# 信息论与编码期末复习资料

## 1 绪论

### 1.1 信息

**信息**：对事物运动状态或存在状态的不确定性的描述。

**消息**：用文字、符号、数据、语言、音符、图片、图像等能够被人们感觉器官所感知的形式，对客观物质运动和主观思维活动状态的一种表达。

**信号**：需要将消息变换成适合于信道传输物理量，这种物理量就称为信号。

消息中包含信息，是信息的载体。同一则信息可由不同的消息形式来载荷、由不同的信号来表示；同一则消息可以载荷不同的信息；同一则信号可以表示不同的信息。

### 1.2 信息论

**信息论**是关于信息的本质和传输规律的科学的理论，是研究信息的计量、发送、传递、交换、接收和储存的一门新兴学科。

一般将信息论分成：狭义信息论（经典信息论）和广义信息论（信息科学）。

**狭义信息论**：主要研究信息的测度、信道容量以及信源和信道编码问题，是信息论的理论基础。

**广义信息论**：以信息作为主要的研究对象、以信息过程的运动规律作为主要研究内容、以信息科学方法论作为主要研究方法、以扩展人的信息功能作为主要目标的一门科学。

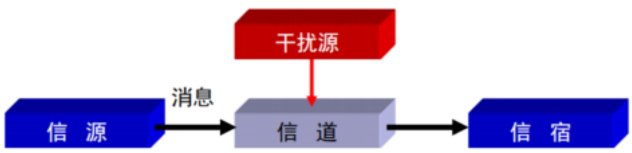
### 1.3 通信系统的模型

#### 1.3.1 通信有效性与可靠性

**有效性(Validity)**：消息若在信源中先去粗取精，则必能提高通信的有效性。对数字通信来说，即信源编码(Source Coding)要解决的主要问题。

**可靠性(Reliability)**：信宿对接收到的“消息”若能够进行判断、评估，去伪存真的处理，则必能提高通信的可靠性。对数字通信来说，即信道编码(Channel Coding) 要解决的主要问题。

#### 1.3.2 简单信道模型



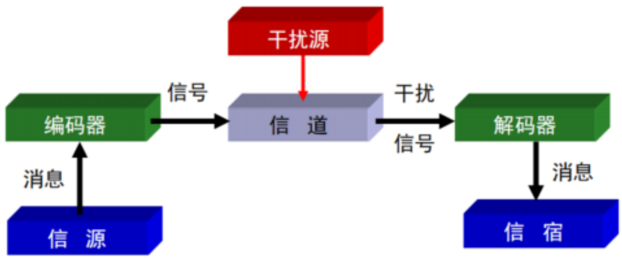

**信源(Source)**：向通信系统中提供消息的人和机器。

**信道(Channel)**：信息传递的通道，承担信息的传输和储存的任务，是构成通信系统的重要组成部分。

**信宿(Destination)**：消息传递的对象，也就是接收消息的人或机器。

**干扰源(Interference)**：通信系统中干扰的集中反映（噪声）。

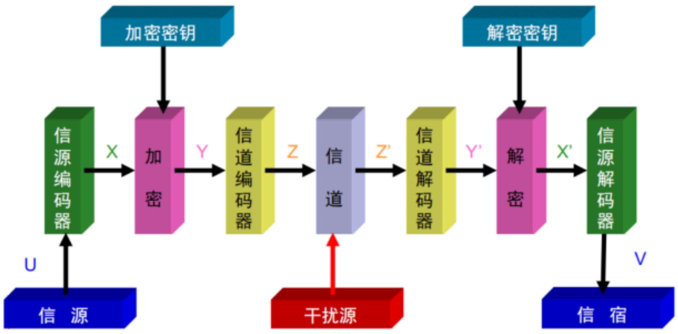
1.3.3 编码信道模型




**编码器(Encoder)**：将消息转换成为信号实现编码的设备。

**解码器(Decoder)**：与编码器相反，是将编码信号进行反变换的设备。

1.3.4 信息传输系统的模型




**信源编码器(Source Encoder)**：将信源发出的消息，变换为由二进制（或者多进制）码元组成的的代码组的设备。

**信源译码器(Source Decoder)**：将相应代码组译成信宿所能接收的消息的形式的设备。

**信道编码器(Channel Encoder)**：为了提高信息传输的可靠性而对消息进行变换和处理的设备。

**信道译码器(Channel Decoder)**：对经过信道传输后的信号进行信道编码逆变换过程的设备。

**加密编码器(Encryption Encoder)**：将信源编码器输出的明文代码组变成密文代码组的设备。

**加密密钥(Encryption Key)**：加密变换过程中使用的参数。

**解密译码器(Decryption Decoder)**：将信道译码器输出的密文代码组变成明文代码组的设备。

**解密密钥(Decryption Key)**：解密变换过程中使用的参数。

# 2 信源及信源熵

## 2.1 信源的描述和分类

消息传递的特点：**不确定性**。收信人无法判断发送人将要描述何种事物的运动状态的消息；收信人无法判断所描述的事物处于何种状态；由于干扰的存在，收信人无法判断收到的消息的正确可靠程度。

根据信源发出的消息在时间上和幅度上的分布情况，信源可分为离散信源和连续信源。

**离散信源**：时间和幅度都离散，如数字、文字、数字信号、数据等。

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix}$$

样本空间

概率分布

$$\sum_{i=1}^{6} P(X=a_i)=1,\quad P(X=a_i)\ge 0$$

