离散信道
及信道容量

信息论与编码原理不挂科 第三讲

离散信道及信道容量

信息论与编码原理不挂科 第三讲

5大模块

7道题目

信息论与编码原理不挂科 第三讲

离散信道及信道容量

模块1 信道的分类及其数学模型

模块2 平均互信息、平均条件互信息

模块3 信道容量及其计算

模块4 离散无记忆扩展信道

模块5 数据处理定理

信道的分类及其数学模型

信道的分类及其数学模型

小节1 信道的分类

小节2 离散信道数学模型

小节3 单符号离散信道

信道的分类「信道是信号的传输媒介，是传送信息的物理通道」

- 按照输入输出信号的幅度和时间特性划分

幅度	时间	信道分类名称
离散	离散	离散信道（数字信道），对讨论编码时有用
连续	离散	连续信道
连续	连续	模拟信道（波形信道），是实际信道，具有重要意义
离散	连续	理论和实用价值较小，不做讨论

信道的分类 「信道是信号的传输媒介，是传送信息的物理通道」

- 按照输入输出间的记忆性划分

无记忆信道 信道在某时刻的输出概率仅依赖于当前时刻的输入，而与先前时刻的输入或输出无关；

有记忆信道 信道在某时刻的输出与其他时刻的输入或输出有关。

- 按照输入输出之间是否是确定关系划分

无干扰信道 输入与输出之间具备确定的相互关系，即传输过程中不会出现差错。

有干扰信道 与无干扰信道对应。

信道的分类 「信道是信号的传输媒介，是传送信息的物理通道」

- 按照统计特性的平稳性划分

平稳信道　统计特性不随时间改变而改变，亦称恒参信道，时不变信道；

非平稳信道　统计特性随时间改变而改变，亦称随参信道，时变信道。

- 按照输入和输出的个数划分

单用户信道　单输入单输出的单向通信；

多用户信道　双向通信或三个或更多用户之间相互通信的情况。

★本讲着重讨论单用户信道，以无记忆，平稳的离散有干扰信道（无反馈）为重点

信道的分类及其数学模型

信道的数学模型

输入信号

输出信号

$$\vec{X} = (X_1X_2\cdots X_N) \longrightarrow P(\vec{Y}\,|\,\vec{X}) \longrightarrow \vec{Y} = (Y_1Y_2\cdots Y_N)$$

$X_i \in \{a_1a_2\cdots a_r\}$

$Y_i \in \{b_1b_2\cdots b_s\}$

干扰

$$[\vec{X}, P(\vec{Y}\,|\,\vec{X}), \vec{Y}]$$

由于干扰存在而导致输出和输入之间会存在一定的概率关系，可用输入序列条件下输出序列的条件概率 $P(\vec{Y}\,|\,\vec{X})$ 表示，转移概率一共有 $r^N \times s^N$ 项。

信道的数学模型

1 无干扰信道（无噪信道）

信道中没有干扰，输出与输入序列之间有确定的一一对应关系 $\vec{Y}=f(\vec{X})$

条件概率满足 $P(\vec{Y}\mid\vec{X})=\begin{cases}1 & \vec{Y}=f(\vec{X})\\ 0 & \vec{Y}\neq f(\vec{X})\end{cases}$

2 有干扰无记忆信道 $P(\vec{Y}\mid\vec{X})=P(Y_1Y_2\cdots Y_N\mid X_1X_2\cdots X_N)=\prod_{i=1}^{N}p(y_i\mid x_i)$

