信息论与编码

Information Theory and Coding

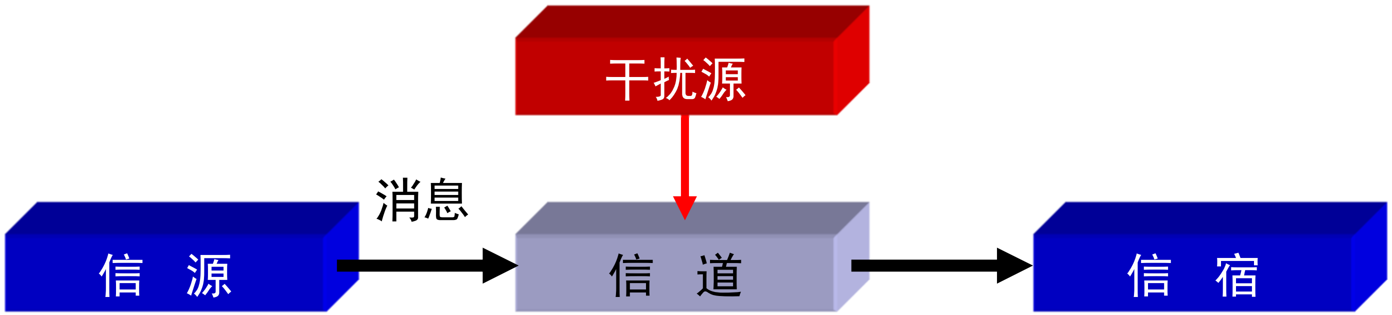
西南交通大学

电子信息工程专业

2020

第3章 信道与信道容量

3-1 引言



简单信道模型

【1】什么是信道?

信道是传送信息的载体—信号所通过的通道。

信息是抽象的，信道则是具体的。

【2】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。

【3】信道的分类

- 无噪声信道—信宿收到信源发出的全部消息，获取全部信息量(可等效为信源本身)。
- 有噪声信道—传递作用，体现信源信宿间的噪声随机干扰作用。

【4】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**。

【5】研究信道的主要问题

- 信道的建模（信道的统计特信的描述）
- 信道的传输信息的能力（信道容量）的计算
- 在有噪声的信道中能否实现可靠传输?
- 如何在有噪声的信道中实现可靠传输?

3-2 信道的数学模型

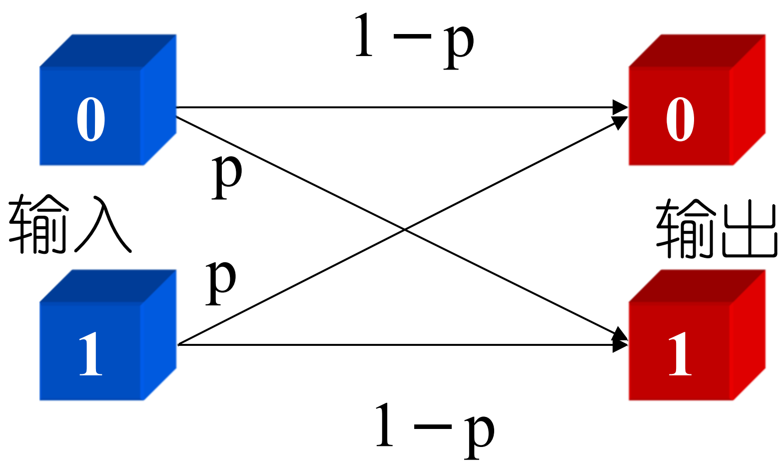
3.2.1 二进制对称信道

(BSC Binary Symmetric Channel)

输入符号集合$X=\{0,1\}$

输出符号集合$Y=\{0,1\}$

构成最常用的二进制单消息对称信道



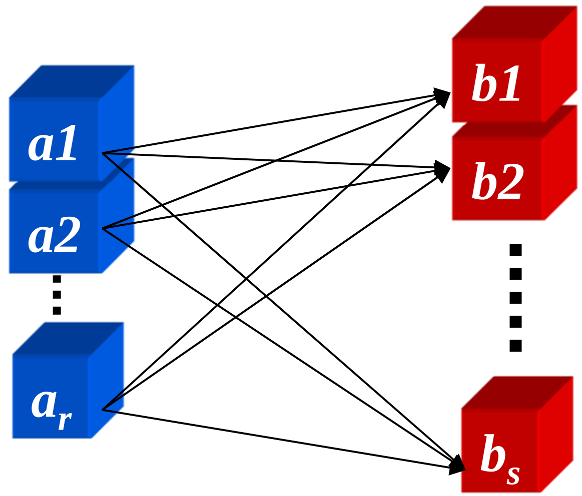
3-2 信道的数学模型

3.2.2 离散无记忆信道

（ DMC Discrete Memory-less Channel ）

输入符号集$X:\{a_1,a_2,\cdots,a_r\}$

输出符号集$Y:\{b_1,b_2,\cdots,b_s\}$



输入输出特性由$r\times s$种

条件概率描述：

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

3-2 信道的数学模型

a）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。

b）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。

转移概率矩阵

$$\mathbf{P}=\begin{bmatrix} p(b_1/a_1) & p(b_2/a_1) & \cdots & p(b_s/a_1) \\ p(b_1/a_2) & p(b_2/a_2) & \cdots & p(b_s/a_2) \\ \vdots & \vdots & \vdots & \vdots \\ p(b_1/a_r) & p(b_2/a_r) & \cdots & p(b_s/a_r) \end{bmatrix}$$

3-2 信道的数学模型

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**。

信道矩阵具有以下特性：

[1] $0 \le p(b_j / a_i) \le 1$

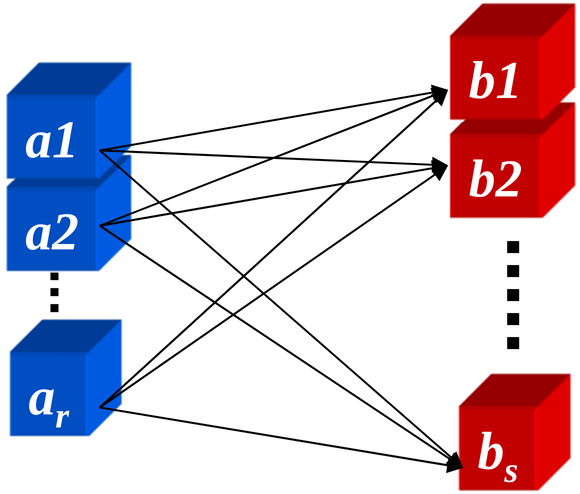
当$p(b_j / a_i)=0$时，

表示输入a_i时不可能输出b_j

当$p(b_j / a_i)=1$时，

表示输入a_i输出b_j是必然事件

[2] $\sum_{j=1}^{s} p(b_j / a_i) = 1$



信道传递图

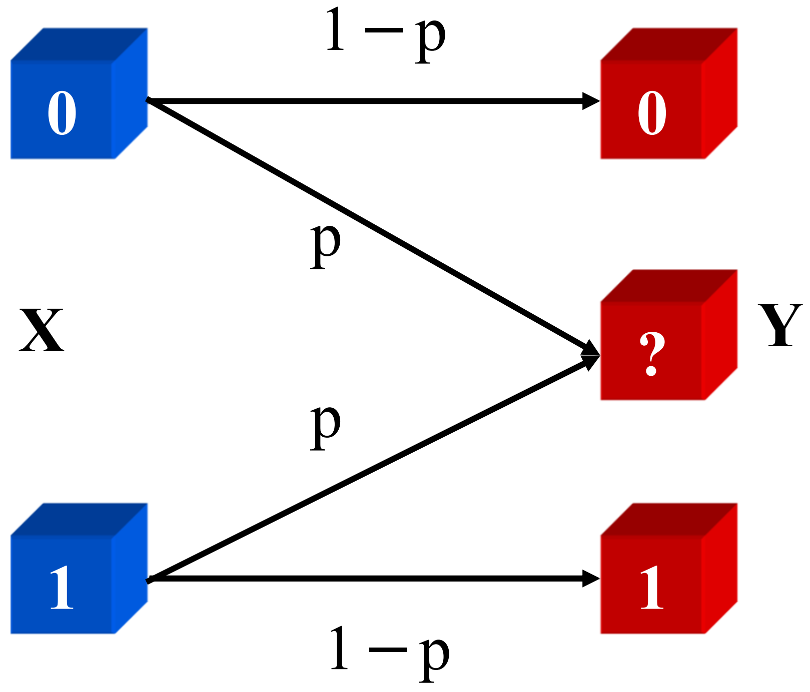
3-2 信道的数学模型

例题： 二进制删除信道(Binary Erasure Channel)

输入符号集合X={0,1}

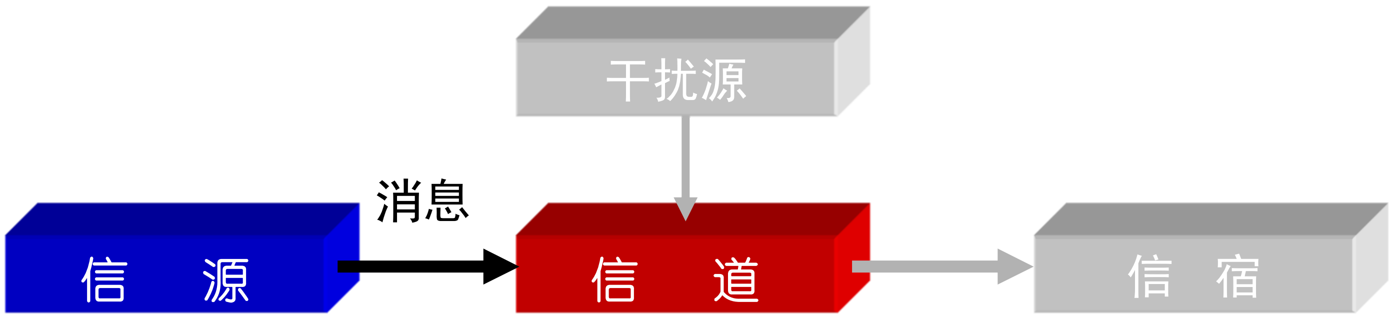
输出符号集合Y={0,?,1}

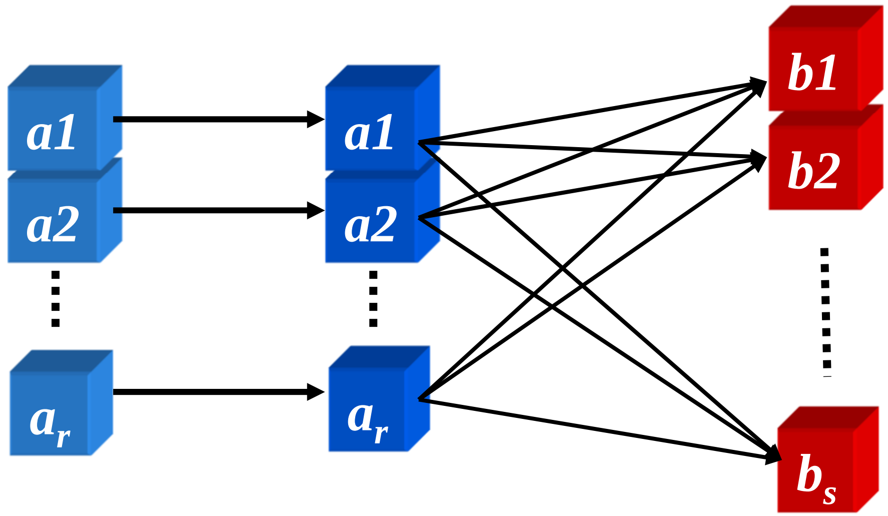
$$[\mathbf{P}]=\begin{bmatrix}1-p & p & 0\\ 0 & p & 1-p\end{bmatrix}$$



