.27}$$

$$I(X;YZ)=I(X;Z)+I(X;Y\mid Z)=I(X;Y)+I(X;Z\mid Y) \tag{3.3.28}$$

$$I(X;YZ)=I(YZ;X) \tag{3.3.29}$$

$$I(X;Y\mid Z)\geqslant 0 \tag{3.3.30}$$

$$I(X;YZ)\geqslant I(X;Z) \tag{3.3.31}$$

当且仅当 $P(x\mid z)=P\left(x\mid yz\right)$ 时，式（3.3.30）和式（3.3.31）中的等号成立。即在 z 出现条件下，x 和 y 相互独立时，X 和 Y 互相之间提供的信息量为零。

【例 3.6】　设随机变量 X 和 Y 的联合概率分布如表 3.1 所示。

表 3.1　随机变量 *X* 和 *Y* 的联合概率

X	Y	
	$b_1=0$	$b_2=1$
$a_1=0$	1/3	1/3
$a_2=1$	0	1/3

已知随机变量 $Z=X\oplus Y$，试计算：

（1）$H(X)$，$H(Y)$，$H(Z)$，$H(XY)$，$H(YZ)$，$H(XYZ)$；

（2）$H(X\mid Y)$，$H(Z\mid XY)$；

（3）$I(X;Y)$，$I(X;Z\mid Y)$ 和 $I(Z;XY)$。

解： 由已知条件可得联合概率 $P(yz),P(xyz)$ 分别如表 3.2 和表 3.3 所示。

表 3.2　联合概率 $P(yz)$

yz	$P(yz)$
00	1/3
01	0
10	1/3
11	1/3

表 3.3　联合概率 $P(xyz)$

xy	$P(xyz)$	
	$z=0$	$z=1$
00	1/3	0
01	0	1/3
10	0	0
11	1/3	0

（1）$H(X)=H\left(\frac{2}{3},\frac{1}{3}\right)=0.918$（比特/符号）

$H(Y)=H\left(\frac{1}{3},\frac{2}{3}\right)=0.918$（比特/符号）

$H(Z)=H\left(\frac{2}{3},\frac{1}{3}\right)=0.918$（比特/符号）

$H(XY)=H\left(\frac{1}{3},\frac{1}{3},0,\frac{1}{3}\right)=1.585$（比特/二个符号）

$H(YZ)=H\left(\frac{1}{3},0,\frac{1}{3},\frac{1}{3}\right)=1.585$（比特/二个符号）

$H(XYZ)=H\left(\frac{1}{3},\frac{1}{3},\frac{1}{3},0,0,0,0,0\right)=1.585$（比特/三个符号）

（2）$H(X|Y)=H(XY)-H(Y)=0.667$（比特/符号）

$H(Z|XY)=H(XYZ)-H(XY)=0$

（3）$I(X;Y)=H(X)+H(Y)-H(XY)=0.251$（比特/符号）

$I(X;Z|Y)=H(X|Y)-H(X|YZ)=[H(XY)-H(Y)]-[H(XYZ)-H(YZ)]$

$=(1.585-0.918)-(1.585-1.585)$

$=0.667$（比特/符号）

$I(Z;XY)=H(Z)-H(Z|XY)=0.918$（比特/符号）

3.4 信道容量的定义

3.4.1 信息传输率和信息传输速率

为了研究在通信过程中信道传递信息的能力，这里先引入信道的信息传输率。

信道的信息传输率是指信道中平均每个符号所能传送的信息量，即平均互信息，单位为“比特/信道符号”。如果不涉及编码，信源输出符号直接送入信道进行传输，则一个信源符号就对应一个信道符号，因此为了简单起见，在不引起歧义的前提下，信息传输率的单位采用“比特/符号”。可见，信道的信息传输率表示为

$$R=I(X;Y)=H(X)-H(X|Y)=H(Y)-H(Y|X) \tag{3.4.1}$$

式中，X 表示信源发出的符号，Y 表示信宿收到的符号，平均互信息 $I(X;Y)$ 就是接收到符号 Y 后平均每个符号获得的关于 X 的信息量。从另一个角度看，也就是信道为信源传递的平均信息量。

将图 3.12 所示的维拉图在通信系统场景下作进一步展示，如图 3.13 所示。在通信系统中，信源符号的平均信息量为 $H(X)$，而信宿符号 Y 的平均信息量（信宿符号的平均不确定性）为 $H(Y)$，在实际信道中，由于信道特性不理想，$H(X)$ 和 $H(Y)$ 并没有固定的约束关系。信道对所传递的信息施加的影响可以从两个方面来看：一方面，一部分关于信源的信息熵 $H(X|Y)$ 在传输过程中被遗失（损失熵）；另一方面，信道又加入了一部分来自噪声的信息熵 $H(Y|X)$（噪声熵），因此信宿收到的符号提供的平均信息量只有一部分是来自信

源的，这部分就是 $I(X;Y)$ 。

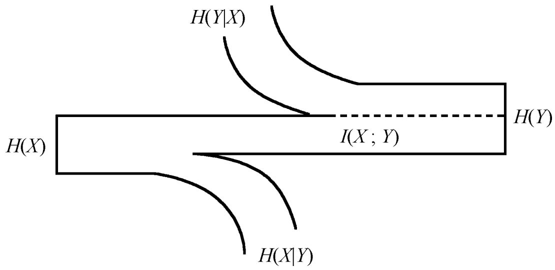

图 3.13　通信系统中各种熵和平均互信息的关系

例如两种极端情况下的信道：一种是无噪信道，它的信道矩阵是单位矩阵。因为 $H(Y|X)=H(X|Y)=0$，平均互信息则为 $I(X;Y)=H(X)=H(Y)$；另一种是无用信道，其信道输入端 X 和输出端 Y 统计独立，即 $P(b_j|a_i)=P(b_j)$，接收到 Y 后不可能消除有关输入端 X 的任何不确定性，所以获得的信息量为 0。例如二元对称信道中的转移概率 $p=0.5$ 时，无论信源输出符号是怎样的概率分布，此时

$$H(Y) = H(Y|X) = 1 \text{ 比特/符号} \tag{3.4.2}$$

可以算出平均互信息 $I(X;Y)=0$ 。

如果关注的是单位时间内信道平均传输的信息量，则定义**信道的信息传输速率** $R_{\rm b}$，表示为

$$R_{\rm b}=\frac{I(X;Y)}{T_{\rm s}}=R_{\rm s}I(X;Y)=R_{\rm s}[H(X)-H(X|Y)] \tag{3.4.3}$$

其中 $T_{\rm s}$ 表示平均传输一个符号需要的时间（单位：秒），$R_{\rm s}$ 表示信道的符号速率（单位：符号/秒）。信息传输速率表示信道每秒传输的平均信息量，单位为 bit/s。

3.4.2　信道容量

由定理 3.1 可知，$I(X;Y)$ 是输入变量 X 的概率分布 $P(x)$ 的 $\bigcap$ 形凸函数。因此对于一个固定信道，总存在某种概率分布 $P(x)$ 的信源，使传输每个符号平均获得的信息量最大，也就是每个固定信道都有一个最大的信息传输率。定义这个最大的信息传输率为**信道容量** C 。即

$$C=R_{\max}=\max_{P(x)}\{I(X;Y)\} \text{（比特/符号）} \tag{3.4.4}$$

而相应的输入概率分布称为**最佳输入分布**。

如果关注的是单位时间内信道平均传输的最大信息量，则信道容量定义为

$$C_{\rm t}=(R_{\rm b})_{\max}=R_{\rm s}\max_{P(x)}\{I(X;Y)\} \text{（比特/秒）} \tag{3.4.5}$$

这里 $C_{\rm t}$ 仍称为信道容量，只是增加一个下标 t 以示区别。式（3.4.5）中的 $R_{\rm s}$ 称为符号速率，是指信道每秒钟传输的符号数。

信道容量 C 只是信道转移概率的函数，只与信道的统计特性有关，而与输入信源的概率分布无关。即对于一个特定的信道，其信道容量 C 是确定的，是不随输入信源的概率分布变化而改变的。信道容量 C 取值的大小，直接反映了信道质量的高低。所以，信道容量是完全描述信道特性的参量，是信道能够传输的最大信息量。

对于离散信道而言，计算信道容量只需给出信道的转移概率矩阵即可。从数学上来说，式（3.4.4）所示的信道容量C就是平均互信息$I[P(x)]$对$P(x)$取极大值。因为任何信源概率分布必须遵循约束条件

$$\sum_X P(x)=1 \tag{3.4.6}$$

所以，信道容量C就是平均互信息$I[P(x)]$在约束条件式（3.4.6）的约束下，对信源概率分布$P(x)$取条件极大值。

【例 3.7】 已知某信道的转移概率矩阵为

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

试计算信道容量，并说明达到信道容量的最佳输入分布。

分析： 利用公式$I(X;Y)=H(Y)-H(Y|X)$计算平均互信息$I(X;Y)$，然后求其最大值，即得信道容量。

解： 设信源的概率空间为$\begin{bmatrix}X\\P(x)\end{bmatrix}=\begin{bmatrix}a_1 & a_2\\ \omega & 1-\omega\end{bmatrix}$，则信道输入符号和信道输出符号的联合概率$P(a_ib_j)$如表 3.4 所示。

表 3.4　例 3.7 中的联合概率 $P(a_ib_j)$

a_i	$P(a_ib_j)$		
	$b_j=b_1$	$b_j=b_2$	$b_j=b_3$
a_1	$\omega/2$	$\omega/2$	0
a_2	$(1-\omega)/2$	$(1-\omega)/4$	$(1-\omega)/4$