**连续信源**：时间或幅度连续分布，如语音、图像、图形、模拟信号等。

$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} (a,b) \\ p(x) \end{bmatrix} \text{或} \begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} (-\infty,\infty) \\ p(x) \end{bmatrix}$$

X的取值范围

概率密度函数

$$\int_a^b p(x)\mathrm{d}x = 1 \text{或} \int_{-\infty}^{\infty} p(x)\mathrm{d}x = 1$$

**离散无记忆信源**：所发出的各个符号是相互独立的，发出符号序列中的各个符号之间没有统计关联性，各个符号的出现概率是其自身的先验概率。

**离散有记忆信源**：所发出的各个符号的概率是有关联的。

**发出单个符号的信源**：信源每次只发出一个符号代表一个消息。

**发出符号序列的信源**：信源每次发出一组含两个以上符号的符号序列代表一个消息。

- 多符号离散信源
  - 非平稳信源
  - 平稳信源
    - 无记忆信源
      - 单符号离散平稳无记忆信源
      - 符号序列离散平稳无记忆信源
    - 有记忆信源
      - 非马尔可夫信源（记忆长度无限）
      - 马尔可夫信源（记忆长度有限）

## 2.2 单符号离散信源

### 2.2.1 自信息量

自信息量是信源发出符号的先验概率的函数。

设信源输出$r$个消息$\{a_1,a_2,\dots,a_r\}$，每个消息的出现概率分别为$\{p(a_1),p(a_2),\dots,p(a_r)\}$，用随机变量$X$表

示这个信源，则其信源空间表示为：

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

其中$0 \le p(a_i) \le 1$，$\sum_{i=1}^{r} p(a_i) = 1$。

每个消息的自信息量为：

$$I(x_i) = \log_2 \left[\frac{1}{p(x_i)}\right] = -\log_2[p(x_i)]$$

其中，对数的底为 2 时，单位为 bit；为$e$时，单位为 nat；为 10 时，单位为 det。

**联合自信息量**

如果两符号$x_i$，$y_j$同时出现，可用联合概率$p(x_i y_j)$来表示，此时的联合自信息量为：

$$I(x_i y_j) = -\log p(x_i y_j)$$

**条件自信息量**

如果符号$x_i$，$y_j$不相互独立，可用条件概率$p(x_i|y_j)$来表示，此时的条件自信息量为：

$$I(x_i|y_j) = -\log p(x_i|y_j)$$

### 2.2.2 信源的信息熵

将自信息量的数学期望称为信源的平均自信息量$H(X)$，即信源$X$的信息熵。

$$H(X) = E[I(X)] = \sum_{i=1}^{r} p(x_i) I(x_i) \quad \text{(bit/symbol)}$$

信源的信息熵的物理含义：表示信源输出前，信源的平均不确定性；表示信源输出后，每个消息所提供的平均信息量；表示随机变量 X 的随机性。

熵函数$H(\boldsymbol{P})$是概率矢量$\boldsymbol{P}$的一种特殊的函数，其函数形式为：

$$H(P) = H(p_1, p_2, \dots, p_r) = -\sum_{i=1}^{r} p_i \log p_i$$

### 2.2.3 信息熵的性质

**对称性：**

$$H(X) = H(p_1, p_2, \dots, p_r) = H(p_2, p_1, \dots, p_r)$$

**确定性：**

$$H(X) = H(0, \dots, 1, \dots, 0) = 0$$

即只要有一个分量为1，其他分量必为0，熵为0。

**非负性（离散信源）：**

$$H(X)=H(p_1,p_2,\dots,p_r)\geq 0$$

对于确定信源，等号成立。

**扩展性：**

$$\lim_{\varepsilon\to 0}H_{r+1}((p_1,p_2,\dots,p_r-\varepsilon,\varepsilon)=H_r(p_1,p_2,\dots,p_r)$$