3-3 互信息量和平均互信息量

3.3.1 互信息量(Mutual Information)





3-3 互信息量和平均互信息量

离散信源X的信源空间为：

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

给定信道矩阵为：

$$\mathbf{P} = \begin{bmatrix} p(b_1/a_1) & p(b_2/a_1) & \cdots & p(b_s/a_1) \\ p(b_1/a_2) & p(b_2/a_2) & \cdots & p(b_s/a_2) \\ \vdots & \vdots & \vdots & \vdots \\ p(b_1/a_r) & p(b_2/a_r) & \cdots & p(b_s/a_r) \end{bmatrix}$$

3-3 互信息量和平均互信息量

【一】输入输出符号间的统计特性表示为：

[1] 输入符号 a_i，输出符号b_j的联合概率；

[2] 输出符号b_j的概率；

[3] 输出符号b_j后，推测输入符号为 a_i的后验概率；

3-3 互信息量和平均互信息量

[1]"输入符号a_i,输出符号b_j"的联合概率

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

$$=p(a_i)p(b_j/a_i)$$

其中：

$p(a_i)$是信源X的“先验概率”分布

$p(b_j/a_i)$是信道的传递概率，也称为“前向概率”

3-3 互信息量和平均互信息量

[2] “输出符号b_j” 的概率

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

$$\begin{aligned}&\begin{bmatrix}p(b_1) & p(b_2) & \cdots & p(b_s)\end{bmatrix}\\ &=\begin{bmatrix}p(a_1) & p(a_2) & \cdots & p(a_r)\end{bmatrix}\cdot\mathbf{P}\end{aligned}$$

$$\begin{bmatrix}p(b_1)\\ p(b_2)\\ \vdots\\ p(b_s)\end{bmatrix}=\mathbf{P}^T\begin{bmatrix}p(a_1)\\ p(a_2)\\ \vdots\\ p(a_r)\end{bmatrix}$$

3-3 互信息量和平均互信息量

[3] "输出符号b_j后，推测输入符号a_i" 的后验概率

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

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

并且$\sum_{i=1}^{r}p(a_i/b_j)=1$

$p(a_i/b_j)$称为"输出b_j后，推测输入符号a_i"的后验概率

课堂小测

信宿收到b_j后，从b_j中获取的关于信源发出a_i的信息量

信宿收到b_j前，对信源发出a_i的先验不确定性

信宿收到b_j后，对信源发出a_i仍然存在的后验不确定性

信宿收到b_j前后，对信源发a_i的不确定性的消除

全概率公式、贝叶斯概率公式

3-3 互信息量和平均互信息量

【二】互信息量

定义：

$I(a_i;b_j)$：信宿收到b_j后，从b_j中获取的关于信源发出a_i的信息量

$I(a_i)$：信宿收到b_j前，对信源发出a_i的先验不确定性

$I(a_i/b_j)$：信宿收到b_j后，对信源发出a_i仍然存在的后验不确定性

$$I(a_i;b_j)=[I(a_i)-I(a_i/b_j)]$$

$=$信宿收到b_j前后，对信源发a_i的不确定性的消除

3-3 互信息量和平均互信息量

先验不确定性

$I(a_i)=$信源发a_i先验概率$p(a_i)$的倒数的对数；

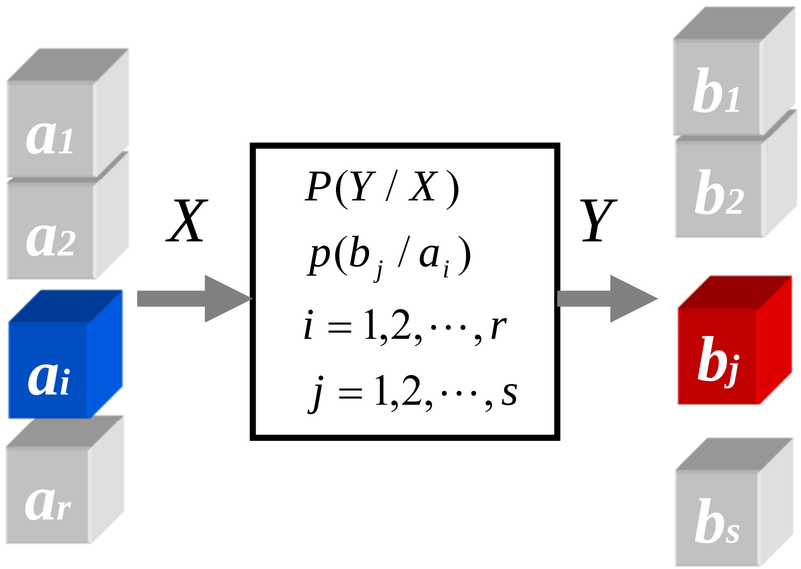
后验不确定性

$I(a_i/b_j)=$后验概率$p(a_i/b_j)$的倒数的对数；

$$I(a_i;b_j)=[I(a_i)-I(a_i/b_j)]$$

$$=\log\frac{1}{p(a_i)}-\log\frac{1}{p(a_i/b_j)}$$

$$=\log\frac{p(a_i/b_j)}{p(a_i)}$$

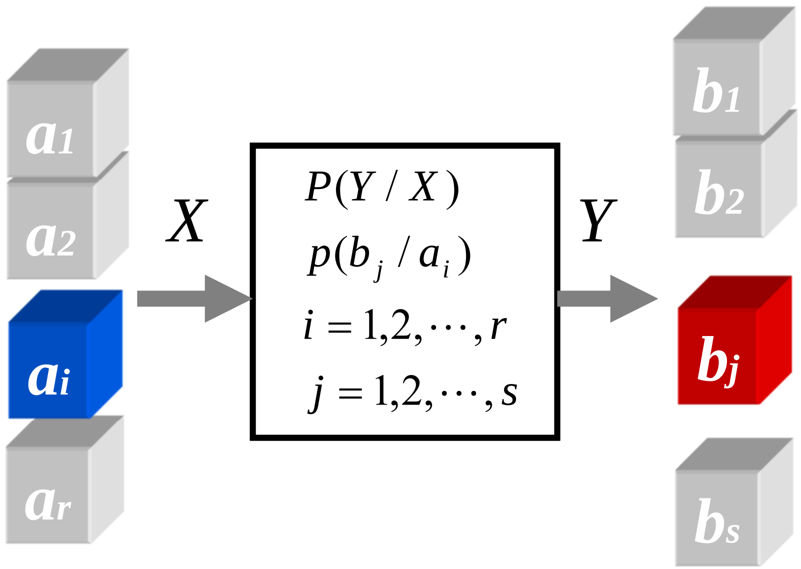


3-3 互信息量和平均互信息量

$$I(a_i;b_j)=[I(a_i)-I(a_i/b_j)]=\log\frac{p(a_i/b_j)}{p(a_i)}$$

$$=\log\frac{p(a_i/b_j)p(b_j)}{p(a_i)p(b_j)}=\log\frac{p(a_i,b_j)}{p(a_i)p(b_j)}$$

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

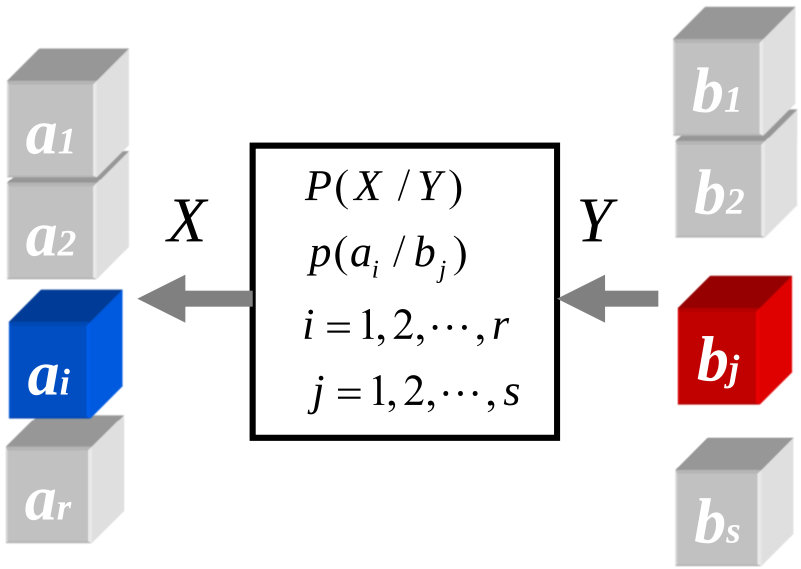


3-3 互信息量和平均互信息量

$$I(a_i;b_j)=[I(a_i)-I(a_i/b_j)]=\log\frac{p(a_i,b_j)}{p(a_i)p(b_j)}$$

$$=\log\frac{p(b_j/a_i)p(a_i)}{p(a_i)p(b_j)}=\log\frac{p(b_j/a_i)}{p(b_j)}$$

$$=\log\frac{1}{p(b_j)}-\log\frac{1}{p(b_j/a_i)}$$

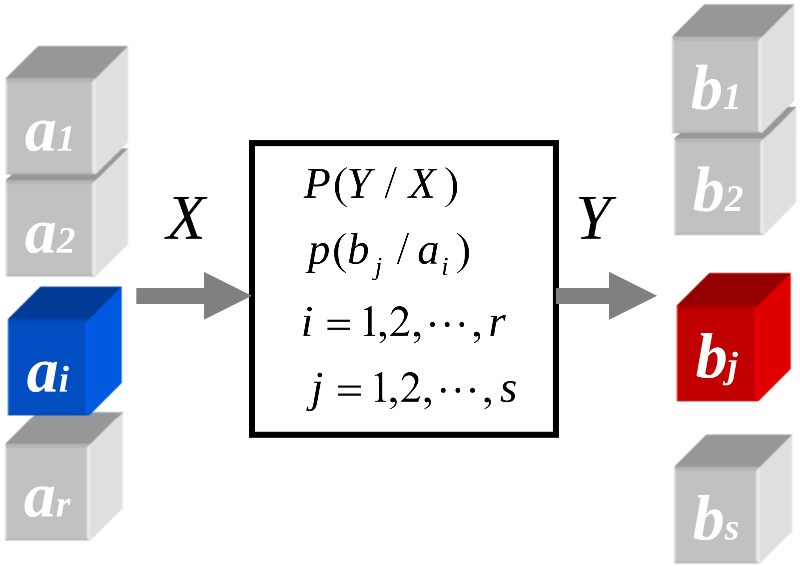


3-3 互信息量和平均互信息量

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

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

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



显然$I(a_i;b_j)=I(b_j;a_i)$

说明了由传递概率$P(X/Y)$、$P(Y/X)$联系的随机变量X,Y从b_j中获取的关于a_i的信息量$I(a_i;b_j)$与从a_i中获取的关于b_j的信息量$I(b_j;a_i)$，“你中有我，我中有你”且二者相等因此称$I(a_i;b_j)$和$I(b_j;a_i)$为a_i、b_j之间的互信息量。