则信道输出符号的概率为

$$P(b_1)=1/2,\qquad P(b_2)=(1+\omega)/4,\qquad P(b_3)=(1-\omega)/4$$

所以

$$\begin{aligned}H(Y)&=H\left(\frac{1}{2},\frac{1+\omega}{4},\frac{1-\omega}{4}\right)\\&=\frac{3}{2}-\frac{1+\omega}{4}\log(1+\omega)-\frac{1-\omega}{4}\log(1-\omega)\ \text{（比特/符号）}\end{aligned}$$

又因为噪声熵

$$\begin{aligned}H(Y|X)&=P(a_1)H\left(\frac{1}{2},\frac{1}{2},0\right)+P(a_2)H\left(\frac{1}{2},\frac{1}{4},\frac{1}{4}\right)\\&=\frac{3}{2}-\frac{\omega}{2}\ \text{（比特/符号）}\end{aligned}$$

所以

$$\begin{aligned}I(X;Y)&=H(Y)-H(Y|X)\\&=-\frac{1+\omega}{4}\log(1+\omega)-\frac{1-\omega}{4}\log(1-\omega)+\frac{\omega}{2}\ \text{（比特/符号）}\end{aligned}$$

解 $\dfrac{\partial I(X;Y)}{\partial \omega}=0$，即 $\dfrac{1}{2}-\dfrac{1}{4}\log\dfrac{1+\omega}{1-\omega}=0$。

因此当 $\omega=0.6$ 时，平均互信息 $I(X;Y)$ 取最大值，即信道容量为

$$C=\max I(X;Y)=0.16 \text{（比特/符号）}$$

对应的最佳输入分布为

$$\begin{bmatrix} X \\ P(x) \end{bmatrix}=\begin{bmatrix} a_1 & a_2 \\ 0.6 & 0.4 \end{bmatrix}$$

3.5　单符号离散信道及其信道容量

对于一般单符号离散信道，信道容量的计算是比较复杂的，从数学上来说，就是对平均互信息 $I(X;Y)$ 求极大值的问题。但对于某些特殊信道，可利用其特点，运用信息理论的基本概念，简化信道容量的计算，直接得到信道容量的数值。下面先讨论几种典型的单符号离散信道的信道容量，然后讨论一般单符号离散信道的信道容量的计算。

3.5.1　典型信道的信道容量

1．无噪信道

因为信道矩阵是单位矩阵，这种信道的信道疑义度 $H(X|Y)$ 和噪声熵 $H(Y|X)$ 都等于零，即

$$H(Y|X)=H(X|Y)=0 \tag{3.5.1}$$

所以这类信道的平均互信息为

$$I(X;Y)=H(X)=H(Y) \tag{3.5.2}$$

那么信道容量为

$$C=\max_{P(x)}\{I(X;Y)\}=\max_{P(x)}\{H(X)\}=\log r \quad \text{（比特/符号）} \tag{3.5.3}$$

可见，当信源独立等概分布时，无噪信道达到信道容量，等于 $\log r$，其中 r 为信源 X 的符号个数。

2．对称离散信道的信道容量

若单符号离散信道的信道矩阵 $\boldsymbol{P}$ 中每一行都是同一符号集 $\{p_1', p_2', \cdots, p_s'\}$ 诸元素的不同排列，并且每一列也都由同一符号集 $\{q_1', q_2', \cdots, q_r'\}$ 中诸元素的不同排列组成，则这种信道称为**对称离散信道**。一般 $r \neq s$。例如

$$\boldsymbol{P}_1=\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} \quad \text{和} \quad \boldsymbol{P}_2=\begin{bmatrix} 0.2 & 0.3 & 0.5 \\ 0.5 & 0.2 & 0.3 \\ 0.3 & 0.5 & 0.2 \end{bmatrix}$$

所对应的信道是对称离散信道。但信道矩阵

$$\boldsymbol{P}_3=\begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6}\\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{3}\end{bmatrix} \quad 和 \quad \boldsymbol{P}_4=\begin{bmatrix}0.7 & 0.2 & 0.1\\ 0.2 & 0.1 & 0.7\end{bmatrix}$$

则不具有对称性，因而所对应的信道不是对称离散信道。这是因为在 $\boldsymbol{P}_3$ 和 $\boldsymbol{P}_4$ 中，虽然每行都是同一集合的不同排列，但每列不都是同一集合的不同排列。

定理 3.3 对于对称离散信道，当输入等概时达到信道容量，且信道容量为

$$C=\log s-H(p_1',p_2',\cdots,p_s') \tag{3.5.4}$$

其中，s 为信道的输出符号数，$H(p_1',p_2',\cdots,p_s')$ 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。

证明： 根据熵函数的对称性，对称离散信道的噪声熵为

$$\begin{aligned}H(Y\mid X)&=-\sum_{i=1}^{r}\sum_{j=1}^{s}P(a_i)P(b_j\mid a_i)\log P(b_j\mid a_i)\\&=\sum_{i=1}^{r}P(a_i)\{-\sum_{j=1}^{s}P(b_j\mid a_i)\log P(b_j\mid a_i)\}\\&=\sum_{i=1}^{r}P(a_i)H(p_1',p_2',\cdots,p_s')=H\{p_1',p_2',\cdots,p_s'\}\end{aligned} \tag{3.5.5}$$

由此可见，对称离散信道的噪声熵 $H(Y\mid X)$ 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 $H(p_1',p_2',\cdots,p_s')$。

考虑到行元素集合 $\{p_1',p_2',\cdots,p_s'\}$ 是给定的对称离散信道本身的固有参数，与输入信源 X 无关。所以对称离散信道容量为

$$\begin{aligned}C&=\max_{P(x)}\left\{I(X;Y)\right\}=\max_{P(x)}\left\{H(Y)-H(Y\mid X)\right\}\\&=\max_{P(x)}\{H(Y)\}-H(p_1',p_2',\cdots,p_s')\\&=\log s-H(p_1',p_2',\cdots,p_s')=\log s-H(\boldsymbol{P}\text{的行矢量})\end{aligned} \tag{3.5.6}$$

接下来的问题是，当输入信源 X：$\{a_1,a_2,\cdots,a_r\}$ 的概率分布 $P(X)$：$\{P(a_1),P(a_2),\cdots,P(a_r)\}$ 是什么分布时，才能使输出随机变量 Y 达到等概分布，从而使输出随机变量 Y 的熵 $H(Y)$ 达到最大值 $\log s$。

由于输出随机变量 Y 的概率分布为

$$P(b_j)=\sum_{i=1}^{r}P(a_i)P(b_j\mid a_i),\quad (j=1,2,\cdots,s) \tag{3.5.7}$$

式（3.5.7）中，$P(b_j\mid a_1),P(b_j\mid a_2),\cdots,P(b_j\mid a_r)$ 都是对称离散信道的列元素集合 $\{q_1',q_2',\cdots,q_r'\}$ 中的 r 个元素。由于其每一列也都由同一符号集 $\{q_1',q_2',\cdots,q_r'\}$ 的诸元素的不同排列组成，所以，要使输出随机变量 Y 等概，即

$$P(b_1)=P(b_2)=\cdots=P(b_s)=\frac{1}{s} \tag{3.5.8}$$

则必须要求输入信源 X：$\{a_1,a_2,\cdots,a_r\}$ 等概分布，即

$$P(a_1)=P(a_2)=\cdots=P(a_r)=\frac{1}{r} \tag{3.5.9}$$

[证毕]

可见，对于对称离散信道，只有当输入信源等概分布时，才能达到信道容量 C，信道容量 C 的值只取决于信道矩阵 $\boldsymbol{P}$ 中行元素集合 $\{p_1',p_2',\cdots,p_s'\}$ 和信道的输出符号数 s。此结论说明信道容量 C 是信道本身固有的特征参量。

【例 3.8】 设某对称离散信道的信道矩阵为

$$\boldsymbol{P}=\begin{bmatrix} 0.5 & 0.25 & 0.125 & 0.125 \\ 0.125 & 0.5 & 0.25 & 0.125 \\ 0.125 & 0.125 & 0.5 & 0.25 \\ 0.25 & 0.125 & 0.125 & 0.5 \end{bmatrix}$$

求其信道容量。

解：由对称信道的信道容量公式（3.5.4）得

$$C=\log 4-H(0.5,0.25,0.125,0.125)=0.25\ （比特/符号）$$

在这个信道中，每个符号平均能够传输的最大信息为 0.25bit，而且只有当信道输入是等概分布时才能达到这个最大值。

均匀信道是对称信道中一类特殊的信道。若输入符号数 r 与输出符号数 s 相等，即 $r=s$，且信道中总的错误转移概率为 p，平均分配给 $r-1$ 个输出符号，即信道矩阵为

$$\boldsymbol{P}=\begin{bmatrix} \bar{p} & \frac{p}{r-1} & \frac{p}{r-1} & \cdots & \frac{p}{r-1} \\ \frac{p}{r-1} & \bar{p} & \frac{p}{r-1} & \cdots & \frac{p}{r-1} \\ \frac{p}{r-1} & \frac{p}{r-1} & \bar{p} & \cdots & \frac{p}{r-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{p}{r-1} & \frac{p}{r-1} & \frac{p}{r-1} & \cdots & \bar{p} \end{bmatrix} \tag{3.5.10}$$

其中 $p+\bar{p}=1$，则此信道称为**均匀信道**或**强对称信道**。一般信道的信道矩阵中各行之和为 1，但各列之和不一定等于 1，而均匀信道中各列之和也等于 1。