有干扰 指的是信道中有干扰，输出与输入序列之间无确定的对应关系，此时应用**一般的概率分布**表征信道的统计特性；

无记忆 指的是信道在某时刻的输出序列仅依赖于对应时刻的输入序列，而与其他时刻的输入或输出无关，则称信道为无记忆的。

信道的数学模型

3 有记忆有干扰信道

更为一般的情况，信道既有干扰又有记忆，在这一类信道中，某一瞬间的输出符号不但与对应时刻的输入符号有关，还与此前其他时刻的输入符号和输出符号有关。

有记忆有干扰信道的两种处理方法

（1）把记忆较强的N个符号当做一个矢量符号，而各矢量之间认为无记忆，则可按照无记忆信源分析，但会有误差；

（2）把条件概率矩阵视为马尔可夫链，则可以利用平稳信源的性质进行分析。

信道的分类及其数学模型

小节1 信道的分类

小节2 离散信道数学模型

小节3 单符号离散信道

单符号离散信道「离散无记忆平稳信道」

离散无记忆信道一般都看做离散无记忆恒参（平稳）信道；

故对离散无记忆信道的研究**只需研究单个符号的传输**即可，所以也叫**单符号离散信道。**

输入信号 输出信号

$$X_i \in \{a_1 a_2 \cdots a_r\} \longrightarrow P(Y|X) \longrightarrow Y_i \in \{b_1 b_2 \cdots b_s\}$$

干扰

$$[X, P(Y|X), Y]$$

其中，输入和输出的传递概率为 $P(y|x) = P(y = b_j | x = a_i)$ $(i = 1,2,\cdots,r \quad j = 1,2,\cdots,s)$

信道传递概率矩阵

其中，输入和输出的传递概率为 $P(y|x) = P(y = b_j|x = a_i)$ $(i = 1,2,\cdots,r \quad j = 1,2,\cdots,s)$

以输入符号为条件，各传递概率组合成一个矩阵，即可得到信道传递概率矩阵：

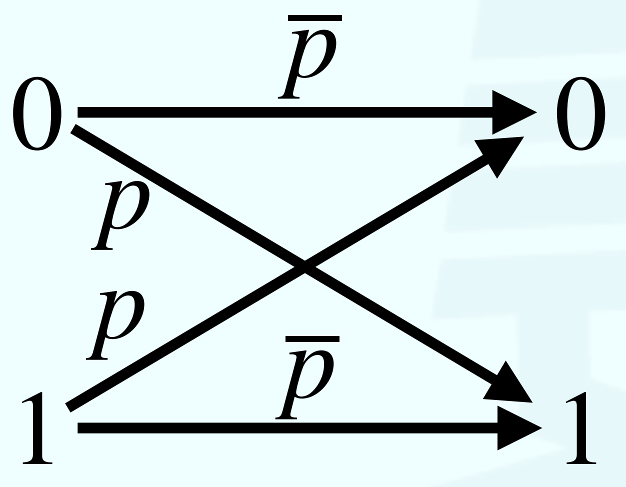
$$\begin{bmatrix} p(b_1|a_1) & p(b_2|a_1) & \cdots & p(b_s|a_1) \\ p(b_1|a_2) & p(b_2|a_2) & \cdots & p(b_s|a_2) \\ \vdots & \vdots & \ddots & \vdots \\ p(b_1|a_r) & p(b_2|a_r) & \cdots & p(b_s|a_r) \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rs} \end{bmatrix}$$

在有干扰的情况下 $\sum_{j=1}^{s} p(b_j|a_i) = 1$

单符号离散信道举例

例1：二元对称信道（简称BSC）

输入符号集A = {0,1}，输出符号集B ={0,1}



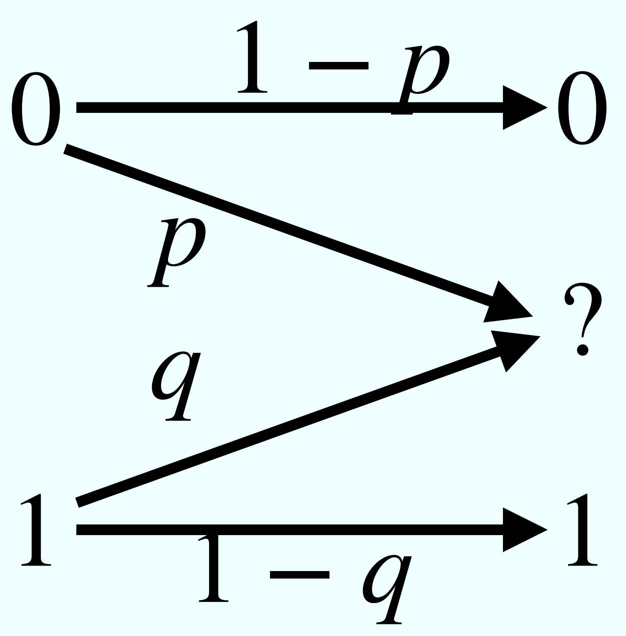
$$p(0\,|\,0) = \bar{p} \qquad p(1\,|\,0) = p$$