3-3 互信息量和平均互信息量

【三】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，$I(a_i;b_j)=\log\dfrac{1}{p(a_i)}=I(a_i)$

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

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

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

3-3 互信息量和平均互信息量

【三】先验概率和后验概率对互信息量的影响

互信息量的定义及其定量估算公式是解决信息度量的基本出发点，其中的自信息量及其信息函数只是互信息量中当后验概率等于1时的特例。互信息量将信息传输理论推向定量分析的范畴，为在深层次上揭示信息传递规律奠定了基础。

3-3 互信息量和平均互信息量

3.3.2 平均互信息量

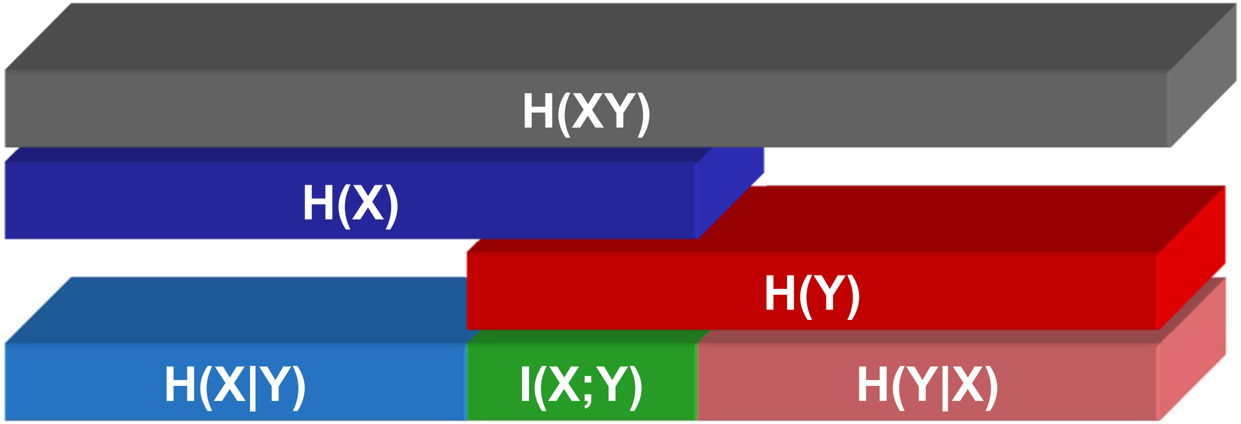
定义：信道每传递一个符号所传输的平均互信息量 $I(X;Y)$ 应该等于互信息量 $I(a_i;b_j)$ 在 X 和 Y 的联合概率空间 $P(XY)$ 中的统计平均值，也就是