根据式（3.5.4）得强对称离散信道的信道容量为

$$\begin{aligned} C&=\log r-H\left(\bar{p},\frac{p}{r-1},\frac{p}{r-1},\cdots,\frac{p}{r-1}\right) \\ &=\log r+\bar{p}\log\bar{p}+\underbrace{\frac{p}{r-1}\log\frac{p}{r-1}+\cdots+\frac{p}{r-1}\log\frac{p}{r-1}}_{共(r-1)项} \\ &=\log r+\bar{p}\log\bar{p}+p\log\frac{p}{r-1}=\log r-p\log(r-1)-H(p) \end{aligned} \tag{3.5.11}$$

式中，p 是总的错误转移概率，$\bar{p}$ 是正确转移概率。

二元对称信道就是 $r=2$ 的均匀信道，由式（3.5.11）可计算出信道容量为

$$C=1-H(p)\ （比特/符号）$$

当 $p=0.5$ 时的二元对称信道为无用信道，此时信道矩阵 $\boldsymbol{P}=\begin{bmatrix}0.5 & 0.5\\0.5 & 0.5\end{bmatrix}$，即便信道输入概率分布为最佳输入分布，信源的信息也全部损失在信道中，信道容量 $C=0$ 。

3．准对称离散信道的信道容量

若信道矩阵中，每行都是第一行元素的不同排列，每一列并不都是第一列元素的不同排列，但是该信道矩阵按列可以划分成几个互不相交的子集合，而每个子矩阵（由子集所对应的信道矩阵中的列所组成）具有下述性质：

（1）每一行都是第一行的一种排列；

（2）每一列都是第一列的一种排列。

则该信道为**准对称离散信道**。例如，信道矩阵 $\boldsymbol{P}=\begin{bmatrix}0.8 & 0.1 & 0.1\\0.1 & 0.1 & 0.8\end{bmatrix}$ 可以划分成两个对称的子矩阵 $\boldsymbol{P}_1=\begin{bmatrix}0.8 & 0.1\\0.1 & 0.8\end{bmatrix}$ 和 $\boldsymbol{P}_2=\begin{bmatrix}0.1\\0.1\end{bmatrix}$，因此它是准对称信道。图 3.6 所示的二元删除信道是准对称信道。

可以证明：准对称离散信道的最佳输入分布是等概分布，即准对称离散信道的信道容量就是输入符号独立等概时的平均互信息。因为准对称离散信道 $H(Y|X)=H(p_1',p_2',\cdots,p_s')$，则平均互信息 $I(X;Y)=H(Y)-H(p_1',p_2',\cdots,p_s')$。

定理 3.4 对于准对称离散信道，当输入等概时达到信道容量，其信道容量为

$$C=H(Y)-H(p_1',p_2',\cdots,p_s') \tag{3.5.12}$$

$$=\log r-\sum_{k=1}^{n}N_k\log M_k-H(p_1',p_2',\cdots,p_s') \tag{3.5.13}$$

式中，$H(Y)$ 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合个数，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。式（3.5.13）的证明需要借助后面的定理 3.6，感兴趣的读者可以查阅参考文献[21]。

【例 3.9】 设某信道矩阵为

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

求其信道容量。

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。

解：输入等概时，由信道矩阵可得联合概率 $P(xy)$，如表 3.5 所示。

表 3.5 联合概率 $P(a_ib_j)$

a_i	$P(a_ib_j)$		
	$b_j=b_1$	$b_j=b_2$	$b_j=b_3$
a_1	0.25	0.15	0.1
a_2	0.15	0.25	0.1

容易得到输出符号的概率分别为 0.4，0.4，0.2，所以

$$H(Y)=H(0.4,0.4,0.2)$$

因此，信道容量为

$$C=\max I(X;Y)=H(Y)-H(Y|X)$$

$$\xlongequal{\text{输入等概}}H(0.4.0.4,0.2)-H(0.5,0.3,0.2)=0.0365\ \text{（比特/符号）}$$

信道容量也可以通过式（3.5.13）计算得到。为此将 $\boldsymbol{P}=\begin{bmatrix}0.5 & 0.3 & 0.2\\ 0.3 & 0.5 & 0.2\end{bmatrix}$ 划分成两个对称的子矩阵

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

因为

$$\begin{aligned}&r=2, N_1=0.5+0.3=0.8, M_1=0.5+0.3=0.8,\\&N_2=0.2, M_2=0.2+0.2=0.4, n=2\end{aligned}$$

所以该准对称离散信道的信道容量为

$$\begin{aligned}C&=\log r-\sum_{k=1}^{n}N_k\log M_k-H(p_1',p_2',\cdots,p_s')\\&=\log 2-(0.8\log 0.8+0.2\log 0.4)-H(0.5.0.3,0.2)\\&=0.0365\text{（比特/符号）}\end{aligned}$$

3.5.2　串联信道及其信道容量

假设有一离散单符号信道 I，其输入变量为 X，输出变量为 Y，并设另有一离散单符号信道 II，其输入变量为 Y，输出变量为 Z，这两信道串接起来组成如图 3.14 所示的串联信道。其中，X 取值于集合 $\{a_1,a_2,\cdots,a_r\}$，Y 取值于集合 $\{b_1,b_2,\cdots,b_s\}$，Z 取值于集合 $\{c_1,c_2,\cdots,c_t\}$。信道 I 的转移概率记为 $P(y|x)=P(b_j|a_i)$，而信道 II 的转移概率一般与前面的 X 和 Y 都有关，记为 $P(z|xy)=P(c_k|a_ib_j)$。

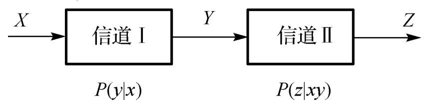

图 3.14　串联信道

图 3.14 所示的两个串联信道可以等价成一个总的离散信道，其输入为 X，取值于 $\{a_1,a_2,\cdots,a_r\}$，输出为 Z，取值于 $\{c_1,c_2,\cdots,c_t\}$，此信道的转移概率为

$$P(z|x)=\sum_{Y}P(y|x)\cdot P(z|xy)\qquad x\in X, y\in Y, z\in Z \tag{3.5.14}$$

如果这两信道的输入和输出（即 X、Y 和 Z 序列）构成马尔可夫链，那么信道 II 的输出只与输入 Y 有关，与前面的输入 X 无关。即在 y 出现的条件下，x 和 z 相互独立，满足

$$P(z|xy)=P(z|y)\quad\text{（对所有 } x,\ y,\ z\text{）} \tag{3.5.15}$$

此时串联信道的信道矩阵为

$$\Big[\underset{r\times t}{P(z|x)}\Big]=\Big[\underset{r\times s}{P(y|x)}\Big]\cdot\Big[\underset{s\times t}{P(z|y)}\Big] \tag{3.5.16}$$

可以证明，对于 N 个单符号信道组成的串联信道，若其输入输出变量之间组成一个马尔可夫链，设信道矩阵分别为 $\boldsymbol{P}_1,\boldsymbol{P}_2,\cdots,\boldsymbol{P}_N$，则串联信道矩阵为

$$P_{\text{串}}=P_1P_2\cdots P_N=\prod_{i=1}^{N}P_i \tag{3.5.17}$$

计算串联信道的信道容量并不困难。利用所求的串联信道矩阵，就可以按照前面介绍的方法来计算串联信道的信道容量。

【例 3.10】 设有两个离散二元对称信道，其组成的串联信道如图 3.15 所示，求该串联信道的信道容量。

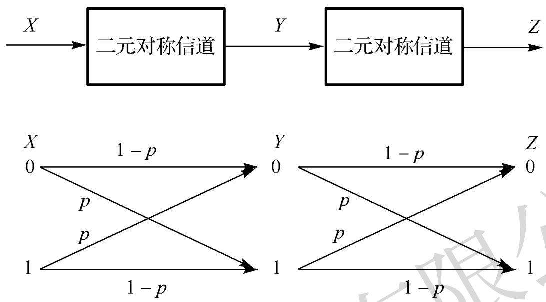

图 3.15　二元对称信道的串联信道

解： 两个二元对称信道的信道矩阵均为

$$P_1=P_2=\begin{bmatrix}1-p & p\\ p & 1-p\end{bmatrix}$$

由于 X 、Y 和 Z 组成马尔可夫链，则串联信道的信道矩阵为

$$P_{\text{串}}=P_1P_2=\begin{bmatrix}1-p & p\\ p & 1-p\end{bmatrix}^2=\begin{bmatrix}(1-p)^2+p^2 & 2p(1-p)\\ 2p(1-p) & (1-p)^2+p^2\end{bmatrix}$$

因此该串联信道仍然是一个二元对称信道。则串联信道的信道容量为

$$C_{\text{串}}=1-H[2p(1-p)]$$

3.5.3　信息处理定理

在通信系统中，信息从信源到信宿流动的过程中通常要经历大量的变换。例如通信系统中常见的编码/解码、调制/解调、加密/解密等，有必要探讨这些数据处理会对信息造成何种影响。对信息的处理过程在数学模型上与串联信道是相似的，可以使用相同的方式进行描述，如图 3.14 所示。下面将着重考察满足马尔可夫链的数据处理流程。