即增加一个小概率事件，信源熵保持不变。

**连续性：**

$$\lim_{\varepsilon\to 0}H_r((p_1,p_2,\dots,p_{r-1}-\varepsilon,p_r+\varepsilon)=H_r(p_1,p_2,\dots,p_r)$$

即概率分量微小变动，信源熵保持不变。

**递推性：**

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

**香农辅助定理：**

$$H(p_1,p_2,\dots,p_r)=-\sum_{i=1}^{r}p_i\log p_i\leq -\sum_{i=1}^{r}p_i\log q_i$$

即符号种数为$r$的信源$X$的熵函数，一定不大于符号数同样为$r$的另一信源中每一个符号的自信息量在信源$X$的概率空间中的统计平均值。

**最大熵定理：**

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

**可加性：**

若信源$X$和$Y$统计独立，则：

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

## 2.3 离散序列信源

### 2.3.1 离散平稳信源

设$Q$，$T$为任意时刻，若信源$X$的概率分布与时间无关，即有：

$$P(X_Q)=P(X_T),\dots,P(X_QX_{Q+1}\dots X_{Q+N})=P(X_TX_{T+1}\dots X_{T+N})$$

将信源$X$称为$N$维离散平稳信源。

### 2.3.2 离散平稳无记忆信源

如果$N$维离散平稳信源$\boldsymbol{X}=X_1X_2\dots X_N$中，各时刻随机变量$X_k$之间统计独立，则将信源$X$称为离散平稳无记忆信源；将信源$\boldsymbol{X}$称为$N$维离散平稳无记忆信源。

把$N$维离散平稳无记忆信源$\boldsymbol{X}$称为离散平稳无记忆信源$X$的$N$次扩展信源，记为$X^N=X_1X_2\dots X_N$。

$$\begin{bmatrix} X^N \\ P(\boldsymbol{X}) \end{bmatrix}=\begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{r^N} \\ p(\alpha_1) & p(\alpha_1) & \cdots & p(\alpha_{r^N}) \end{bmatrix}$$

其中$0\le p(\alpha_i)\le 1$，$\sum_{i=1}^{r^N}p(\alpha_i)=1$，$\alpha_i=a_{i1}a_{i2}\dots a_{iN}$。

$N$次扩展信源的熵：

$$H(\boldsymbol{X})=H(X^N)=-\sum_{i=1}^{r^N}p(\alpha_i)\log p(\alpha_i)\quad(\text{bit/message})$$

离散平稳无记忆信源的序列熵：

$$H(\boldsymbol{X})=H(X^N)=H(X_1X_2\dots X_N)=NH(X)$$

### 2.3.3 离散平稳有记忆信源

如果$N$维离散平稳信源$\boldsymbol{X}=X_1X_2\dots X_N$中，各时刻随机变量$X_k$之间并非统计独立，则将信源$X$称为离散平稳有记忆信源；将信源$\boldsymbol{X}$称为$N$维离散平稳有记忆信源。

**联合熵**

$$H(X_1X_2)=-\sum_{i=1}^{r}\sum_{j=1}^{r}p(a_ia_j)\log p(a_ia_j)=-\sum_{i=1}^{r}\sum_{j=1}^{r}p(a_ia_j)\log p(a_i)-\sum_{i=1}^{r}\sum_{j=1}^{r}p(a_ia_j)\log p(a_j|a_i)$$

**条件熵**

$$H(X_2|X_1)=-\sum_{i=1}^{r}\sum_{j=1}^{r}p(a_ia_j)\log p(a_j|a_i)$$

联合熵的计算：

$$H(\boldsymbol{X})=H(X_1X_2)=H(X_1)+H(X_2|X_1)=H(X_2)+H(X_1|X_2)$$

$N$维离散平稳有记忆信源的序列熵信源输出为$N$长序列时，信源的序列熵为：

$$H(\boldsymbol{X})=H(X_1X_2\dots X_N)=H(X_1)+H(X_2|X_1)+H(X_3|X_1X_2)+\cdots+H(X_N|X_1X_2\dots X_{N-1})$$