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

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

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

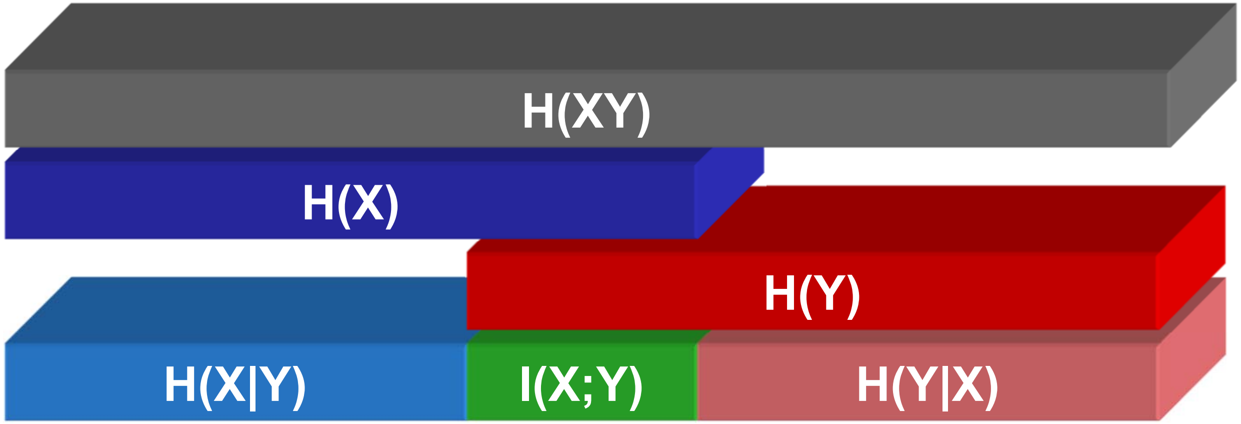
3-3 互信息量和平均互信息量



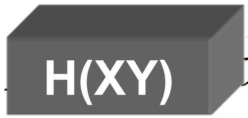
(1) $I(X;Y)=H(X)-H(X/Y)$

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，称为**疑义度**。

3-3 互信息量和平均互信息量

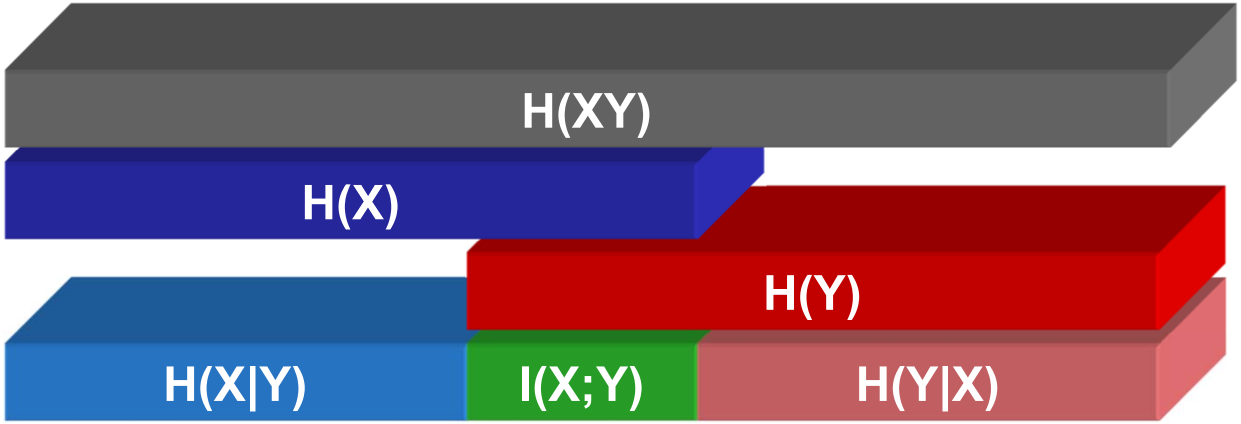


$$(2)\ I(X;Y)=H(X)+H(Y)-H(XY)$$



表示随机变量X通过信道传递在输出端输出Y，信道两端同时出现X、Y平均不确定性，称为**共熵**。

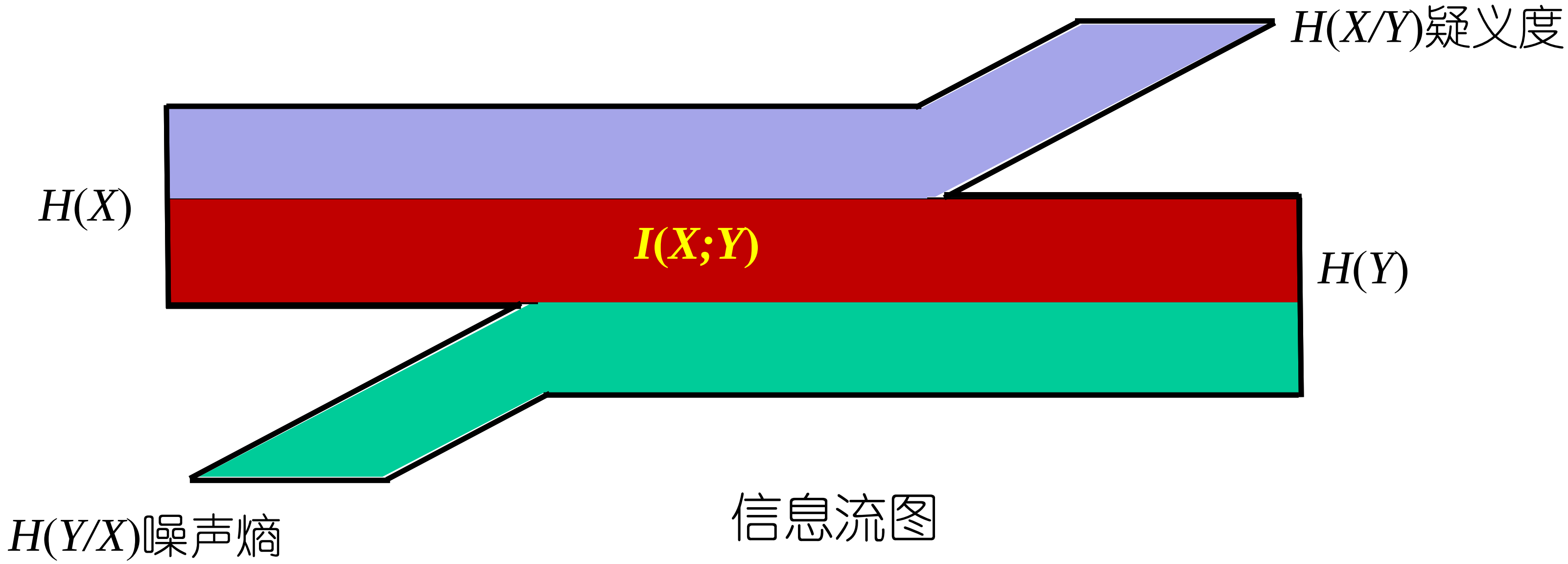
3-3 互信息量和平均互信息量



(3) $I(X;Y)=H(Y)-H(Y/X)$

表示反向信道输出随机变量X后，对输入随机变量Y仍然存在的平均不确定性，称为**噪声熵**。

3-3 互信息量和平均互信息量



信息流图

$H(X)$表示发送的信息量

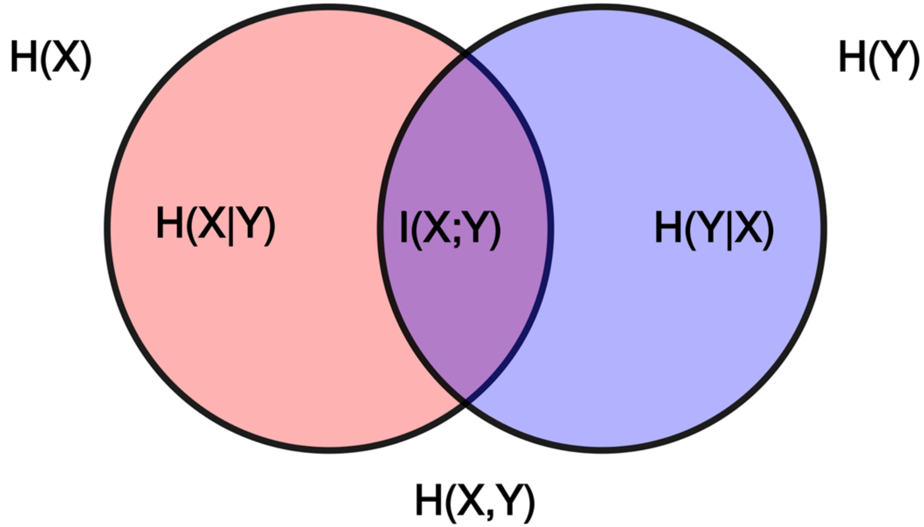
$H(Y)$表示接收的信息量

$I(X;Y)$表示信道中传递的信息量

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

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

3-3 互信息量和平均互信息量



$H(X)$表示发送的信息量

$H(Y)$表示接收的信息量

$I(X;Y)$表示信道中传递的信息量

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

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

3-3 互信息量和平均互信息量

例题：掷骰子，如果结果是1，2，3或4，则掷一次硬币；如果结果是5或6则掷两次硬币，试计算从抛掷硬币的结果可以得到多少掷骰子的信息量.

解：根据抛掷硬币出现正面的次数Y来获得关于掷骰子结果X的信息。

设掷骰子的结果是1,2,3,4的事件为X＝0，

结果是5, 6　的事件为X=1；

随机变量Y＝0表示抛硬币出现正面0次，

Y＝1，Y=2分别表示出现正面1次和2次；

可见，随机变量X的概率空间

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 2/3 & 1/3 \end{bmatrix}$$

3-3 互信息量和平均互信息量

条件概率矩阵 $\mathbf{P} = P(Y/X) = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}$

随机事件Y的概率分布为：

$$P(Y) = P(X)\mathbf{P} = [2/3 \quad 1/3]\begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/4 \end{bmatrix} = \begin{pmatrix} \frac{5}{12} & \frac{1}{2} & \frac{1}{12} \end{pmatrix}$$

因此，Y的信息熵 $H(Y) = -\sum_{j=1}^{3} p(y_j)\log p(y_j) = 1.325 bit/sym$

噪声熵 $H(Y/X) = -\sum_{i=1}^{2}\sum_{j=1}^{3} p(x_i)p(y_j/x_i)\log p(y_j/x_i)$

$$= -\frac{2}{3}\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) - \frac{1}{3}\left(\frac{1}{4}\log\frac{1}{4} + \frac{1}{2}\log\frac{1}{2} + \frac{1}{4}\log\frac{1}{4}\right) = 1.166 bit/sym$$

因此：$I(X;Y) = H(Y) - H(Y/X) = 0.159 bit/sym$

3-3 互信息量和平均互信息量

平均互信息的性质

【1】非负性 $I(X;Y) >= 0$

根据凸函数的性质：$f\left(\sum_{i=1}^{r} p_i x_i\right) \geq \sum_{i=1}^{r} p_i f(x_i)$

即定义域中r个变量在由r个概率分量构成的完备的概率空间中的统计平均值，大于或等于这r个变量的函数值在这一概率空间中的统计平均值。

$$-I(X;Y) = \sum_{i}^{r}\sum_{j}^{s} p(x_i y_j)\log\frac{p(x_i)p(y_j)}{p(x_i y_j)}$$

$$\leq \log\sum_{i}^{r}\sum_{j}^{s} p(x_i y_j)\frac{p(x_i)p(y_j)}{p(x_i y_j)}$$

$$= \log\sum_{i}^{r}\sum_{j}^{s} p(x_i)p(y_j) = 0$$

3-3 互信息量和平均互信息量

由于$I(X;Y)\geq 0$，因此由

$$I(X;Y)=H(X)-H(X/Y)\geq 0$$

$$\Rightarrow H(X/Y)\leq H(X)$$ 疑义度≤输入熵

$$I(X;Y)=H(X)+H(Y)-H(XY)\geq 0$$

$$\Rightarrow H(XY)\leq H(X)+H(Y)$$

$$I(X;Y)=H(Y)-H(Y/X)\geq 0$$

$$\Rightarrow H(Y/X)\leq H(Y)$$ 噪声熵≤信宿熵

即 **有条件的熵小于等于无条件的熵**

3-3 互信息量和平均互信息量

【2】对称性 $I(X;Y) = I(Y;X)$

$$
\begin{aligned}
I(X;Y) &= \sum_{i}^{r}\sum_{j}^{s} p(x_i y_j)\log\frac{p(x_i / y_j)}{p(x_i)} \\
&= \sum_{i}^{r}\sum_{j}^{s} p(x_i y_j)\log\frac{p(x_i y_j)}{p(x_i)p(y_j)} \\
&= \sum_{i}^{r}\sum_{j}^{s} p(x_i y_j)\log\frac{p(y_j / x_i)}{p(y_j)} \\
&= I(Y;X)
\end{aligned}
$$

3-3 互信息量和平均互信息量

【3】极值性 $I(X;Y)<=H(X)$； $I(X;Y)<=H(Y)$

由于 $I(X;Y)=H(X)-H(X/Y)=H(Y)-H(Y/X)$

考虑到条件熵$H(X/Y)$和$H(Y/X)$的非负性，

即由于$0\le p(a_ib_j)\le 1, 0\le p(a_i/b_j)\le 1 \Rightarrow p(a_ib_j)\log p(a_i/b_j)\le 0$

则$H(X/Y)=-\sum_{i=1}^{r}\sum_{j=1}^{s}p(a_ib_j)\log p(a_i/b_j)\ge 0$

因此可有以下两种表现形式：

$$I(X;Y)=H(X)-H(X/Y)\le H(X)$$

$$I(X;Y)=H(Y)-H(Y/X)\le H(Y)$$

可以得到$I(X;Y)\le H(X)$； $I(X;Y)\le H(Y)$

3-3 互信息量和平均互信息量

【3】极值性 $I(X;Y) <= H(X)$; $I(X;Y) <= H(Y)$

说明：从一个事件获得 的关于另一个事件得到信息量，至多只能是另一个事件的平均自信息量，不会超过事件本身所含有的信息量。

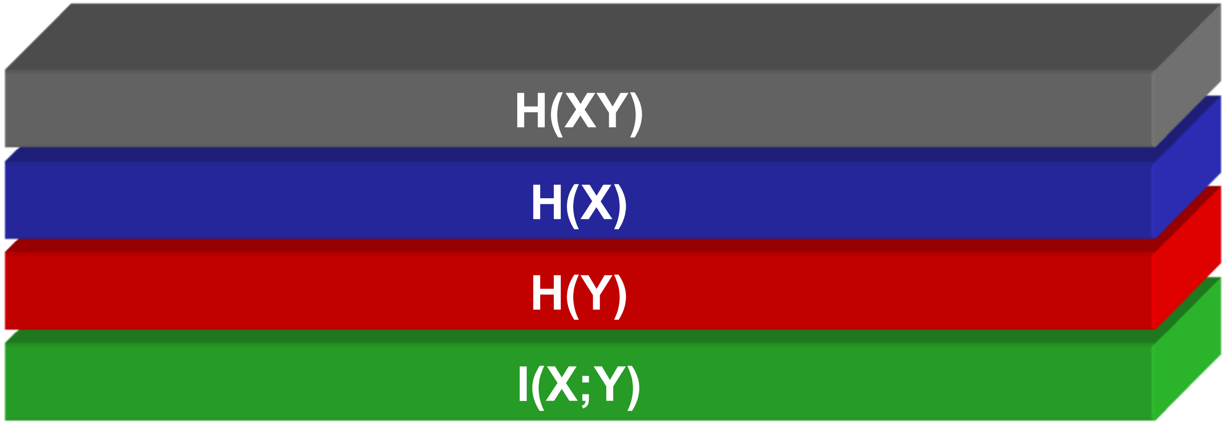
最好情况：通信后 $I(X;Y) = H(X) = H(Y)$ 无损信道

最差情况：通信后 $I(X;Y) = 0$ 全损信道

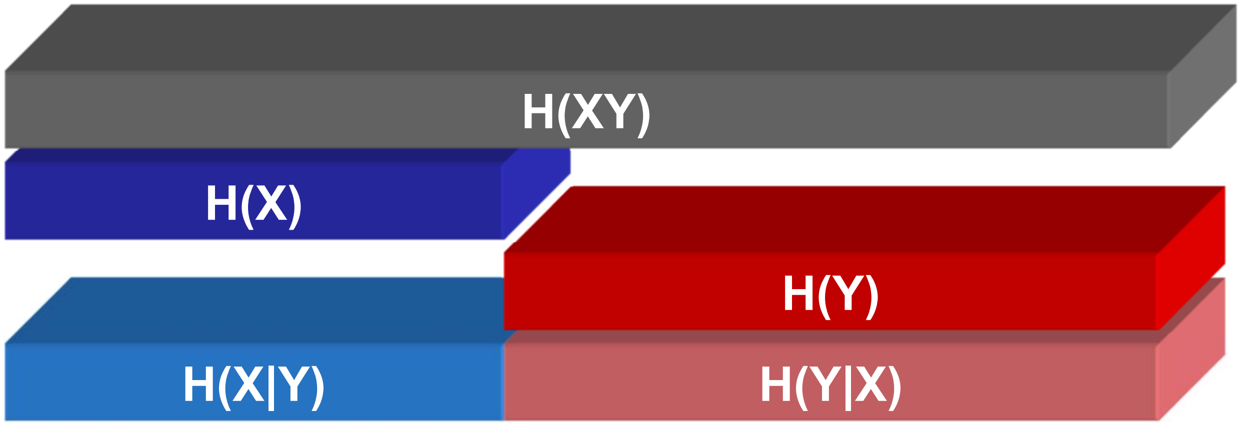
【4】上凸性 对于传递概率给定的信道，其平均交互信息量$I(X;Y)$是输入信源X的概率分布$P(X)$上凸函数