定理 3.5（信息处理定理） 如果 X、Y 和 Z 组成一个马尔可夫链，有

$$I(X;Y)\geqslant I(X;Z) \tag{3.5.18}$$

$$I(Y;Z)\geqslant I(X;Z) \tag{3.5.19}$$

证明： 因为 X、Y 和 Z 组成马尔可夫链，因此

$$I(X;Z|Y)=0$$

因为

$$I(X;YZ)=I(X;Y)+I(X;Z|Y)=I(X;Z)+I(X;Y|Z)$$

而且 $I(X;Y|Z) \geqslant 0$，可得

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

式（3.5.18）得证。

因为 X、Y 和 Z 组成马尔可夫链，因此

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

又因为 $I(XY;Z) \geqslant I(X;Z)$，所以

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

式（3.5.19）得证。

信息处理定理说明：当消息通过串联处理时，其输入消息和输出消息之间的平均互信息不会超过输入消息与中间消息之间的平均互信息，也不会超过中间消息与输出消息之间的平均互信息。也就是说，通过信息处理后，一般只会增加信息的损失，最多保持原来获得的信息，不可能比原来获得的信息有所增加，因此也称为**信息不增性原理**。

3.5.4　一般离散信道的信道容量

除了上面介绍的几种特殊离散信道，在实际通信系统中，还存在大量的离散信道，其转移概率 $P(y|x)$ 不符合以上任何一种特殊形式。对于这类离散信道，其信道容量和最佳输入分布的求解比较困难，本节将从信道容量的数学定义出发，导出达到信道容量需要满足的条件。

根据信道容量的定义，信道容量是在固定信道转移概率 $P(y|x)$ 的条件下，对所有可能的输入概率分布 $P(x)$ 求得的平均互信息 $I(X;Y)$ 的最大值。由定理 3.1 可知，平均互信息 $I(X;Y)$ 是输入概率分布 $P(x)$ 的 $\cap$ 形凸函数，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

$$\sum_{i=1}^{r} P(a_i) = 1 \tag{3.5.20}$$

所以，求信道容量 C 就是在约束条件（3.5.20）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

$$F[P(a_1),P(a_2),\cdots,P(a_r)] = I(X;Y) - \lambda\sum_{X} P(a_i) \tag{3.5.21}$$

式中，λ 为拉格朗日乘子。

当

$$\frac{\partial F}{\partial P(a_i)} = \frac{\partial[I(X;Y) - \lambda\sum_{X} P(a_i)]}{\partial P(a_i)} = 0 \tag{3.5.22}$$

时求得的 $I(X;Y)$ 的值即为信道容量。

由于

$$I(X;Y)=\sum_{i=1}^{r}\sum_{j=1}^{s}P(a_i)P(b_j\,|\,a_i)\log\frac{P(b_j\,|\,a_i)}{P(b_j)}$$
$$=\sum_{i=1}^{r}\sum_{j=1}^{s}P(a_i)P(b_j\,|\,a_i)\left(\log P(b_j\,|\,a_i)-\log P(b_j)\right)$$

而
$$P(b_j)=\sum_{i=1}^{r}P(a_i)P(b_j\,|\,a_i)$$

所以
$$\frac{\partial}{\partial P(a_i)}\log P(b_j)=\left[\frac{\partial}{\partial P(a_i)}\ln P(b_j)\right]\log \mathrm{e}=\frac{P(b_j\,|\,a_i)}{P(b_j)}\log \mathrm{e} \tag{3.5.23}$$

对式（3.5.22）整理得
$$\frac{\partial F}{\partial P(a_i)}=\frac{\partial[I(X;Y)-\lambda\sum_{X}P(a_i)]}{\partial P(a_i)}$$
$$=\sum_{j=1}^{s}P(b_j\,|\,a_i)\log\frac{P(b_j\,|\,a_i)}{P(b_j)}-\sum_{k=1}^{r}\sum_{j=1}^{s}P(a_k)P(b_j\,|\,a_k)\frac{P(b_j\,|\,a_i)}{P(b_j)}\log \mathrm{e}-\lambda \tag{3.5.24}$$

而
$$\sum_{k=1}^{r}\sum_{j=1}^{s}P(a_k)P(b_j\,|\,a_k)\frac{P(b_j\,|\,a_i)}{P(b_j)}=\sum_{k=1}^{r}\sum_{j=1}^{s}P(a_kb_j)\frac{P(b_j\,|\,a_i)}{P(b_j)}$$
$$=\sum_{j=1}^{s}P(b_j)\frac{P(b_j\,|\,a_i)}{P(b_j)}=\sum_{j=1}^{s}P(b_j\,|\,a_i)=1 \tag{3.5.25}$$

因此，式（3.5.24）可以化简为
$$\frac{\partial F}{\partial P(a_i)}=\frac{\partial[I(X;Y)-\lambda\sum_{X}P(a_i)]}{\partial P(a_i)}=\sum_{j=1}^{s}P(b_j\,|\,a_i)\log\frac{P(b_j\,|\,a_i)}{P(b_j)}-\log \mathrm{e}-\lambda \tag{3.5.26}$$

令 $\frac{\partial F}{\partial P(a_i)}=0$，得
$$\sum_{j=1}^{s}P(b_j\,|\,a_i)\log\frac{P(b_j\,|\,a_i)}{P(b_j)}=\lambda+\log \mathrm{e}\qquad (i=1,2,\cdots,r) \tag{3.5.27}$$

式（3.5.27）两边分别乘以 $P(a_i)$，并求和得
$$\sum_{i=1}^{r}\sum_{j=1}^{s}P(a_i)P(b_j\,|\,a_i)\log\frac{P(b_j\,|\,a_i)}{P(b_j)}=\lambda+\log \mathrm{e} \tag{3.5.28}$$

式（3.5.28）左边即为平均互信息的极大值 C，即
$$C=\lambda+\log \mathrm{e} \tag{3.5.29}$$

结合式（3.5.29），把式（3.5.27）中前 r 个方程改写成

$$\sum_{j=1}^{s}P(b_j|a_i)\log P(b_j|a_i)-\sum_{j=1}^{s}P(b_j|a_i)\log P(b_j)=C \qquad (i=1,2,\cdots,r)$$

移项后得

$$\sum_{j=1}^{s}P(b_j|a_i)[C+\log P(b_j)]=\sum_{j=1}^{s}P(b_j|a_i)\log P(b_j|a_i) \qquad (i=1,2,\cdots,r) \tag{3.5.30}$$

将 $\beta_j=C+\log P(b_j)$ 代入式（3.5.30），得

$$\sum_{j=1}^{s}P(b_j|a_i)\beta_j=\sum_{j=1}^{s}P(b_j|a_i)\log P(b_j|a_i) \qquad (i=1,2,\cdots,r) \tag{3.5.31}$$

这是含有 s 个未知数 β_j、有 r 个方程的非齐次线性方程组。

如果设 $r=s$，信道矩阵 $\boldsymbol{P}$ 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为

$$C=\log\sum_{j}2^{\beta_j} \quad \text{（比特/符号）} \tag{3.5.32}$$

由这个 C 值就可解得对应的输出概率分布 $P(b_j)$ 为

$$P(b_j)=2^{\beta_j-C} \qquad (j=1,2,\cdots,s) \tag{3.5.33}$$

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

观察式（3.5.27）可以发现，该式的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（3.5.29）可知

$$I(x_i;Y)=\sum_{j=1}^{s}P(b_j|a_i)\log\frac{P(b_j|a_i)}{P(b_j)}=C \tag{3.5.34}$$

由此可以导出以下定理。

定理 3.6　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

$$\begin{cases} I(x_i;Y)=C, & p_i\neq 0 \\ I(x_i;Y)\leqslant C, & p_i=0 \end{cases} \tag{3.5.35}$$

这时的 $I(X;Y)$ 就是信道容量 C。

定理 3.6 说明当达到信道容量时，信源符号集里的每一个概率不为零的符号对于接收端贡献的平均互信息是相等的。考虑对于一个特定的信道，当信源的分布特征不满足式（3.5.35）时，必定存在一些符号提供的平均互信息比较大，而一个经过良好设计的信源势必会更多地发送这些传输特性“优良”的符号，然而过多的发送将改变信源的输入概率分布，使得这些符号的概率上升，从而降低这些符号提供的平均互信息，同时提高其他符号提供的平均互信息。最终，所有发送概率不为零的符号提供的平均互信息将达到一个相同的水平。

定理 3.6 有助于检验一种指定的分布 $P(x)$ 是否为最佳分布，从而判断其对应的 $I(X;Y)$ 是否达到信道容量 C。

【例 3.11】 设离散信道矩阵为

$$\boldsymbol{P}=\begin{bmatrix}1 & 0 & 0\\ \dfrac{1}{3} & \dfrac{1}{3} & \dfrac{1}{3}\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$$

求信道容量。

解：根据信道矩阵可知，该信道不是对称信道也不是准对称信道。

设输入符号集为 $\{a_1,a_2,a_3,a_4\}$，输出符号集为 $\{b_1,b_2,b_3\}$。仔细观察信道矩阵 $\boldsymbol{P}$，可以发现输入符号 a_1,a_3,a_4 与输出符号 b_1,b_2,b_3 是一一对应的，而输入符号 a_2 等概地映射到三个输出符号。考虑如果将 a_2 的概率置零，则信道演变为一个理想信道，从而可将 a_1,a_3,a_4 设置为等概分布，即

$$P(a_2)=0,\ P(a_1)=P(a_3)=P(a_4)=1/3$$

以上分布是否为最佳分布，所对应的 $I(X;Y)$ 是否达到信道容量？根据定理 3.6，可以分别求出所有概率非零的符号对应的互信息。

$$I(x=a_1;Y)=\sum_{j=1}^{3}P(b_j\,|\,a_1)\log\frac{P(b_j\,|\,a_1)}{P(b_j)}=\log 3$$

$$I(x=a_2;Y)=\sum_{j=1}^{3}P(b_j\,|\,a_2)\log\frac{P(b_j\,|\,a_2)}{P(b_j)}=0$$

$$I(x=a_3;Y)=\sum_{j=1}^{3}P(b_j\,|\,a_3)\log\frac{P(b_j\,|\,a_3)}{P(b_j)}=\log 3$$

$$I(x=a_4;Y)=\sum_{j=1}^{3}P(b_j\,|\,a_4)\log\frac{P(b_j\,|\,a_4)}{P(b_j)}=\log 3$$