### 2.3.4 有记忆与无记忆信源熵的差别

针对二维离散平稳有记忆信源$\boldsymbol{X}=X_1X_2$和二维离散平稳无记忆信源$X^2=X_1X_2$有：

$$H(X^2)=H(X_1)+H(X_2),H(\boldsymbol{X})=H(X_1)+H(X_2|X_1)$$

所以：

$$H(\boldsymbol{X})\le H(X^2)$$

#### 2.3.5 平均符号熵

用于评估离散平稳信源$X$每发一个符号所提供的平均信息量。

$$H_N(\boldsymbol{X}) = \frac{H(X_1X_2 \dots X_N)}{N}$$

$$H_2(\boldsymbol{X}) \le H_2(X^2) = H(X)$$

### 2.4 马尔可夫信源

#### 2.4.1 马尔可夫信源的定义

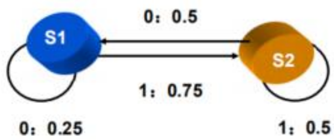
离散平稳有记忆信源$X:\{a_1a_2 \dots a_r\}$发出时间域延伸到无穷远的随机变量序列$\boldsymbol{X} = \cdots X_1X_2 \dots X_mX_{m+1} \dots$，如果$\boldsymbol{X}$中的任一时刻$(m+1)$的随机变量$X_{m+1}$只依赖于它前面已发的$m$个随机变量$X_1X_2 \dots X_m$，与更前面的随机变量无关，就将这种信源称为$m$阶马尔可夫信源。

设符号集为$X$和状态为$S$。信源输出的信息符号还与信源所处的状态有关。每一时刻信源发出一个符号后，所处的状态将发生转移。

时刻1到时刻$m$的随机变量序列提供的某一具体消息$S_i = \{a_{k_1}, a_{k_2}, \dots, a_{k_m}\}$可以看成是状态$S_i$；时刻2到时刻$m+1$的随机变量序列提供的某一具体消息$S_j = \{a_{k_2}, a_{k_3}, \dots, a_{k_{m+1}}\}$可以看成是状态$S_j$。若信源处于某一状态$S_i$，当它发出一个符号$a_{k_{m+1}}$后，所处的状态就变为$S_j$；任何时刻信源处在什么状态完全由前一时刻的状态和发出的符号决定。

马尔可夫链的状态转移图：每个圆圈代表一种状态，状态之间的有向线代表某一状态向另一状态的转移。有向线一侧的符号和数字分别代表发出的符号和条件概率。

$$P = \begin{bmatrix} 0.25 & 0.75 \\ 0.5 & 0.5 \end{bmatrix}$$




#### 2.4.2 马尔可夫信源的数学描述

信源输出依赖长度为$m+1$的随机序列就转化为对应的状态序列，这种状态序列符合简单的马尔可夫链的性质。

确定离散平稳有记忆信源$X$的符号集，再确定$m$阶数$(m>0)$。

$m$阶马尔可夫离散信源的数学模型可由一组信源符号集和一组条件概率确定：

$$\begin{bmatrix} X \\ P(X_{m+1}) \end{bmatrix} = \begin{bmatrix} a_1, a_2, \dots, a_r \\ p\left(a_{k_{m+1}} | a_{k_1}a_{k_2} \dots a_{k_m}\right) \end{bmatrix}$$

满足：

$$\sum_{km+1=1}^{r} p\left(a_{k_{m+1}} | a_{k_1}a_{k_2} \dots a_{k_m}\right) = 1$$

② 给定$r^{2m}$个由任一消息（状态）$S_i$到下一个消息（状态）$S_j$的：

一步转移概率：

$$p\left(a_{k_{m+1}}|a_{k_1}a_{k_2}\dots a_{k_m}\right)=p\left(a_{k_{m+1}}|S_i\right)=p\left(S_j|S_i\right)$$