3-3 互信息量和平均互信息量

无损信道



全损信道



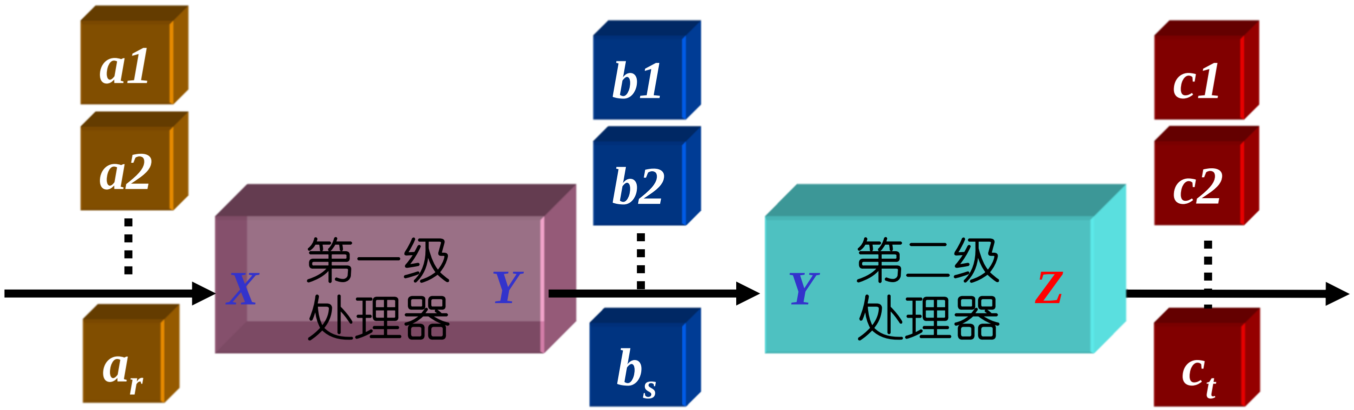
课堂测试：

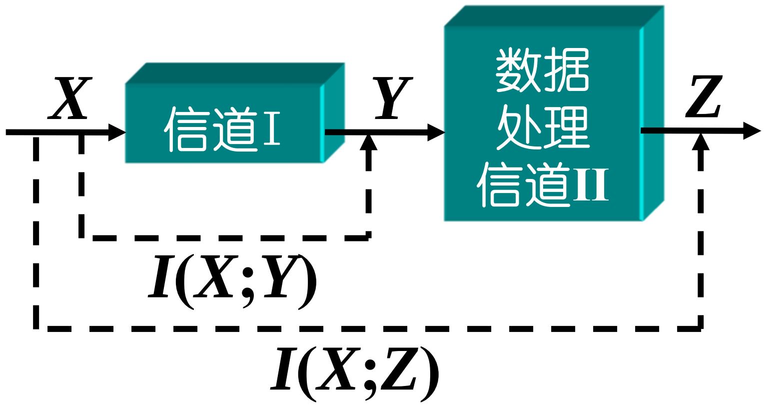
已知信源空间和信道矩阵分别为

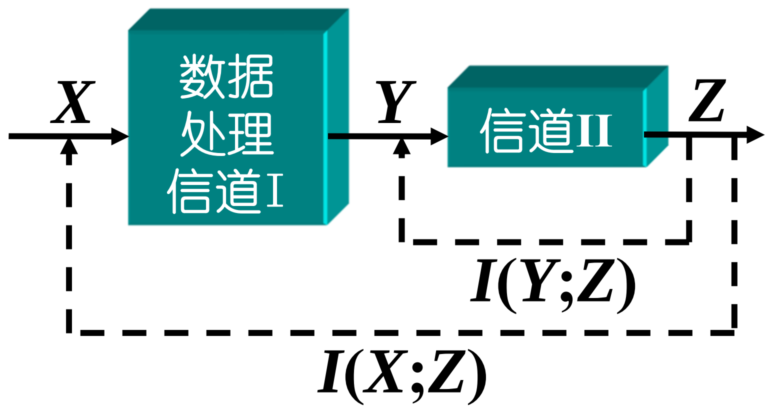
$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 = 0 & x_2 = 1 \\ 1/2 & 1/2 \end{bmatrix} \qquad \mathbf{P} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$$

求平均互信息量 I(X;Y)

3-4 数据处理中信息的变化







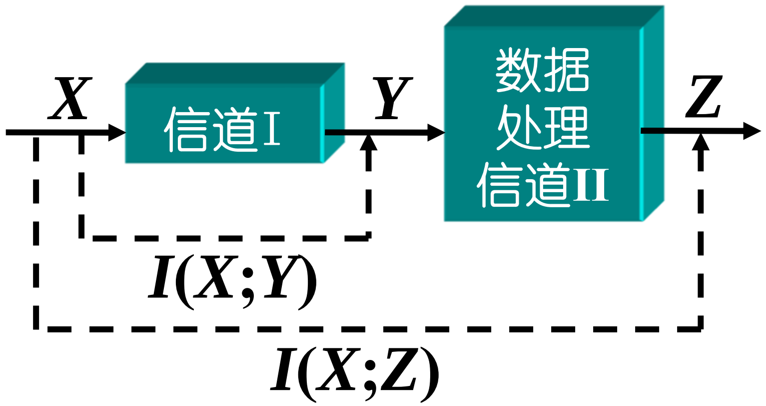
$I(X;Z)$? $I(X;Y)$　　　　$I(X;Z)$? $I(Y;Z)$

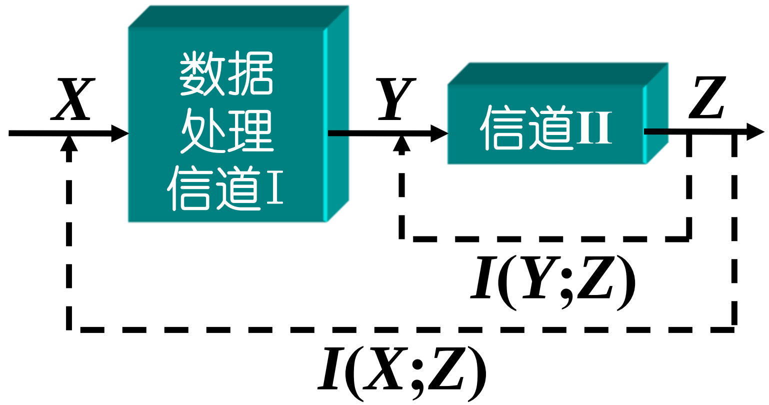
3-4 数据处理中信息的变化

定理： 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。

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

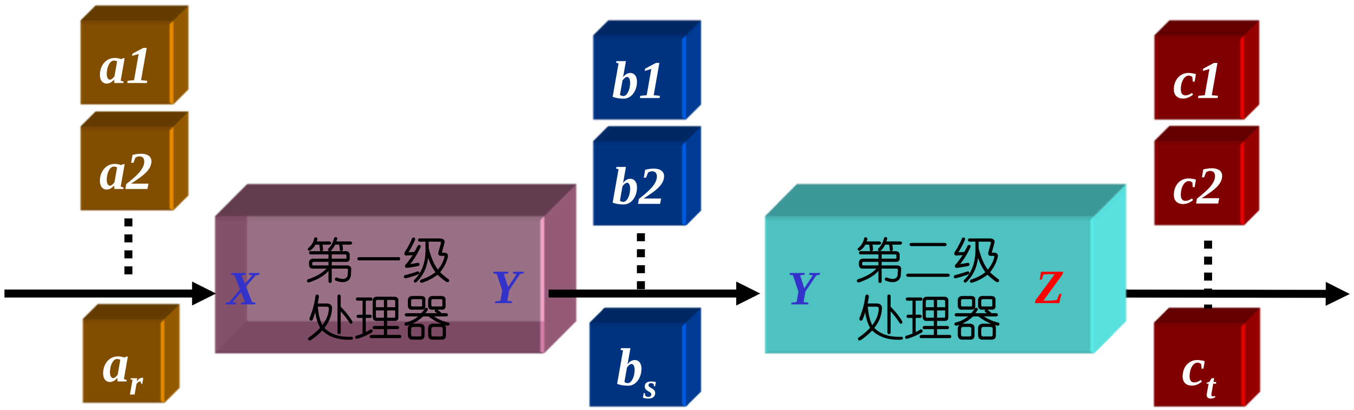
$$I(X;Z) \le I(Y;Z)$$





$$I(X;Z)\ ?\ I(X;Y) \qquad\qquad I(X;Z)\ ?\ I(Y;Z)$$

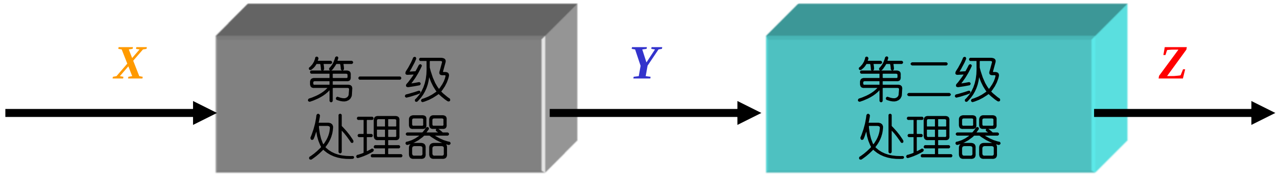
3-4 数据处理中信息的变化



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：$I(a_i;b_j/c_k)=\log\dfrac{p(a_i/b_jc_k)}{p(a_i/c_k)}$

3-4 数据处理中信息的变化



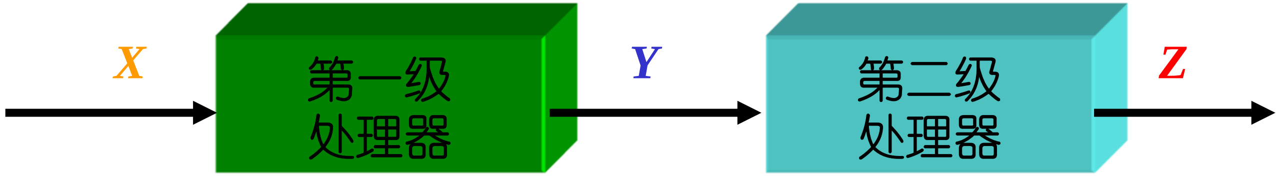
在已知事件c_k给定的条件下，收到b_j后得到关于某事件a_i的条件互信息为：$I(a_i;b_j/c_k)=\log\dfrac{p(a_i/b_jc_k)}{p(a_i/c_k)}$

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

$$I(a_i;b_j/c_k)=\log\frac{1}{p(a_i/c_k)}+\log\frac{1}{p(b_j/c_k)}-\log\frac{1}{p(a_ib_j/c_k)}$$

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

3-4 数据处理中信息的变化



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：

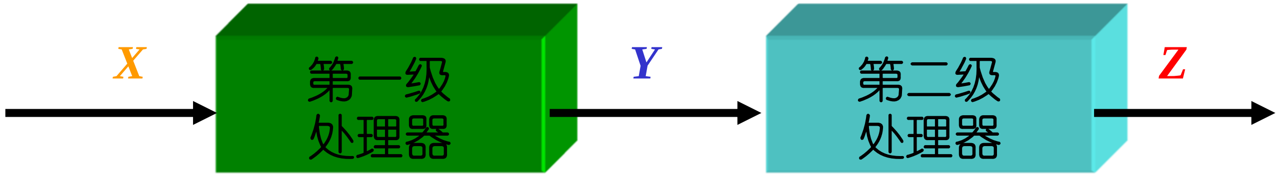
$$I(a_i;b_jc_k)=\log\frac{p(a_i/b_jc_k)}{p(a_i)}=\log\frac{p(a_i/c_k)p(a_i/b_jc_k)}{p(a_i)p(a_i/c_k)}$$