$$p(0\,|\,1) = p \qquad p(1\,|\,1) = \bar{p}$$

$$P = \begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix}$$

例2：二元删除信道（简称BEC）

输入符号集A = {0,1}，输出符号集B ={0,?,1}



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

常用概率及其关系

1 先验概率 $p(a_i)$

2 前向概率（传递概率） $P(b_j|a_i) = P(y = b_j|x = a_i)$

3 后向概率（后验概率） $P(a_i|b_j) = P(x = a_i|y = b_j)$

4 联合概率（后验概率） $p(a_ib_j) = p(X = a_i, Y = b_j) = p(a_i)p(b_j|a_i) = p(b_j)p(a_i|b_j)$

5 先验概率、前向概率与后验概率的关系——贝叶斯公式 $P(a_i|b_j) = \dfrac{P(a_ib_j)}{p(b_j)} = \dfrac{p(b_j|a_i)p(a_i)}{\sum_i^r (b_j|a_i)p(a_i)}$

6 输出符号概率——全概率公式 $p(b_j) = \sum_{i=1}^{r} p(a_ib_j) = \sum_{i}^{r} p(a_i)p(b_j|a_i) \longleftrightarrow P_Y = P_X P_{Y|X}$

常用概率及其关系

根据上述概率公式，我们可以得出结论：

有关信道的所有概率都可以由先验概率 $p(a_i)$ 和前向概率 $P(b_j|a_i)$ 表示。

输入符号概率，输出符号概率，以及转移概率的关系为：

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

平均互信息、平均条件互信息

小节1 信道疑义度

小节2 互信息

小节3 条件互信息

平均互信息、平均条件互信息

小节1 信道疑义度

小节2 互信息

小节3 条件互信息

信道疑义度的定义

表示接收端收到信道输出的一个符号之后对信道输入的的符号仍然存在的平均不确定性。

信道疑义度的定义为 $H(X|Y)=\sum_{j} p(y_j)H(X|y_j)$

- 理想传输时，输出与输入一一对应，因此已知输出对于输入就没有不确定度了，有 $H(X|Y)=0$
- 一般情况下，有 $H(X|Y)<H(X)$，即收到输出后对于输入的不确定度一般总会减少。
- 当 $H(X|Y)=H(X)$ 时，则表示收到输出变量后对输入变量的不确定性一点也没有减少

平均互信息、平均条件互信息

互信息量

互信息量的定义

互信息量表示事件y_j所给出的关于x_i的信息量，其定义式为

$$I(x_i;y_j)=I(x_i)-I(x_i|y_j)=\log\frac{p(x_i|y_j)}{p(x_i)}$$

$I(x_i)$表示先验不确定度：收到 y_j 之前关于输入x_i 存在的不确定度；

$I(x_i|y_j)$表示后验不确定度：收到 y_j 之后关于输入x_i 仍然存在的不确定度。

互信息量的含义

表示事件y_j 出现前后关于输入x_i 存在的不确定度的减少量；

表示事件 y_j 出现以后信宿获得的关于输入x_i 的信息量。

互信息量的计算

$I(x_i;y_j)=I(x_i)-I(x_i|y_j)$　　$I(x_i;y_j)=I(y_j)-I(y_j|x_i)$　　$I(x_i;y_j)=I(x_i)+I(y_j)-I(x_iy_j)$