$n$步转移概率：

$$p\{S(m+n)=j\mid S(m)=i\}=p_{ij}^{(n)}(m)$$

③ 状态转移矩阵：由于系统在任一时刻处于状态空间$S:\{S_1,S_2,\dots,S_{r^m}\}$中任意一个状态，将给定的状态一步转移概率$p\left(S_j|S_i\right)$按照对应关系，排成一个状态一步转移矩阵：

$$\boldsymbol{P}=\{p_{ij}\}=\begin{bmatrix}p_{11} & \cdots & p_{1r^m}\\ \vdots & \ddots & \vdots\\ p_{r^m1} & \cdots & p_{r^mr^m}\end{bmatrix},\boldsymbol{P}^{(n)}=[\boldsymbol{P}]^n$$

### 2.4.3 状态极限概率与极限熵

状态极限概率与$m$阶$M$信源极限熵：

$$H_\infty=\sum_{i=1}^{r^m}\sum_{j\in J}p(S_i)p\left(S_j|S_i\right)\log p\left(S_j|S_i\right)=\sum_i p(S_i)H(X|S_i)$$

状态概率$p(S_i)$是$m$阶$M$信源达到稳定时出现状态$S_i$的概率，称为状态极限概率（稳态分布概率$W_i$）。状态极限概率可由下式计算出：

$$\boldsymbol{WP}=\boldsymbol{W},\sum_{i=1}^{r^m}W_i=1,\boldsymbol{W}=(W_1,W_2,\dots,W_{r^m})$$

### 2.4.4 各态遍历性

有限齐次马尔可夫信源的各态遍历性：对于有限平稳的$m$阶$M$信源，由信源$X:\{a_1,a_2,\dots,a_r\}$的$m$个符号组成的消息(状态)$S_i=\left\{a_{i_1},a_{i_2},\dots,a_{i_m}\right\}$和消息(状态)$S_j=\left\{a_{j_1},a_{j_2},\dots,a_{j_m}\right\}$存在一个正整数$n\geq 1$,且经过$n$步，从状态$S_i$转移到状态$S_j$的$n$步转移概率$p_{ij}^{(n)}>0$，则这种$m$阶$M$信源具有各态遍历性。

对于具有遍历性的$m$阶$M$信源，对于每个$j=1,2,\dots,r^m$都存在不依赖于起始状态$S_i$的状态极限概率$\lim_{n\to\infty}p_{ij}^{(n)}=p\left(S_j\right)$，而且状态极限概率是在约束条件$p\left(S_j\right)>0$，$\sum_{j=1}^{r^m}p\left(S_j\right)=1$的约束下方程组$p\left(S_j\right)=\sum_{j=1}^{r^m}p(S_i)p_{ij}$的唯一解。

**不可约性**

如果状态空间$S:\{S_1,S_2,\dots,S_{r^m}\}$中任意两个状态$S_i,S_j$都存在至少一个正整数$k$，使得$p_{ij}^{(k)}>0$，即从状态$S_i$开始，总有可能转移到另一状态$S_j$，状态空间中各个状态之间，都能相互到达，那么状态空间$S$组成的集合称为不可约闭集。

**非周期性**

如果状态空间$S:\{S_1,S_2,\dots,S_{r^m}\}$是一个不可约闭集， 并且从每一状态出发，经过$k$步转移回到本状态的

所有可能的步数$k$中，即使$p_{ij}^{(k)} > 0$的所有$k$中，不存在大于1的公因子，则称不可约闭集$S:\{S_1, S_2, \dots, S_{r^m}\}$具有非周期性。

马尔可夫信源各态遍历性的判断的两种方法：

① $n$步状态转移矩阵$\boldsymbol{P}^{[n]}$中，不存在"0"元素，则可判断具有各态遍历性；如果$n$为任意正整数的$\boldsymbol{P}^{[n]}$中，都有"0"元素，则不具有各态遍历性。

② 状态转移图中的状态集合是不可约非周期闭集，则可判断具有各态遍历性；若不是不可约非周期闭集，则判断不具有各态遍历性。

### 2.5 相对熵率和冗余度

为了衡量信源符号之间的依赖程度，将离散平稳有记忆信源的极限熵$H_\infty$，与把这个信源当作离散平稳无记忆等概率信源达到的最大熵值$H_0 = \log r$的比值，定义为这个离散平稳有记忆信源的相对熵率（信息效率）：