可见，此分布对应的互信息满足式（3.5.35）

$$\begin{cases}I\left(x_i;Y\right)=\log 3, & p_i\neq 0\\ I\left(x_i;Y\right)=0<\log 3, & p_i=0\end{cases}$$

所以，此分布为最佳分布，对应的信道容量为 $C=\log 3$（比特/符号）。

实际上，例 3.11 中的信道虽然不属于特殊信道，却可以通过直接观察其信道矩阵的特征，轻易地演变成理想信道，因此其最佳分布和信道容量还是比较容易求得的。在本例中，定理 3.6 验证了关于最佳分布的推断，并导出了信道容量。然而，更多实际的信道是无法通过直接观察推测出最佳分布的，因此需要通过求解式（3.5.31）来解决。

【例 3.12】 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

$$\boldsymbol{P}=\begin{bmatrix}\frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4}\\ \frac{1}{2} & 0 & \frac{1}{2} & 0\\ 0 & \frac{1}{2} & \frac{1}{2} & 0\\ \frac{1}{8} & \frac{1}{8} & \frac{1}{4} & \frac{1}{2}\end{bmatrix}$$

求其信道容量及最佳输入分布。

解：观察该信道转移概率矩阵可知，该信道不是特殊信道，也难以通过直接观察获得最佳分布，因此无法通过定理 3.6 来计算信道容量。所以，根据式（3.5.31）得

$$\begin{cases}\frac{1}{4}\beta_1+\frac{1}{2}\beta_2+\frac{1}{4}\beta_4=\frac{1}{4}\log\frac{1}{4}+\frac{1}{2}\log\frac{1}{2}+\frac{1}{4}\log\frac{1}{4}\\ \frac{1}{2}\beta_1+\frac{1}{2}\beta_3=\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\\ \frac{1}{2}\beta_2+\frac{1}{2}\beta_3=\frac{1}{2}\log\frac{1}{2}+\frac{1}{2}\log\frac{1}{2}\\ \frac{1}{8}\beta_1+\frac{1}{8}\beta_2+\frac{1}{4}\beta_3+\frac{1}{2}\beta_4=\frac{1}{8}\log\frac{1}{8}+\frac{1}{8}\log\frac{1}{8}+\frac{1}{4}\log\frac{1}{4}+\frac{1}{2}\log\frac{1}{2}\end{cases}$$

解得

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

由式（3.5.32），得

$$C=\log\sum_j 2^{\beta_j}=0.7039\text{（比特/符号）}$$

由式（3.5.33），得

$$P(b_1)=P(b_2)=0.2735\text{，}\quad P(b_3)=0.3445\text{，}\quad P(b_4)=0.1085$$

又由 $P(b_j)=\sum_{i=1}^{r}P(a_i)P(b_j\,|\,a_i)$，可列写下式

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

因此

$$[P(a_1),P(a_2),P(a_3),P(a_4)]=[0.2735\quad 0.2735\quad 0.3445\quad 0.1085]\begin{bmatrix}\frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4}\\ \frac{1}{2} & 0 & \frac{1}{2} & 0\\ 0 & \frac{1}{2} & \frac{1}{2} & 0\\ \frac{1}{8} & \frac{1}{8} & \frac{1}{4} & \frac{1}{2}\end{bmatrix}^{-1}$$

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

因此，最佳分布为

$$P(a_1)=0.2698;\quad P(a_2)=0.3915;\quad P(a_3)=0.2566;\quad P(a_4)=0.0821$$

下面，使用定理 3.6 验证以上结果。

$$\begin{aligned}I(x=a_1;Y)&=\sum_{j=1}^{4}P(b_j\,|\,a_1)\log\frac{P(b_j\,|\,a_1)}{P(b_j)}\\&=\frac{1}{4}\log\frac{1/4}{0.2735}+\frac{1}{2}\log\frac{1/2}{0.2735}+\frac{1}{4}\log\frac{1/4}{0.1085}\\&=0.7039\text{（比特/符号）}\end{aligned}$$

同理可以计算

$$I(x=a_2;Y)=\sum_{j=1}^{4}P(b_j\,|\,a_2)\log\frac{P(b_j\,|\,a_2)}{P(b_j)}=0.7039\text{（比特/符号）}$$

$$I(x=a_3;Y)=\sum_{j=1}^{4}P(b_j\,|\,a_3)\log\frac{P(b_j\,|\,a_3)}{P(b_j)}=0.7039\text{（比特/符号）}$$

$$I(x=a_4;Y)=\sum_{j=1}^{4}P(b_j\,|\,a_4)\log\frac{P(b_j\,|\,a_4)}{P(b_j)}=0.7039\text{（比特/符号）}$$

显然，满足式（3.5.35）

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

而每个符号贡献的互信息也正好是求解出的信道容量，证实了该求解过程是正确的。例 3.12 展示了一般离散信道的信道容量和最佳输入分布的求解过程，该算法是否对任意信道条件都适用？下面将通过例 3.13 来分析。

【例 3.13】 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

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

求其信道容量及最佳输入分布。

解： 本例与例 3.12 相似，因此尝试使用相同的求解过程。根据式（3.5.31）得

$$\begin{cases}\frac{1}{2}\beta_1+\frac{1}{4}\beta_3+\frac{1}{4}\beta_4=\frac{1}{2}\log\frac{1}{2}+\frac{1}{4}\log\frac{1}{4}+\frac{1}{4}\log\frac{1}{4}\\ \frac{1}{2}(\beta_1+\beta_4)=2\times\frac{1}{2}\log\frac{1}{2}\\ \beta_2=0\\ \frac{1}{2}(\beta_3+\beta_4)=2\times\frac{1}{2}\log\frac{1}{2}\end{cases}$$

解得

$$\beta_1=\beta_3=-2,\quad \beta_2=\beta_4=0$$

由式（3.5.33），得

$$P(b_1)=P(b_3)=0.1,\quad P(b_2)=P(b_4)=0.4,$$

又由 $P(b_j)=\sum_{i=1}^{4}P(a_i)P(b_j\,|\,a_i)$，可列写下式

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

因此

$$\begin{aligned}[P(a_1),P(a_2),P(a_3),P(a_4)]&=[P(b_1),P(b_2),P(b_3),P(b_4)]\boldsymbol{P}^{-1}\\&=[-0.4\quad 0.6\quad 0.4\quad 0.4]\end{aligned}$$

导出的最佳分布为

$$P(a_1)=-0.4;\quad P(a_2)=0.6;\quad P(a_3)=0.4;\quad P(a_4)=0.4$$

很明显，这不是一个合法的概率分布。究其原因，是因为前面通过拉格朗日法求极值时只限定了约束条件 $\sum_{i=1}^{r}P(a_i)=1$，而并未限定 $P(a_i)\geqslant 0$，从而导致负概率的出现。当这种情况出现时，表明达到极值的条件已经超出了合法的概率区间，因此，合法的最大 $I(X;Y)$ 应该出现在约束条件的边界上。因而，势必有符号的先验概率应该取值为 0，从趋势上看，可将求解出负概率值的符号的先验概率置零。因此，取 $P(a_1)=0$，并在此条件下，再次进行上述运算。

因为 $P(a_1)=0$，即信源以 0 概率发出符号 a_1，因此信道转移概率矩阵 $P(y\,|\,x)$ 的第一行对于该信源是没有意义的，在根据式（3.5.31）列写方程时不考虑。

$$\begin{cases}\dfrac{1}{2}(\beta_1+\beta_4)=2\times\dfrac{1}{2}\log\dfrac{1}{2}\\ \beta_2=0\\ \dfrac{1}{2}(\beta_3+\beta_4)=2\times\dfrac{1}{2}\log\dfrac{1}{2}\end{cases}$$

显然，这是一个欠定方程，有无数个解，可以暂时将 β_1 视为已知参量，因此有

$$\beta_2=0,\quad \beta_3=\beta_1,\quad \beta_4=-2-\beta_1$$

由式（3.5.32），得

$$C=\log\sum_j 2^{\beta_j}=\log(2\times 2^{\beta_1}+1+2^{-2-\beta_1})\text{（比特/符号）}\tag{3.5.36}$$

由式（3.5.33），得

$$\begin{cases}P(b_1)=P(b_3)=2^{\beta_1-C}\\ P(b_2)=2^{\beta_2-C}=2^{-C}\\ P(b_4)=2^{\beta_4-C}=2^{-2-\beta_1-C}\end{cases}\tag{3.5.37}$$

又由 $P(b_j)=\sum_{i=2}^{4}P(a_i)P(b_j|a_i)$，可列写下式

$$\left[P(a_1),P(a_2),P(a_3),P(a_4)\right]\boldsymbol{P}=\left[P(b_1),P(b_2),P(b_3),P(b_4)\right]$$

因此

$$\begin{cases}\frac{1}{2}P(a_2)=P(b_1)\\P(a_3)=P(b_2)\\\frac{1}{2}P(a_4)=P(b_3)\\\frac{1}{2}P(a_2)+\frac{1}{2}P(a_4)=P(b_4)\end{cases}\tag{3.5.38}$$