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

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

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

3-4 数据处理中信息的变化



在已知事件c_k以后，获得的关于事件a_ib_j的**相关互信息量**为：

$$I(a_ib_j;c_k)=\log\frac{p(a_ib_j/c_k)}{p(a_ib_j)}=\log\frac{p(a_ib_jc_k)}{p(a_ib_j)p(c_k)}$$

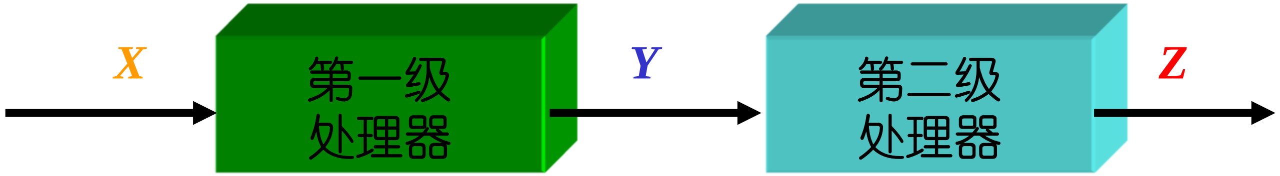
$$=\log\frac{p(c_k/a_ib_j)p(a_ib_j)}{p(a_ib_j)p(c_k)}=\log\frac{p(c_k/a_ib_j)p(c_k/b_j)}{p(c_k)p(c_k/b_j)}$$

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

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

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

3-4 数据处理中信息的变化



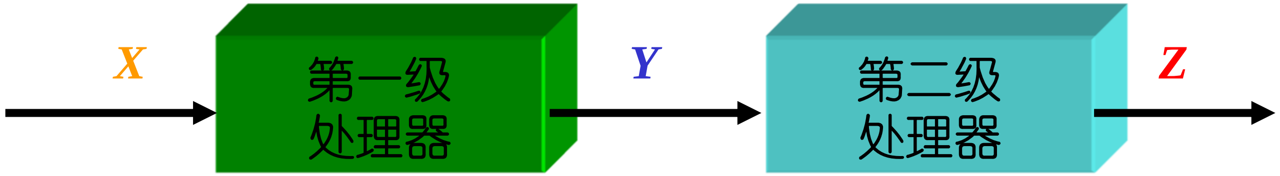
在已知事件c_k以后，获得的关于事件a_ib_j的相关互信息为：

$$I(a_ib_j;c_k)=\log\frac{p(a_ib_j/c_k)}{p(a_ib_j)}=\log\frac{p(a_ib_jc_k)}{p(a_i)p(b_j/a_i)p(c_k)}$$

$$=\log\frac{p(a_ic_k)p(b_j/a_ic_k)}{p(a_i)p(b_j/a_i)p(c_k)}=\log\frac{p(c_k)p(a_i/c_k)p(b_j/a_ic_k)}{p(a_i)p(b_j/a_i)p(c_k)}$$

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

3-4 数据处理中信息的变化



在已知事件c_k以后，获得的关于事件a_ib_j的相关互信息为：

$$I(a_ib_j;c_k)=\log\frac{p(a_ib_j/c_k)}{p(a_ib_j)}=\log\frac{p(a_ib_jc_k)}{p(b_j)p(a_i/b_j)p(c_k)}$$

$$=\log\frac{p(b_jc_k)p(a_i/b_jc_k)}{p(b_j)p(a_i/b_j)p(c_k)}=\log\frac{p(c_k)p(b_j/c_k)p(a_i/b_jc_k)}{p(b_j)p(a_i/b_j)p(c_k)}$$

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

3-4 数据处理中信息的变化

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

$$I(X;YZ)=\sum_{X}\sum_{Y}\sum_{Z}p(a_ib_jc_k)I(a_i;b_jc_k)$$

$$I(XY;Z)=\sum_{X}\sum_{Y}\sum_{Z}p(a_ib_jc_k)I(a_ib_j;c_k)$$

3-4 数据处理中信息的变化

得到
$$I(X;YZ)=I(X;Y)+I(X;Z/Y)$$
$$=I(X;Z)+I(X;Y/Z)$$

得到
$$I(XY;Z)=I(X;Z)+I(Y;Z/X)$$
$$=I(Y;Z)+I(X;Z/Y)$$

最终得到

$$I(X;Z)=I(X;Y)+I(X;Z/Y)-I(X;Y/Z)$$

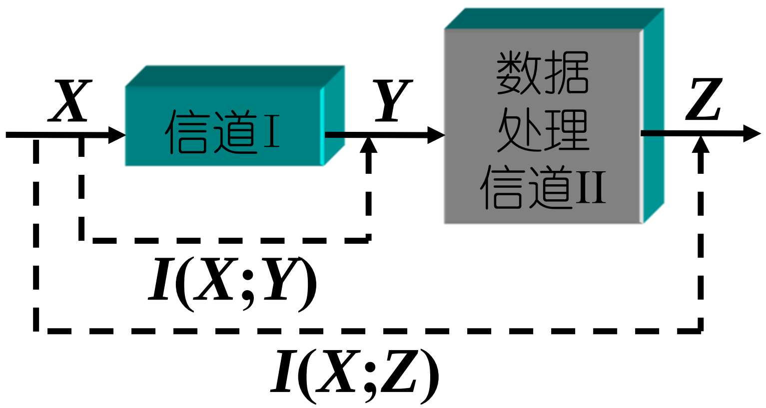
$$I(X;Z)=I(Y;Z)+I(X;Z/Y)-I(Y;Z/X)$$

3-4 数据处理中信息的变化

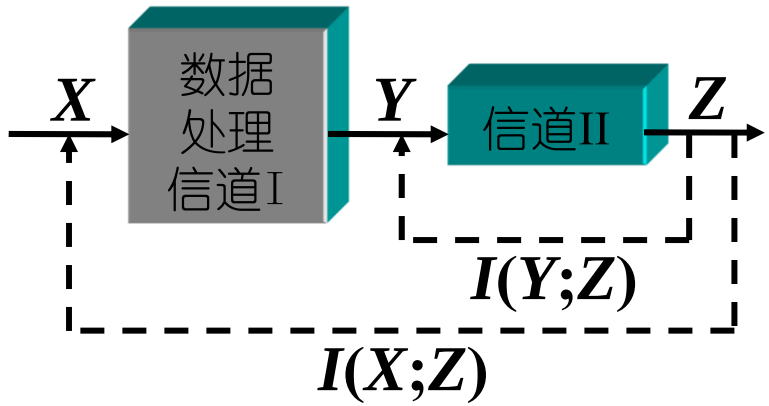
根据平均互信息量的非负性和XYZ为马尔可夫链的性质（在Y条件下X和Z相互独立），则$I(X;Z/Y)=0$，于是得到

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

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



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



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

3-4 数据处理中信息的变化

定理： 当消息通过多级处理器时，随着处理器数目的增加，输入消息与输出消息之间的平均互信息量趋于减小，丢失的信息可能越多。

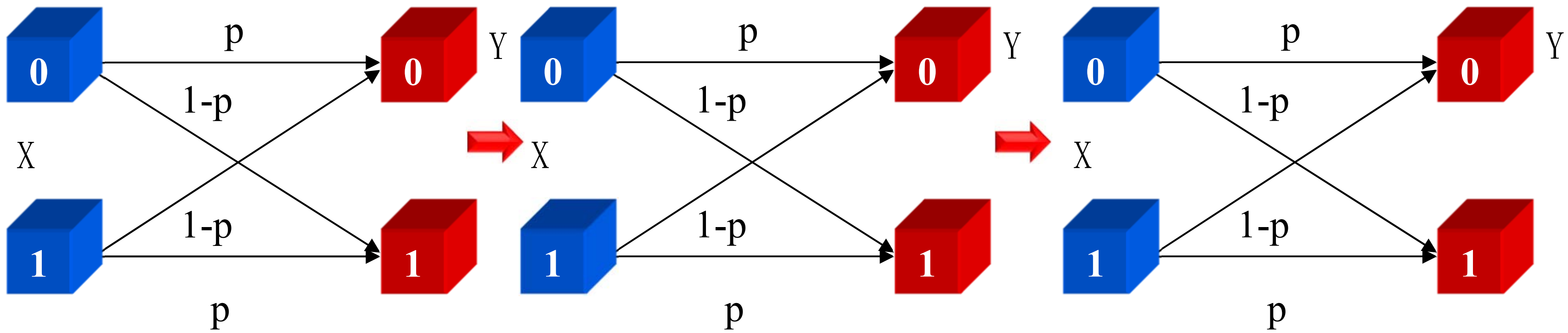
也可表述为：

任何无源数据处理都不会增加信息量（信息不增性原理）。

将这个结论称为**信息处理定理。**

3-4 数据处理中信息的变化

例题： 设有一个二进制对称离散信道，串接组成下图所示串接信道，随机变量序列X-Y-Z-W是Markov链，如果信源X：{0,1}是等概率信源，试求平均交互信息量I(X;Y),I(X;Z),I(X;W)并比较其大小。



3-4 数据处理中信息的变化

解题：

1、由题可知信源空间和信道I的信道矩阵分别为：

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 = 0 & x_2 = 1 \\ 1/2 & 1/2 \end{bmatrix} \qquad \mathbf{P} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$$

可知输出符号Y的先验概率为：

$$\left[p(y_1 = 0), p(y_2 = 1) \right] = [p(x_1 = 0), p(x_2 = 1)] \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix} = \left[\frac{1}{2}, \frac{1}{2} \right]$$

因此,

$$H(Y) = -\sum_{j=1}^{2} p(y_j) \log p(y_j) = H(\frac{1}{2}, \frac{1}{2}) = 1bit / sym$$

3-4 数据处理中信息的变化

由信道I信道矩阵P，可知

$$H(Y/X) = -\sum_{i=1}^{2}\sum_{j=1}^{2} p(x_i)p(y_j/x_i)\log p(y_j/x_i)$$

$$= -\frac{1}{2}\big(p\log p + (1-p)\log(1-p)\big) - \frac{1}{2}\big(p\log p + (1-p)\log(1-p)\big)$$

$$= H(p)bit/sym$$

因此，信道I符号间的平均互信息量为：

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

$$= H(\frac{1}{2},\frac{1}{2}) - H(p)$$

$$= 1 - H(p)bit/sym$$

3-4 数据处理中信息的变化

由信道间符号为马尔科夫链，故信道I和II的串接信道（X-Z）间的信道矩阵为PI与PII的连乘，即：

$$\mathbf{P}_{I,II}=\mathbf{P}_{I}\bullet\mathbf{P}_{II}=\begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}\bullet\begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$$