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

将1减去相对熵率$\eta$所得之差定义为离散平稳有记忆信源的剩余度（冗余度）：

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

## 3 信道与信道容量

### 3.1 信道

信道：传送信息的载体—信号所通过的通道。信息是抽象的，信道则是具体的。

信道的作用：在信息系统中信道主要用于传输与存储信息；而在通信系统中则主要用于传输。

信道的分类：无噪声信道、有噪声信道。

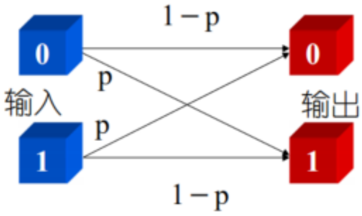
研究信道的目的：在通信系统中研究信道，主要是为了描述、度量、分析不同类型信道，计算其容量，即极限传输能力，并分析其特性。

### 3.2 信道的数学模型

#### 3.2.1 二进制对称信道(BSC, Binary Symmetric Channel)

输入符号集合$X = \{0, 1\}$；

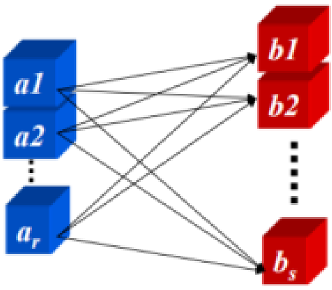
输出符号集合$Y = \{0, 1\}$。




3.2.2 离散无记忆信道(DMC, Discrete Memory-less Channel)

输入符号集合$X = \{a_1, a_2, \dots, a_r\}$；

输出符号集合$Y = \{b_1, b_2, \dots, b_s\}$。




输入输出特性由$r \times s$种条件概率描述：

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

条件概率$p(b_j|a_i)$体现了信道对输入符号$a_i$的传递作用，也将条件概率称为**传递概率**。

信道的总体传递特性，由随机变量$X$和随机变量$Y$的$r \times s$个条件概率$p(b_j|a_i)$构成的**转移概率矩阵**表示。

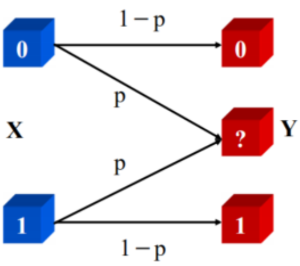
$$\boldsymbol{P} = \begin{bmatrix} p(b_1|a_1) & \cdots & p(b_s|a_1) \\ \vdots & \ddots & \vdots \\ p(b_1|a_r) & \cdots & p(b_s|a_r) \end{bmatrix}$$

由于转移概率矩阵完全描述了离散无记忆信道的传递特性，又称为**信道矩阵**。

3.2.3 二进制删除信道(BEC, Binary Erasure Channel)