将式（3.5.37）代入式（3.5.38），求解得

$$P(a_2)=P(a_4)=0.2929,\quad P(a_3)=0.4142,\quad \beta_1=-1.5$$

将 $\beta_1=-1.5$ 代入式（3.5.36），得

$$C=\log\sum_j 2^{\beta_j}=1.2716\text{（比特/符号）}$$

同样，可以通过定理 3.6 验证求得的最优分布和信道容量，具体验证过程留给读者完成。

对比例 3.12 和例 3.13，这两个例子给出的信道均满足 $r=s$，且信道矩阵 $\boldsymbol{P}$ 为满秩矩阵。因此，对这种情况下的一般离散信道容量的计算步骤总结如下：

（1）由 $\sum_{j=1}^{s}P(b_j|a_i)\beta_j=\sum_{j=1}^{s}P(b_j|a_i)\log P(b_j|a_i)$，求出 β_j；

（2）由 $C=\log\sum_j 2^{\beta_j}$，求出 C；

（3）由 $P(b_j)=2^{\beta_j-C}$，求出 $P(b_j)$；

（4）由 $P(b_j)=\sum_{i=1}^{r}P(a_i)P(b_j|a_i)$，求出 $P(a_i)$。

值得注意的是，按上述方法所求出的 $P(a_i)$ 并不一定满足概率的条件，必须对解进行检查。如果解得所有 $P(a_i)\geqslant 0$，则此解就是正确的解。如果有某些 $P(a_i)<0$，则此解无效。它表明所求的极限值 C 出现的区域不满足概率条件。那么，这时最大值必在边界上，即某些输入符号的概率 $P(a_i)=0$。因此，必须设某些输入符号的概率 $P(a_i)=0$，然后重新进行计算。

若 $r<s$，求解非齐次线性方程组就比较困难，即使已求出，也无法保证求得的输入符号概率都大于或等于零。因此，必须反复进行试运算，这时运算变得十分复杂。在这样的应用场景下，借助计算机的强大数值运算能力的迭代算法则成为重要的求解工具。1972 年，S.Arimoto 和 R.E.Blahut 提出了一种离散无记忆信道容量的迭代算法，该算法能在有限的运算步骤内以任意指定的精度逼近信道容量 C。

3.6　多符号离散信道及其信道容量

3.6.1　多符号离散信道的平均互信息

在 3.2 节中讨论了多符号离散信道的数学模型，信道输入 $\boldsymbol{X}=(X_1X_2\cdots X_N)$，信道输出 $\boldsymbol{Y}=(Y_1Y_2\cdots Y_N)$。传输过程中传递的平均互信息为

$$I(\boldsymbol{X};\boldsymbol{Y})=I(X_1X_2\cdots X_N;Y_1Y_2\cdots Y_N)$$

下面将讨论 $I(\boldsymbol{X};\boldsymbol{Y})$ 与其中各随机变量的平均互信息之和 $\sum_{l=1}^{N}I(X_l;Y_l)$ 之间的关系。

定理 3.7　若信道的输入随机矢量为 $\boldsymbol{X}=(X_1X_2\cdots X_N)$，通过信道传输，接收到的随机矢量为 $\boldsymbol{Y}=(Y_1Y_2\cdots Y_N)$。如果信道是无记忆的，即信道转移概率满足

$$P(\boldsymbol{y}|\boldsymbol{x})=P(y_1y_2\cdots y_N|x_1x_2\cdots x_N)=\prod_{l=1}^{N}P(y_l|x_l) \tag{3.6.1}$$

则存在

$$I(\boldsymbol{X};\boldsymbol{Y})\leqslant\sum_{l=1}^{N}I(X_l;Y_l) \tag{3.6.2}$$

式中，X_l 和 Y_l 是随机矢量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 中对应的第 l 位随机变量。

定理 3.8　若信道的输入随机矢量为 $\boldsymbol{X}=(X_1X_2\cdots X_N)$，通过信道传输，接收到的随机矢量为 $\boldsymbol{Y}=(Y_1Y_2\cdots Y_N)$，而信道的转移概率为 $P(\boldsymbol{y}|\boldsymbol{x})$，如果信源是无记忆的，即

$$P(\boldsymbol{x})=P(x_1x_2\cdots x_N)=\prod_{l=1}^{N}P(x_l)$$

则存在

$$I(\boldsymbol{X};\boldsymbol{Y})\geqslant\sum_{l=1}^{N}I(X_l;Y_l) \tag{3.6.3}$$

式中，X_l 和 Y_l 是随机矢量 $\boldsymbol{X}$ 和 $\boldsymbol{Y}$ 中对应的第 l 位随机变量。

定理 3.7 和定理 3.8 可利用詹森不等式证明得到，请参见附录 A。

如果信源与信道都是无记忆的，则式（3.6.2）和式（3.6.3）同时满足，即

$$I(\boldsymbol{X};\boldsymbol{Y})=\sum_{l=1}^{N}I(X_l;Y_l) \tag{3.6.4}$$

3.6.2　离散无记忆信道的信道容量

对于一般离散无记忆信道的 N 次扩展信道，信道容量为

$$C_{1,2,\cdots,N}=\max_{P(x)}I(\boldsymbol{X};\boldsymbol{Y})$$

当信源无记忆时，由式（3.6.4）得

$$C_{1,2,\cdots,N}=\max_{P(\boldsymbol{x})}\sum_{l=1}^{N}I(X_l;Y_l)=\sum_{l=1}^{N}\max_{P(x_l)}I(X_l;Y_l)=\sum_{l=1}^{N}C_l \tag{3.6.5}$$

式（3.6.5）中 $C_l=\max\limits_{P(x_l)}I(X_l;Y_l)$，这是某时刻 l 通过离散无记忆信道传输的最大信息量。

如果离散无记忆信道是平稳（即时不变）的，则任何时刻信道的转移概率保持不变，即 $P(y_l|x_l)=P(y|x)$，因为信道输入 $\boldsymbol{X}=(X_1X_2\cdots X_N)$ 中的随机变量 $X_l(l=1,2,\cdots,N)$ 取自同一信源符号集，并具有同一种概率分布，所以任何时刻通过离散无记忆信道传输的最大信息量都相同，即 $C_l=C\ (l=1,2,\cdots,N)$。因此，离散无记忆的 N 次扩展信道的信道容量为

$$C_{1,2,\cdots,N}=NC \tag{3.6.6}$$

式（3.6.6）说明离散平稳无记忆的 N 次扩展信道的信道容量等于原单符号离散信道的信道容量的 N 倍。当且仅当输入信源是无记忆的，且每一输入变量 $X_l(l=1,2,\cdots,N)$ 的分布各自达到最佳分布 $P(x)$ 时，才能达到这个信道容量 NC。

【例 3.14】 求图 3.5 所示的二元无记忆离散对称信道的二次扩展信道的信道容量。

解： 因为图 3.5 所示的单符号二元对称信道的信道容量为

$$C=\log 2-H(p)=1-H(p)$$

所以二次扩展信道的信道容量

$$C_{1,2}=2[\log 2-H(p)]$$

信道容量也可以通过二元无记忆离散对称信道的二次扩展信道矩阵得到。因为该二次扩展信道矩阵仍是对称信道，由式（3.2.15）所示的信道矩阵，可知信道容量

$$C_{1,2}=\log 4-H(\bar{p}^2,\bar{p}p,p\bar{p},p^2)$$

例如，当 $p=0.1$ 时，单符号二元对称信道的信道容量 $C=0.531$（比特/符号），则二次扩展信道的信道容量为 $C_{1,2}=1.062$（比特/序列）

3.6.3 独立并联信道及其信道容量

独立并联信道如图 3.16 所示。设有 N 个信道，它们的输入分别是 $(X_1X_2\cdots X_N)$，它们的输出分别是 $(Y_1Y_2\cdots Y_N)$，它们的转移概率分别是 $P(y_1|x_1),P(y_2|x_2),\cdots,P(y_N|x_N)$。在这 N 个独立并联信道中，每一个信道的输出 Y_l 只与本信道的输入 X_l 有关，与其他信道的输入、输出都无关。那么，这 N 个信道的联合转移概率满足

$$P(y_1y_2\cdots y_N|x_1x_2\cdots x_N)=P(y_1|x_1)P(y_2|x_2)\cdots P(y_N|x_N) \tag{3.6.7}$$

这相当于信道是无记忆时应满足的条件。因此可以把定理 3.7 的结论推广应用到 N 个独立并联信道中来。定理 3.7 推广得

$$I(X_1X_2\cdots,X_N;Y_1Y_2\cdots Y_N)\leqslant\sum_{l=1}^{N}I(X_l;Y_l) \tag{3.6.8}$$

即联合平均互信息不大于各自信道的平均互信息之和。

因此得到独立并联信道的信道容量

$$C_{1,2,\cdots,N}=\max_{P(x_1\cdots x_N)}I(X_1\cdots X_N;Y_1\cdots Y_N)=\sum_{l=1}^{N}C_l \tag{3.6.9}$$

式中，C_l 是第 l 个独立信道的信道容量，即 $C_l=\max\limits_{P(x_l)}I(X_l;Y_l)$。

当 N 个输入符号相互独立，且每个输入符号 X_l 的概率分布达到各个信道容量的最佳输入分布时，独立并联信道的信道容量等于各信道容量之和，即

$$C_{1,2,\cdots,N}=\sum_{l=1}^{N}C_l \tag{3.6.10}$$

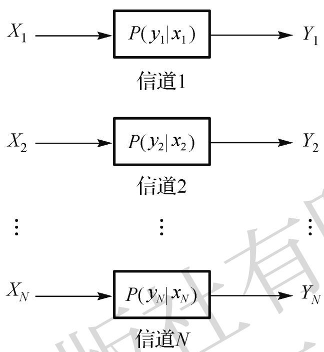

图 3.16　N 个独立并联信道

3.7　信源与信道的匹配

从通信系统的模型来看，信源发出的消息符号一定要通过信道来传输。对于一个信道，其信道容量是一定的，只有当输入符号的概率分布 $P(x)$ 满足一定条件时才能达到信道容量 C。这就是说，只有一定的信源才能使某一信道的信息传输率达到最大。