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

信道II输出符号Z的概率分布：

$$\left[p(z_1=0),p(z_2=1)\right]=[p(x_1=0),p(x_2=1)]\mathbf{P}_{I,II}=\left[\frac{1}{2},\frac{1}{2}\right]$$

符号Z的信息熵：

$$H(Z)=H(\frac{1}{2},\frac{1}{2})=1bit/sym$$

3-4 数据处理中信息的变化

由信道I和II的串接信道矩阵可得:

$$H(Z/X) = -\sum_{i=1}^{2}\sum_{k=1}^{2} p(x_i)p(z_k/x_i)\log p(z_k/x_i)$$

$$= -\frac{1}{2}\left[(p^2+(1-p)^2)\log(p^2+(1-p)^2)+2p(1-p)\log(2p(1-p))\right]$$

$$-\frac{1}{2}\left[(p^2+(1-p)^2)\log(p^2+(1-p)^2)+2p(1-p)\log(2p(1-p))\right]$$

$$= H(2p(1-p)) = H\left[0.5-0.5(1-2p)^2\right] bit/sym$$

信道I，II之间符号的平均互信息量:

$$I(X;Z) = H(Z) - H(Z/X)$$

$$= 1 - H\left[0.5-0.5(1-2p)^2\right] bit/sym$$

3-4 数据处理中信息的变化

同理可得信道I和III的串接信道符号间平均互信息量：

$$I(X;W)=H(W)-H(W/X)$$
$$=1-H\left[0.5-0.5(1-2p)^3\right]bit/sym$$

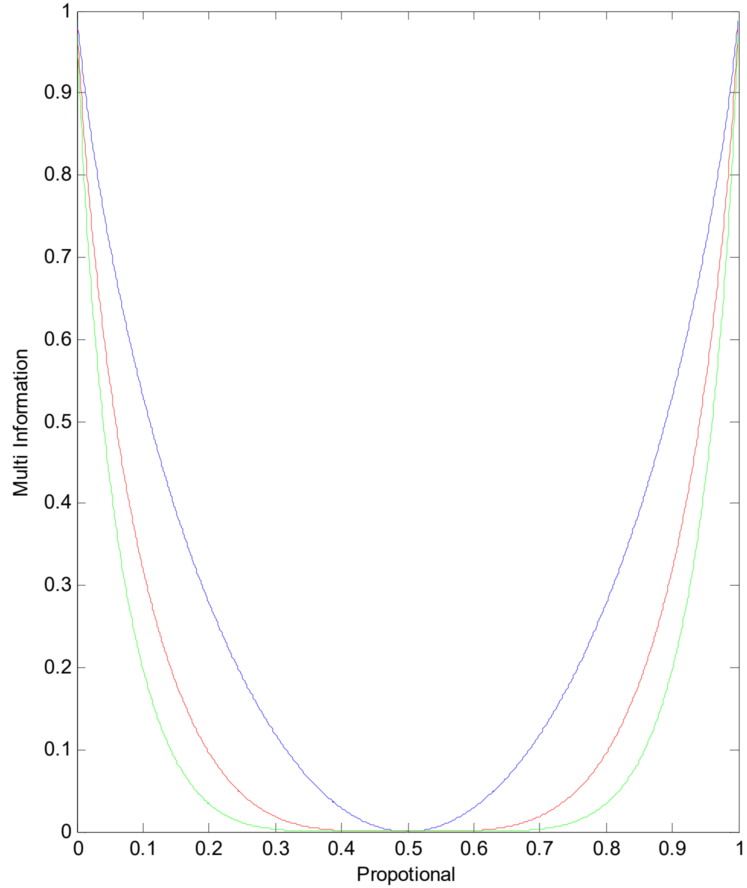
因此，可见信道I，II，III

符号间平局互信息量：

$$I(X;W)\le I(X;Z)\le I(X;Y)$$

同时，可知随着串接信道数目N增加，符号间平均互信息为：

$$I(X;M)=H(M)-H(M/X)$$
$$=1-H\left[0.5-0.5(1-2p)^N\right]bit/sym$$



3-5 信道容量

3.5.1 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

$$
\begin{aligned}
R = I(X;Y) &= H(X) - H(X/Y) \\
&= H(X) + H(Y) - H(XY) \\
&= H(Y) - H(Y/X) \qquad = I(Y;X) \\
&= -\sum p(a_i)\log p(a_i) + \sum\sum p(a_ib_j)\log p(a_i/b_j) \\
&= -\sum p(a_i)\log p(a_i) + \sum\sum p(a_i)p(b_j/a_i)\log \frac{p(a_i)p(b_j/a_i)}{p(b_j)} \\
&= -\sum p(a_i)\log p(a_i) + \sum\sum p(a_i)p(b_j/a_i)\log \frac{p(a_i)p(b_j/a_i)}{\sum_i p(a_ib_j)}
\end{aligned}
$$

3-5 信道容量

引理： 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.

定义： 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
$(Channel\ Capacity)$

$$C=\max_{p(x)}\{I(X;Y)\}\qquad 单位\ bit/Symbol$$

信道容量是信道最大传输能力，是信道自身的特性

3-5 信道容量

定义：

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

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

显然，信道的最大信息传输速率$R_{t\max}$

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

3-5 信道容量

3.5.2 一般信道的信道容量计算方法

根据信道容量的定义为在**信道固定的条件下**，平均互信息量$I(X;Y)$对所有可能的输入分布$p(x)$的极大值。由于**平均互信息量**为**输入分布概率**的**上凸函数**，因此存在极大值。

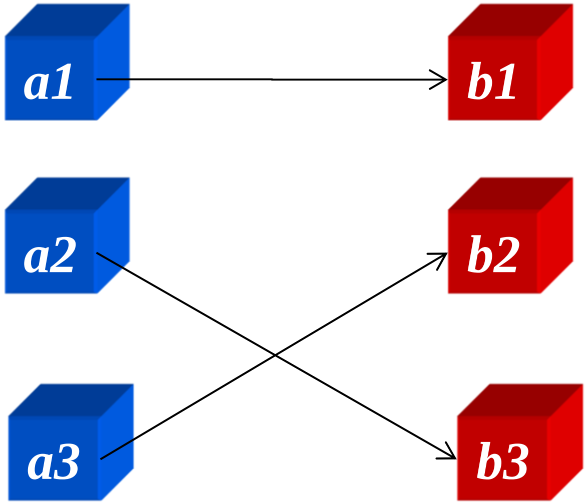
在信道固定的条件下，平均互信息量是r个变量$p(x_i)$的多元函数，且满足约束条件 $\sum_{i=1}^{r} p(x_i)=1$，因此可采用设立辅助函数： $F=I(X;Y)-\lambda\sum_{i} p(x_i)$ 利用拉格朗日乘子法来求解条件极值。

当 $\partial F/\partial p(x_i)=0$ 时求得的平均互信息量$I(X;Y)$即为信道容量。

3-5 信道容量

3.5.3 特殊信道的信道容量计算方法

【1】无噪无损信道（一一对应关系）



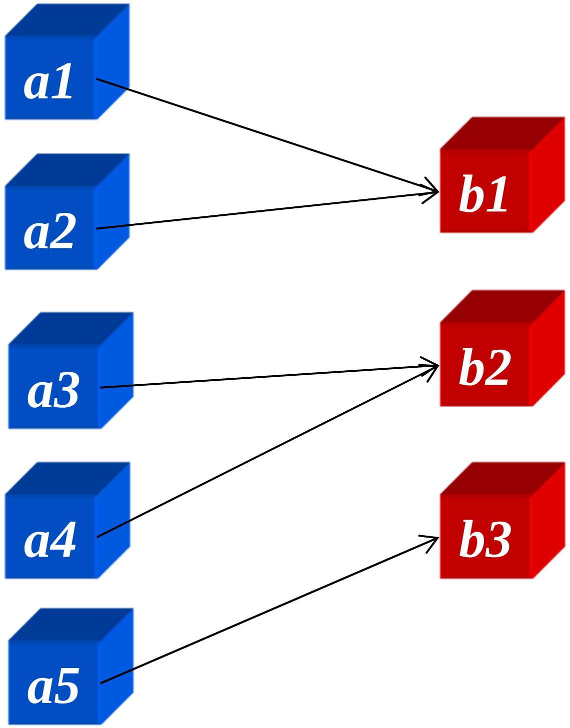
$$P=\begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}$$

疑义度$H(X/Y)=0,$

$$\begin{aligned}C&=\max_{p(x)}\{I(X;Y)\}\\&=\max_{p(x)}\{H(X)-H(X/Y)\}\\&=\max_{p(x)}\{H(X)\}\\&=\log r\quad(bit/Symbol)\end{aligned}$$

3-5 信道容量

【2】无噪有损信道（多个输入对应一个输出）



噪声熵$H(Y/X)=0,$

但疑义度$H(X/Y)\neq 0$

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

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

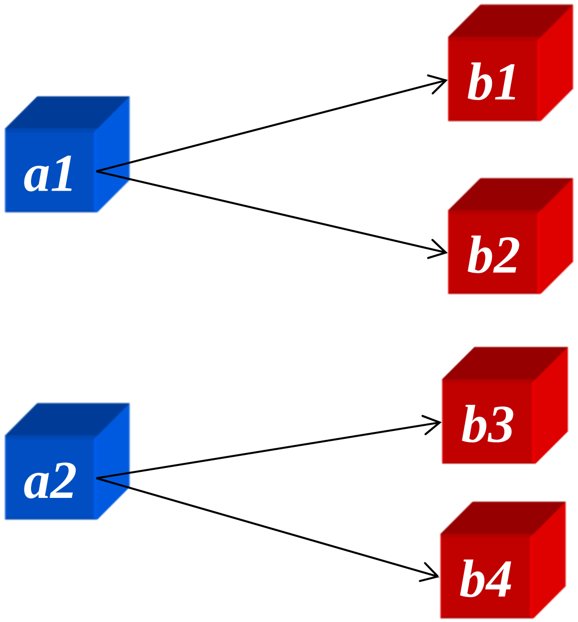
$$=\max_{p(x)}\{H(Y)-H(Y/X)\}$$

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

$$=\log s\quad(bit/Symbol)$$

3-5 信道容量

【3】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）



疑义度$H(X/Y)=0$,

但噪声熵$H(Y/X)\neq 0$

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

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

$$=\max_{p(x)}\{H(X)-H(X/Y)\}$$

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

$$=\log r \quad (bit/Symbol)$$

3-5 信道容量

说明：

- 综上三种情况，若严格区分，凡疑义度（损失熵）等于零的信道称为无损信道；凡噪声熵等于零的信道称为无噪信道，而一一对应的信道则为无噪无损信道。
- 对于无损信道，其信息传输率R就是输入信源X输出每个符号携带的信息量(信源熵H(X))，因此其信道容量为 $C=\log_2 r$ 式中假设输入信源X的符号共有r个，所以等概率分布时信源熵最大。
- 对于无噪信道，其信道容量为 $C=\log_2 s$ 式中假设输出信源Y的符号共有s个，等概率分布时H(Y)最大,而且一定能找到一种输入分布使得输出符号Y达到等概率分布。
- 可见这些信道的信道容量C只取决定于信道的输入符号数r，或输出符号数s，与信源无关。