输入符号集合$X = \{0, 1\}$；

输出符号集合$Y = \{0, ?, 1\}$。




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

## 3.3 互信息量和平均互信息量

### 3.3.1 输入输出符号间的统计特性

① 输入符号$a_i$，输出符号$b_j$的联合概率；

$$p(a_i,b_j)=p(a_i)p(b_j|a_i)$$

② 输出符号$b_j$的概率；

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

③ 输出符号$b_j$后，推测输入符号为$a_i$的后验概率。

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

### 3.3.2 互信息量

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

互信息量$I(a_i;b_j)$表示信宿收到$b_j$前后，对信源发$a_i$的不确定性的消除。

先验不确定性：$I(a_i)$；后验不确定性：$I(a_i|b_j)$。

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

### 3.3.3 先验概率和后验概率对互信息量的影响

① 当$p(a_i|b_j)=1$时，$I(a_i;b_j)=\log\frac{1}{p(a_i)}=I(a_i)$；

② 当$p(a_i)<p(a_i|b_j)<1$时，$I(a_i;b_j)=\frac{\log p(a_i|b_j)}{p(a_i)}>0$；

③ 当$p(a_i|b_j)=p(a_i)$时，$I(a_i;b_j)=\frac{\log p(a_i|b_j)}{p(a_i)}=0$；

④ 当$p(a_i|b_j)<p(a_i)<1$时，$I(a_i;b_j)=\frac{\log p(a_i|b_j)}{p(a_i)}<0$。

### 3.3.4 平均互信息量

信道每传递一个符号所传输的平均互信息量$I(X;Y)$等于互信息量$I(a_i;b_j)$在$X$和$Y$的联合概率空间$P(XY)$中的统计平均值，也就是：

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

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

其中$H(X),H(Y)$分别表示发送和接受的信息量，$I(X;Y)$表示信道中传递的信息量，$H(X|Y)$表示疑义度，$H(Y|X)$表示噪声熵。

### 3.3.5 平均互信息量的性质

① 非负性：$I(X;Y)\geq 0$；

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

即疑义度≤输入熵、噪声熵≤信宿熵；

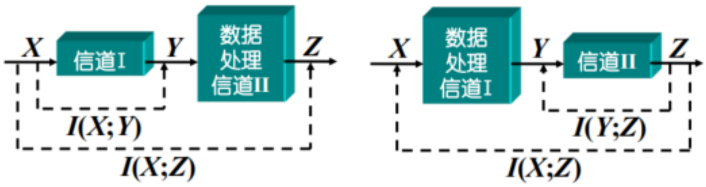
② 对称性：$I(X;Y)=I(Y;X)$；

③ 极值性：$I(X;Y)\leq H(X),I(X;Y)\leq H(Y)$；

④ 上凸性。

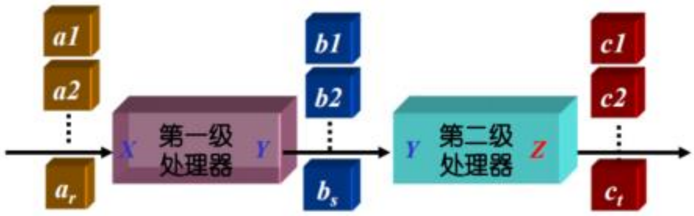
## 3.4 数据处理中信息的变化

### 3.4.1 条件互信息量相关互信息量




如果随机变量$XYZ$构成一个马尔可夫链，则有以下关系成立，并且仅当$p(x|yz)=p(x|z)$或 $p(z|xy)=p(z|x)$时，等号成立：

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




在已知事件$c_k$给定的条件下，收到$b_j$后得到关于某事件$a_i$的**条件互信息量**为：

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

在已知事件$b_j,c_k$以后，总共获得的关于事件$a_i$的**相关互信息量**为：

$$I(a_i;b_jc_k)=\log\frac{p(a_i|b_jc_k)}{p(a_i)}=I(a_i;b_j)+I(a_i;c_k|b_j)$$

在已知事件$c_k$以后，总共获得的关于事件$a_ib_j$的**相关互信息量**为：

$$I(a_ib_j;\ c_k)=I(a_i;c_k)+I(b_j|a_i;c_k)=I(a_i;c_k)+I(b_j;c_k|a_i)=I(b_j;c_k)+I(a_i;c_k|b_j)$$

### 3.4.2 平均条件互信息量和平均相关互信息量

将上述互信息在概率空间$XYZ$中求统计平均，得到**平均条件互信息量**：

$$I(X;\ Y|Z)=\sum_X\sum_Y\sum_Z p(a_ib_jc_k)\,I(a_i;b_j|c_k)$$

**平均相关（平均联合）互信息量**：

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

$$I(XY;\ Z)=I(X;\ Z)+I(Y;\ Z|X)=I(Y;\ Z)+I(X;\ Z|Y)$$

$$I(X;\ Z)=I(X;\ Y)+I(X;\ Z|Y)-I(X;\ Y|Z)=I(Y;\ Z)+I(X;\ Z|Y)-I(Y;\ Z|X)$$

### 3.4.3 信息处理定理

信息处理定理（信息不增性原理）：任何无源数据处理都不会增加信息量。

## 3.5 信道容量

### 3.5.1 信道容量的定义

信道的信息传输率$R$，即平均互信息量，指信道中平均每个符号所能传送的信息量，即：

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

给定转移概率矩阵$P$以后，平均互信息量$I(X;Y)$是信道输入变量$X$的概率分布$P(X)$的上凸函数.

对于一切可能的输入信号概率分布而言，信道传输率$I(X;Y)$所能达到的最大值，称为信道容量：

$$C=\max_{p(x)}\{I(X;Y)\}\quad (\text{bit/symbol})$$

信道容量是信道最大传输能力，是信道自身的特性。

如果已知一个信道传递一个符号需要$t$秒钟时间，则信道每秒钟能传递的平均互信息量称为信道的信息传输速率：

$$R_t=\frac{I(X;Y)}{t}\quad (\text{bit/s})$$

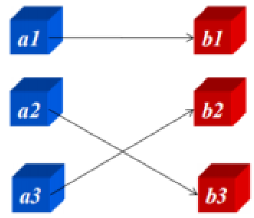
信道的最大信息传输速率：

$$R_{t\max}=\frac{C}{t}=\max_{p(x)}\left\{\frac{I(X;Y)}{t}\right\}\quad (\text{bit/s})$$