可见，信道的信息传输率 R 与信源分布是有密切关系的。一般信源与信道连接时，其信息传输率 $R=I(X;Y)$ 并未达到最大。这样，信道没有得到充分利用，即信道的信息传输率还有提高的可能。当 R 达到信道容量 C 时，称**信源与信道达到匹配**，否则认为信道存在剩余。由此引入信道剩余度的概念。

设信道的信息传输率为 $I(X;Y)$，信道容量为 C，则**信道剩余度**定义为

$$\text{信道剩余度}=\frac{C-I(X;Y)}{C}=1-\frac{I(X;Y)}{C} \tag{3.7.1}$$

特别地，当信道损失 $H(X|Y)=0$ 时，有 $I(X;Y)=H(X)$，则信道容量为

$$C=\max_{p(x)}\{H(X)\}=\log r \tag{3.7.2}$$

式中，r 是信道输入符号的个数。此时的信道剩余度为 $1-\dfrac{H(X)}{\log r}$，与第 2.7 节离散信源的剩余度比较，此时的信道剩余度就是信源剩余度。

可见，对于无噪信道，为了减少信道剩余度，可以对信源进行信源编码，使信道的信息传输率尽可能接近信道容量；对于有噪信道，可以通过信道编码使无差错传输的信息传

输率尽可能接近信道容量。在通信系统设计时，信源编码只考虑信源的统计特性，假定信道无噪声干扰；而信道编码只考虑信道的传输特性，假定信源输出独立等概。可见通过信源编码和信道编码，可使信源和信道达到匹配，信道资源得到充分利用。

本章小结

1．互信息

$$I(a_i;b_j)=I(a_i)-I(a_i\,|\,b_j)$$

当 a_i 和 b_j 相互独立时，互信息为 0。互信息 $I(a_i;b_j)$ 为正值，意味着事件 b_j 的出现有利于肯定事件 a_i 的发生；反之，则是不利的。

2．平均互信息

$$I(X;Y)=H(X)-H(X\,|\,Y)=H(Y)-H(Y\,|\,X)=H(X)+H(Y)-H(XY)$$

当 X 表示信道输入、Y 表示信道输出时，$I(X;Y)$ 就是接收端所获得的关于信源的平均信息量，$H(X\,|\,Y)$ 为信道疑义度，$H(Y\,|\,X)$ 为噪声熵。

平均互信息具有凸函数性。对于一个固定信道，一定存在某种概率分布的信源，使输出端获得的平均信息量为最大；对于一个固定信源，一定存在某种转移概率的信道，使该信源经过此信道传输时，信道的平均互信息达到极小值。

3．信息处理定理

如果 X、Y 和 Z 组成一个马尔可夫链，则有

$$I(X;Y)\geqslant I(X;Z)\text{；}\quad I(Y;Z)\geqslant I(X;Z)$$

4．单符号离散信道容量

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

信道容量是完全描述信道特性的参量，是信道平均每个符号能够传输的最大信息量，单位为“比特/信道符号”。

（1）无噪信道的信道容量 $C=\log r$，其中 r 为信道输入 X 的符号个数。最佳输入分布为独立等概分布。

（2）对称离散信道的信道容量 $C=\log s-H(p_1',p_2',\cdots,p_s')$，其中 s 为信道输出 Y 的符号个数。最佳输入分布为独立等概分布。

（3）准对称离散信道的信道容量 $C=\log r-\sum\limits_{k=1}^{n}N_k\log M_k-H(p_1',p_2',\cdots,p_s')$。最佳输入分布为独立等概分布。

（4）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

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

5．N 次扩展信道

（1）若信道的输入随机矢量为 $\boldsymbol{X} = (X_1 X_2 \cdots X_N)$，通过信道传输，接收到的随机矢量为 $\boldsymbol{Y} = (Y_1 Y_2 \cdots Y_N)$。若信源与信道都是无记忆的，则 $I(\boldsymbol{X};\boldsymbol{Y}) = \sum_{l=1}^{N} I(X_l;Y_l)$。

（2）离散无记忆的 N 次扩展信道的信道容量等于原单符号离散信道的信道容量的 N 倍，即

$$C_N = NC$$

6．串联信道和并联信道

（1）对于 N 个单符号信道组成的串联信道，若其输入输出变量之间组成一个马尔可夫链，设信道矩阵分别为 $\boldsymbol{P}_1, \boldsymbol{P}_2, \cdots, \boldsymbol{P}_N$，则串联信道的总信道矩阵 $\boldsymbol{P}_{\text{总}} = \boldsymbol{P}_1 \boldsymbol{P}_2 \cdots \boldsymbol{P}_N = \prod_{i=1}^{N} \boldsymbol{P}_i$。

（2）当 N 个输入符号相互独立，且每个输入符号 X_l 的概率分布达到各个信道容量的最佳输入分布时，独立并联信道的信道容量等于各信道容量之和，即

$$C_{1,2,\cdots,N} = \sum_{l=1}^{N} C_l$$

7．信源与信道的匹配

当 R 达到信道容量 C 时，称信源与信道达到匹配，否则认为信道存在剩余。

$$\text{信道剩余度} = 1 - \frac{I(X;\ Y)}{C}$$

习　　题

3.1　设信源 X 的概率空间为

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 2 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

每个信源符号通过两条相互独立的信道同时传输，输出分别为 Y 和 Z，两个信道的转移概率模型如图 3.17 所示。计算

（1）$H(Y)$，$H(Z)$；

（2）$H(XY)$，$H(XZ)$，$H(YZ)$，$H(XYZ)$；

（3）$I(X;Y)$，$I(X;Z)$，$I(Y;Z)$；

（4）$I(X;Y|Z)$，$I(X;YZ)$。

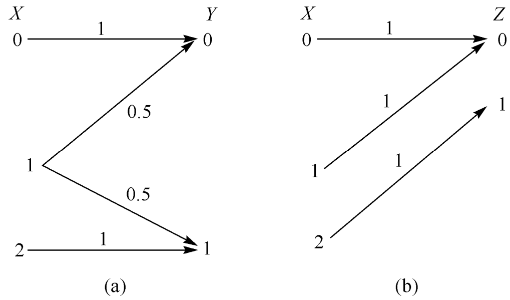

图 3.17 习题 3.1 图

3.2 设 X 和 Y 的联合分布 $P(xy)$ 为

$$P(0,0)=\frac{1}{4}，\ P(0,1)=\frac{1}{4}，\ P(1,0)=\frac{1}{2}，\ P(1,1)=0$$

试计算

（1）$H(X)$，$H(Y)$；

（2）$H(X|Y)$，$H(Y|X)$；

（3）$I(X;Y)$

（4）画出各信息量之间关系的维拉图。

3.3 甲在一个16×16的方格棋盘上随意放一枚棋子，在乙看来棋子放入哪一个位置是不确定的。如果甲告知乙棋子放入棋盘的行号，这时乙获得了多少信息量？

3.4 设有一离散无记忆信源，其概率空间为$\begin{bmatrix} X \\ P(x) \end{bmatrix}=\begin{bmatrix} a_1 & a_2 \\ 0.6 & 0.4 \end{bmatrix}$，它们通过干扰信道，信道输出端的接收符号集为$Y=\{b_1,b_2\}$，信道传递概率如图 3.18 所示。

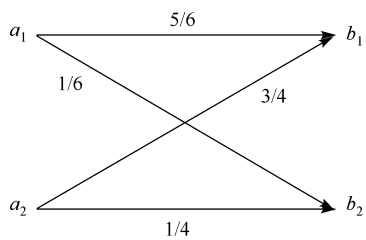

图 3.18 习题 3.4 图

求：

（1）信源 X 中事件 a_1 和 a_2 分别含有的自信息；

（2）收到信息 $b_j(j=1,2)$ 后，获得的关于 $a_i(i=1,2)$ 的信息量；

（3）$H(X)$，$H(Y)$，$H(XY)$；

（4）信道疑义度 $H(X|Y)$；

（5）噪声熵 $H(Y|X)$；

（6）收到消息 Y 后获得的平均互信息。

3.5 有一个四进制信源，每个符号发生的概率分别为 $P(a_1)=1/2$，$P(a_2)=P(a_3)=1/8$，

$P(a_4)=1/4$。试计算：

（1）信源中每个符号平均包含的信息量？

（2）信源每分钟输出 6000 个符号，信源每秒钟输出的信息量是多少？

（3）通过无噪声干扰信道传输，信宿每秒钟接收到多少信息量？

（4）如果信道损失为 0.5 比特/符号，则信宿每秒钟接收到多少信息量？

3.6　在一个二进制信道中，信源消息集 $X=\{0,1\}$，且 $P(0)=P(1)$，信宿消息集 $Y=\{0,1\}$，信道传递概率 $P(0|1)=\frac{1}{8}$，$P(1|0)=\frac{1}{4}$。求：

（1）该情况所能提供的平均信息量 $I(X;Y)$；

（2）在接收端收到 $y_1=0$ 后，所提供的关于传输消息 $x_1=0$ 的互信息 $I(x_1;y_1)$。

3.7　某学生经常缺课，现推测其原因可能是就寝时间过晚。假设 X 表示到课情况，a_1 代表缺课，a_2 代表上课；Y 表示就寝情况，b_1 代表就寝过晚，b_2 代表正常就寝。通过调查，该生的到课情况及就寝时间统计数据如下。

$$P(a_1)=0.2 \quad P(a_2)=0.8$$

$$P(a_1|b_1)=0.5 \quad P(a_2|b_1)=0.5 \quad P(a_1|b_2)=1/8 \quad P(a_2|b_2)=7/8$$