3-5 信道容量

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

定义：如果转移概率矩阵P的每一行都是由同一符号集$\{p_1,p_2,\cdots,p_s\}$诸元素的不同排列组成，每一列也是由同一集合$\{q_1,q_2,\cdots,q_r\}$诸元素的不同排列组成，将这种DMC信道称为**对称DMC信道**。

$$P_1=\begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6}\\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3}\end{bmatrix}\qquad P_2=\begin{bmatrix}\frac{1}{2} & \frac{1}{3} & \frac{1}{6}\\ \frac{1}{6} & \frac{1}{2} & \frac{1}{3}\\ \frac{1}{3} & \frac{1}{6} & \frac{1}{2}\end{bmatrix}\qquad P_3=\begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6}\\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{3}\end{bmatrix}$$

$$P_4=\begin{bmatrix}0.4 & 0.2 & 0.1 & 0.3\\ 0.2 & 0.1 & 0.4 & 0.3\end{bmatrix}$$

3-5 信道容量

对称DMC信道的相关性质：

（1）对称DMC信道的条件熵$H(Y/X)$与信道输入符号的概率分布无关，且$H(Y/X)=H(Y/a_i)$其中$i=1,2,\cdots r$

$$H(Y/X)=-\sum\sum p(a_i b_j)\log p(b_j/a_i)$$
$$=-\sum\sum p(a_i)p(b_j/a_i)\log p(b_j/a_i)$$
$$=-\sum_i p(a_i)\sum_j p(b_j/a_i)\log p(b_j/a_i)$$

(每行元素相同, $\sum_j p(b_j/a_i)\log p(b_j/a_i)$与$i$无关)

$$=-\sum_j p(b_j/a_i)\log p(b_j/a_i)=H(Y/a_i)$$

$$C=\max_{p(a_i)}\{I(X;Y)\}=\max_{p(a_i)}\{H(Y)-H(Y/X)\}$$
$$=\max_{p(a_i)} H(Y)-H(Y/a_i)$$

3-5 信道容量

[2] 对称DMC信道的输入符号等概率分布时，信道的输出符号也等概率分布；反之，若信道的输出符号等概率分布，信道的输入符号必定也是等概率分布

[3] 对称DMC信道的输入符号等概率分布时，对称DMC信道达到其信道容量，为：

$$
\begin{aligned}
C &= \log s - H(Y/a_i) \\
&= \log s - H(p'_1, p'_2, \cdots, p'_s) \\
&= \log s + \sum_{j=1}^{s} p_{ij} \log p_{ij}
\end{aligned}
$$

3-5 信道容量

【5】强对称DMC信道的容量（BSC信道Binary Symmetric Channel）

定义：如果单符号离散对称信道的输入符号集$X:\{a_1,a_2,\cdots,a_r\}$，输出符号集$Y:\{a_1,a_2,\cdots,a_r\}$，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

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

该信道称为**强对称信道**

3-5 信道容量

【5】强对称DMC信道的容量（BSC信道容量）

$$H(Y/X)=H(X/Y)=H\left[(1-\varepsilon),\frac{\varepsilon}{r-1},\cdots,\frac{\varepsilon}{r-1}\right]$$

$$=H(\varepsilon)+\varepsilon\log(r-1)$$

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

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

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

$$=\log r-H(\varepsilon)-\varepsilon\log(r-1)$$

3-5 信道容量

【5】强对称DMC信道的容量（BSC信道容量）

- 对于强对称离散信道，当输入信源为等概分布时，其输出随机变量同时为等概分布，即当输入达到最大熵值时，输出亦达到最大熵值，同时离散信道达到信道容量C。

- 对于强对称离散信道，当输入信源等概分布时，输入信源X中任一符号a_i与输出信宿Y中任一符号b_j之间的正向概率$p(b_j/a_i)$与反向概率$p(a_i/b_j)$是相等的。

- 对于强对称离散信道，当输入信源等概分布时，信道的噪声熵和疑义度相等，且均为：

$$H(Y/X)=H(X/Y)=H\left[(1-\varepsilon),\frac{\varepsilon}{r-1},\frac{\varepsilon}{r-1},\cdots,\frac{\varepsilon}{r-1}\right]$$

$$=H(\varepsilon)+\varepsilon\log(r-1)$$

3-5 信道容量

【6】准对称信道的信道容量

【定义】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$P_l\,(l=1,2,\ldots,m)$，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，称之为**准对称的信道**。

$$P=\begin{bmatrix}\frac{1}{3} & \frac{1}{6} & \frac{1}{2}\\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6}\end{bmatrix}$$

$$P=\left[\begin{array}{cc|cc}1/4 & 1/2 & 1/8 & 1/8\\ 1/2 & 1/4 & 1/8 & 1/8\end{array}\right]$$

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

3-5 信道容量

准对称信道的信道容量计算：

$$C=\max_{p(x_i)} I(X;Y)=\max_{p(x_i)}\left[H(Y)-H(Y/X)\right]$$

$$=\max_{p(x_i)} H(Y)-H(Y/X)=\max_{p(x_i)} H(Y)-H(p'_1,p'_2,\cdots,p'_s)$$

对于离散准对称信道，由于不确定存在一种输入使输出等概

因此 $C\le\max_{p(x_i)} H(Y)-H(p'_1,p'_2,\cdots,p'_s)$

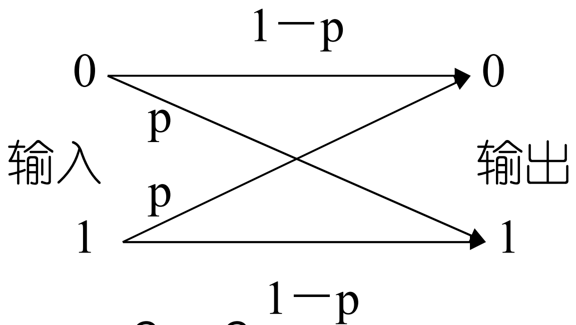
可以证明当输入等概分布时，可以达到信道容量

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

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**

3-5 信道容量

【例题1】二进制对称信道容量计算



$$信道矩阵\mathbf{P}=\begin{bmatrix}1-p & p\\ p & 1-p\end{bmatrix},$$

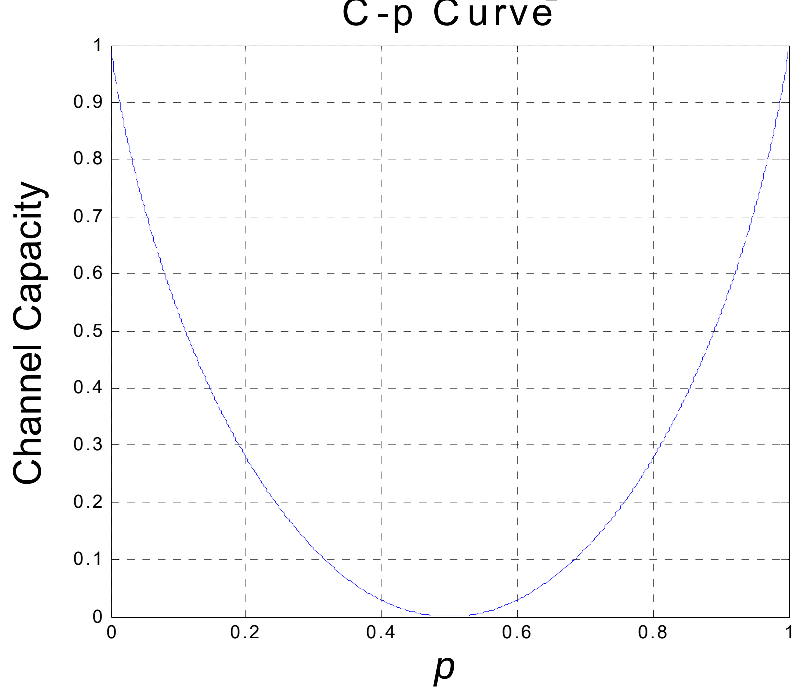
$$C=\max_{p(x)}\{I(X;Y)\}$$

$$=1-H(p)$$

C与p的关系曲线如右

即：$p=0.5$时，

$C=0$



二进制对称信道的匹配信源是二元等概信源。该信道信道容量只是信道传输概率p的函数，不同的二进制对称信道其信道容量也将不同。

3-5 信道容量

【例题2】 设某离散对称信道的信道矩阵为P，计算信道容量

$$P=\begin{bmatrix} 1/2 & 1/3 & 1/6 \\ 1/6 & 1/2 & 1/3 \\ 1/3 & 1/6 & 1/2 \end{bmatrix}$$

解：由信道矩阵可知为对称信道，故

$$C=\max_{p(x_i)} I(X;Y)=\max_{p(x_i)}\left[H(Y)-H(Y/X)\right]$$

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

$$=\sum_{i=1}^{r}p(a_i)\sum_{j=1}^{s}p(b_j/a_i)\log\frac{1}{p(b_j/a_i)}=H(Y/a_i)=H(\frac{1}{2}\quad\frac{1}{3}\quad\frac{1}{6})$$

$$C=\max_{p(x_i)} H(Y)-H(Y/X)=\log s-H(Y/a_i)=\log s-H(\frac{1}{2}\quad\frac{1}{3}\quad\frac{1}{6})$$

$$=\log 3+\frac{1}{2}\log\frac{1}{2}+\frac{1}{3}\log\frac{1}{3}+\frac{1}{6}\log\frac{1}{6}=0.126bit/sym$$

3-5 信道容量

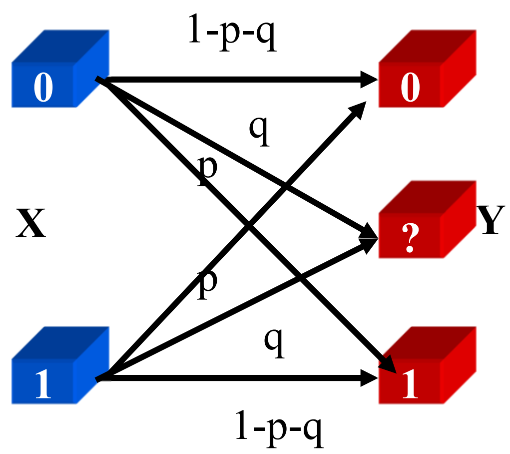
【例题3】 某二元擦除信道矩阵为P

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

计算该信道信道容量C

解：为准对称信道，信道矩阵划分为两个子矩阵组成的形式

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



则 $N_1=1-q \quad M_1=1-q$

$N_2=q \quad M_2=2q$

根据准对称信道容量计算公式

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

$$=\log 2-(1-q)\log(1-q)-q\log(2q)-H(1-p-q,q,p)$$

通信的有效性和可靠性关系