### 3.5.2 特殊信道的信道容量计算方法

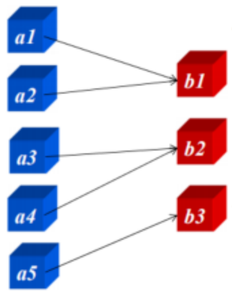
假设输入信源$X$的符号共有$r$个，输出信源$Y$的符号共有$s$个。

**无噪无损信道（一一对应关系）**



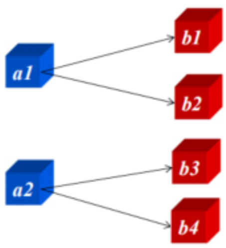

$$C=\log r \quad \text{(bit/symbol)}$$

**无噪有损信道（多个输入对应一个输出）**




$$C=\log s \quad \text{(bit/symbol)}$$

**有噪无损信道（一个输入多个输出，每个输入对应的输出值不重合）**




$$C=\log s \quad \text{(bit/symbol)}$$

**对称 DMC(Discrete Memoryless Channel)信道的容量（输入输出都对称）**

如果转移概率矩阵$\boldsymbol{P}$的每一行都是由同一符号集$\{p_1,p_2,\dots,p_s\}$诸元素的不同排列组成，每一列也是由同一集合$\{q,q_2,\dots,q_r\}$诸元素的不同排列组成，将这种 DMC 信道称为**对称 DMC 信道**。

$$C=\max_{p(a_i)}\{H(Y)-H(Y|X)\}=\max_{p(a_i)}H(Y)-H(Y|a_i)$$

对称 DMC 信道的输入符号等概率分布时，达到其信道容量，为：

$$C=\log s+\sum_{j=1}^{s}p_{ij}\log p_{ij}$$

**强对称 DMC 信道 BSC 信道(Binary Symmetric Channel)**

如果单符号离散对称信道的输入符号集$X:\{a_1,a_2,\dots,a_n\}$，输出符号集$Y:\{a_1,a_2,\dots,a_r\}$每一输入符号的正

确传递概率均为$(1-\varepsilon)$，总的错误传输概率均匀分配在其它$(r-1)$个错误传输概率上，可得到信道矩阵是$r \times r$阶对称矩阵，该信道称为**强对称信道**。

$$P=\begin{bmatrix} (1-\varepsilon) & \frac{\varepsilon}{r-1} & \cdots & \frac{\varepsilon}{r-1} \\ \frac{\varepsilon}{r-1} & (1-\varepsilon) & \cdots & \frac{\varepsilon}{r-1} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\varepsilon}{r-1} & \frac{\varepsilon}{r-1} & \cdots & (1-\varepsilon) \end{bmatrix}$$

$$C=\max_{p(x)}\{H(Y)-H(Y|X)\}=\log r-H(\varepsilon)-\varepsilon\log(r-1)$$

**准对称信道**

如果信道矩阵的每一行元素都是其他每行的置换，且可以分为$m$个$r \times s_l$阶子矩阵$P_l$，$(l=1,2,\dots,m)$，子矩阵的行列都具有排列性，则由$m$个$r \times s_l$阶子矩阵$P_l$组成的$r$行$s$列矩阵所代表的信道，称为**准对称的信道**。

$$P=\begin{bmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \end{bmatrix} \qquad P=\begin{bmatrix} 1/4 & 1/2 & 1/8 & 1/8 \\ 1/2 & 1/4 & 1/8 & 1/8 \end{bmatrix}$$

$$P_1=\begin{bmatrix} 1/4 & 1/2 \\ 1/2 & 1/4 \end{bmatrix} \qquad P_2=\begin{bmatrix} 1/8 & 1/8 \\ 1/8 & 1/8 \end{bmatrix}$$

$$C=\max_{p(x_i)}\{H(Y)-H(Y|X)\}=\max_{p(x_i)}H(Y)-H(Y|X)=\max_{p(x_i)}H(Y)-H(p_1',p_2',\dots,p_s')$$

可以证明当输入等概分布时，可以达到信道容量：

$$C=\log r-\sum_{k=1}^{m}N_k\log M_k-H(p_1',p_2',\dots,p_s')$$

其中$N_k$是$m$个子矩阵中第$k$个子矩阵中行元素之和，$M_k$是第$k$个子矩阵中列元素之和。