（1）事件“缺课”的自信息；

（2）计算在“就寝过晚”条件下，事件“缺课”的自信息；

（3）计算“就寝过晚”和“缺课”的互信息；

（4）试问该生缺课和就寝时间是否有关？关系如何？

3.8　设随机变量 X 和 Y 的联合概率分布为

$$\left[P(a_ib_j)\right]=\begin{bmatrix}\frac{1}{2} & 0\\ \frac{1}{4} & \frac{1}{4}\end{bmatrix}$$

其中 X 取自符号集 $\{a_1=0,a_2=1\}$，Y 取自符号集 $\{b_1=0,b_2=1\}$，已知随机变量 $Z=X+Y$，试计算：

（1）$H(X)$，$H(Y)$，$H(Z)$；

（2）$H(XY)$，$H(XZ)$，$H(YZ)$，$H(XYZ)$；

（3）$H(X|Y)$，$H(Y|X)$，$H(X|Z)$；

（4）$I(X;Y)$，$I(X;Z)$，$I(Y;Z)$。

3.9　设二进制对称信道矩阵为

$$\boldsymbol{P}=\begin{bmatrix}\frac{2}{3} & \frac{1}{3}\\ \frac{1}{3} & \frac{2}{3}\end{bmatrix}$$

（1）若信道输入符号 $P(0)=3/4$，$P(1)=1/4$，求 $H(X)$、$H(X|Y)$、$H(Y|X)$ 和 $I(X;Y)$。

（2）求该信道的信道容量及达到信道容量的最佳输入概率分布。

（3）如果信道输入符号 $P(0)=3/4$，$P(1)=1/4$，计算信道剩余度。

3.10　在一个二元对称离散信道上传输符号 0 和 1，在传输过程中每 100 个符号发生一个错误。已知 $P(0)=P(1)=\dfrac{1}{2}$，信源每秒内发出 1000 个符号，求此信道每秒最大能传输的信息量。

3.11　设某对称离散信道的信道矩阵为

$$\boldsymbol{P}=\begin{bmatrix}0 & 0.5 & 0.5 & 0\\ 0 & 0 & 0.5 & 0.5\\ 0.5 & 0 & 0 & 0.5\\ 0.5 & 0.5 & 0 & 0\end{bmatrix}$$

求其信道容量。

3.12　已知 BEC 信道矩阵为

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

求其信道容量。该信道容量是否可以表示为 $C=1-\varepsilon$ (比特/符号)?

3.13　求下列两个信道的信道容量，并加以比较。

（1）$\boldsymbol{P}_1=\begin{bmatrix}\overline{p}-\varepsilon & p-\varepsilon & 2\varepsilon\\ p-\varepsilon & \overline{p}-\varepsilon & 2\varepsilon\end{bmatrix}$　　（2）$\boldsymbol{P}_2=\begin{bmatrix}\overline{p}-\varepsilon & p-\varepsilon & 2\varepsilon & 0\\ p-\varepsilon & \overline{p}-\varepsilon & 0 & 2\varepsilon\end{bmatrix}$

式中，$p+\overline{p}=1$。

3.14　计算下列三个信道的信道容量，并画出对应的信道转移图，并指出损失熵 $H(X|Y)=0$ 或噪声熵 $H(Y|X)=0$ 的信道。

（1）$\boldsymbol{P}_1=\begin{bmatrix}1 & 0\\ 1 & 0\\ 0 & 1\\ 0 & 1\end{bmatrix}$　　（2）$\boldsymbol{P}_2=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$　　（3）$\boldsymbol{P}_3=\begin{bmatrix}\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0\\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0\\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2}\end{bmatrix}$

3.15　已知二元无记忆离散对称信道的信道矩阵为

$$\boldsymbol{P}=\begin{bmatrix}0.8 & 0.2\\ 0.2 & 0.8\end{bmatrix}$$

计算该信道的三次扩展信道的信道矩阵和信道容量。

3.16　设有扰信道的传输情况如图 3.19 所示，试求这种信道的信道容量。

3.17　若有一离散 Z 形信道，其信道转移概率如图 3.20 所示。试求：

（1）信道容量 C；

（2）$\varepsilon=0$ 和 $\varepsilon=1$ 时的信道容量。

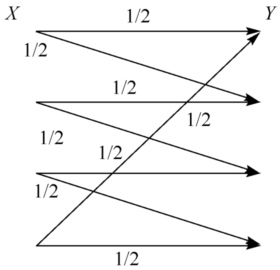

图 3.19　习题 3.16 图

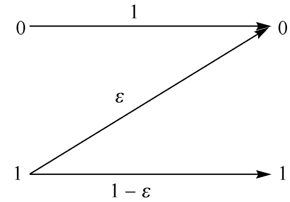

图 3.20　习题 3.17 图

3.18　设某信道的信道矩阵为

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

求其信道容量。

3.19　设某离散无记忆信道的输入 X 的符号集为 $\{0,1,2\}$，输出 Y 的符号集为 $\{0,1,2\}$，如图 3.21 所示，求其信道容量及其最佳的输入概率分布。并求当 $\varepsilon=0$ 和 $\varepsilon=1/2$ 时的信道容量 C。

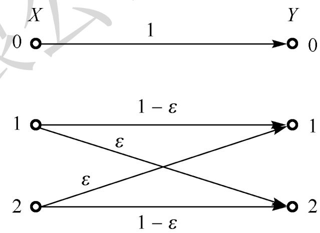

图 3.21　习题 3.19 图

3.20　若有两个串接的离散信道，它们的信道矩阵都是

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

并设第一个信道的输入符号 $X=\{a_1,a_2,a_3,a_4\}$ 是等概分布的。求 $I(X;Z)$ 和 $I(X;Y)$，并加以比较。

3.21　有两个信道的信道矩阵分别为 $\boldsymbol{P}_1=\begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ 0 & \frac{1}{2} & \frac{1}{2}\end{bmatrix}$ 和 $\boldsymbol{P}_2=\begin{bmatrix}1 & 0 & 0\\ 0 & \frac{2}{3} & \frac{1}{3}\\ 0 & \frac{1}{3} & \frac{2}{3}\end{bmatrix}$，它们的串联信道如图 3.22 所示。求证 $I(X;Z)=I(X;Y)$

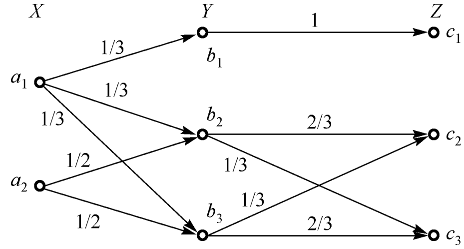

图 3.22　习题 3.21 图

3.22　把 n 个二元对称信道串联起来，信道的串联如图 3.23 所示。每个二元对称信道的错误转移概率为 p，证明这 n 个串联信道可以等效于一个二元对称信道，其错误转移概率为 $\frac{1}{2}[1-(1-2p)^n]$。证明 $\lim\limits_{n\to\infty} I(X_0;X_n)=0$，设 $p\neq 0$ 或 1。

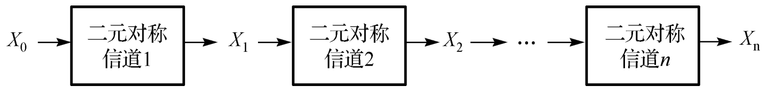

图 3.23　习题 3.22 图

3.23　证明：对于准对称离散信道，当输入等概时达到信道容量。写出信道容量的一般表达式。（提示：先证明当输入等概时 $I(x_i;Y)=$ 常数 C，根据定理 3.6 可知，常数 C 就是所求的信道容量。）

3.24　有两个信道的信道矩阵分别为 $\boldsymbol{P}_1=\begin{bmatrix}0.8 & 0.2\\0.2 & 0.8\end{bmatrix}$ 和 $\boldsymbol{P}_2=\begin{bmatrix}\frac{3}{4} & \frac{1}{4} & 0\\0 & \frac{1}{4} & \frac{3}{4}\end{bmatrix}$，它们的串联信道如图 3.24 所示。设第一个信道的输入符号 $X=\{a_1,a_2\}$ 是等概分布的，求 $I(X;Y)$、$I(Y;Z)$ 和 $I(X;Z)$，并加以比较。

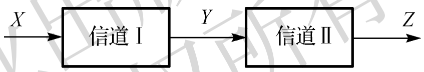

图 3.24　习题 3.24 图

3.25　某带宽为 4000Hz 的无噪信道传送 16 进制符号，该无噪信道的每秒钟最多能传输多少信息量？

3.26　某小学六年级学生人数为 500 人，其中男生有 254 人，女生有 246 人。现在要组建升旗仪仗队，身高要求大于 160cm，体检之后的数据表明，男生身高大于 160cm 的有 110 人，女生身高大于 160cm 的有 123 人。

（1）一个学生“身高大于 160cm”的体检结果提供了多少信息量？

（2）从“身高大于 160cm”的体检结果获得的关于“学生性别为女生”的信息量是多少？

（3）从体检的“身高结果”获得的关于“学生性别”的信息量是多少？

3.27　$I(a_i)\geqslant I(a_i|b_j)$ 是否成立？$H(X)\geqslant H(X|Y)$ 是否成立？

3.28　$I(a_ib_j)\leqslant I(a_i)+I(b_j)$ 是否成立？$H(XY)\leqslant H(X)+H(Y)$ 是否成立？