互信息量的性质

1 互信息具有**互易性**

$I(x_i; y_j) = I(y_j; x_i)$ 观察角度不同，但互信息量具有对称性

2 互信息**可正可负可为0**

$$I(x_i; y_j) = I(x_i) - I(x_i | y_j) = \log \frac{p(x_i | y_j)}{p(x_i)} \begin{cases} >0, & \text{正相关} \\ =0, & \text{不相关（独立）} \\ <0, & \text{负相关} \end{cases}$$

3 互信息**不可能大于其中任意一个事件的自信息**

$I(x_i; y_j) \leq I(x_i)$ $I(x_i; y_j) \leq I(y_j)$

某事件的自信息量是发生其他事件能提供的关于该事件的最大信息量

例题3-1 已设事件 e 表示“降雨”，事件 f 表示“空中有乌云”，且$P(e)=0.125$，$P(e\,|\,f)=0.8$，试求：

（1）“降雨”的自信息；（2）“空中有乌云”条件下“降雨”的自信息；

（3）“无雨”的自信息；（4）“空中有乌云”条件下“无雨”的自信息；

（5）“降雨”与“空中有乌云”的互信息；（6）“无雨”与“空中有乌云”的互信息。

解析3-1

(1) “降雨”的自信息 $I(e)=-\log 0.125=3bit$

(2) “空中有乌云”条件下“降雨”的自信息 $I(e\,|\,f)=-\log 0.8=0.322bit$

(3) “无雨”的自信息 $I(\overline{e})=-\log 0.875=0.193bit$

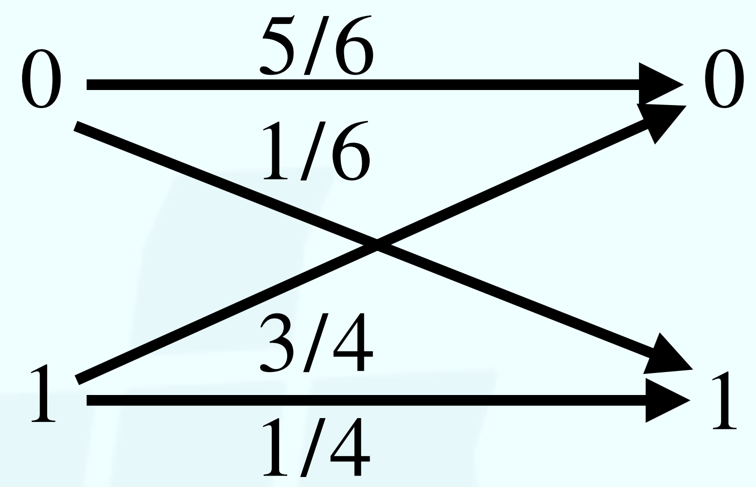
(4) “空中有乌云”条件下“无雨”的自信息 $I(\overline{e}\,|\,f)=-\log 0.2=2.322bit$

(5) “降雨”与“空中有乌云”的互信息 $I(e;f)=I(e)-I(e\,|\,f)=3-0.322=2.678bit$

(6) “无雨”与“空中有乌云”的互信息 $I(\overline{e};f)=I(\overline{e})-I(\overline{e}\,|\,f)=0.193-2.322=-2.129bit$

例题3-2 已一离散无记忆信源，其概率空间为 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0.6 & 0.4 \end{bmatrix}$

信源的发出符号经过一个干扰信道，信道输出端的接收符号集为Y={0,1}，信道传递概率如图所示，试求：



（1）信源符号0和1分别含有的自信息量；

（2）收到y_j后，获得的关于x_i的信息量。

解析3-2 (1) $I(0) = -\log 0.6 = 0.737bit$ $\qquad I(1) = -\log 0.4 = 1.314bit$

(2) $P_{Y|X} = \begin{bmatrix} \frac{5}{6} & \frac{1}{6} \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix}$ $\qquad P_Y = P_X P_{Y|X} = \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \end{bmatrix} \begin{bmatrix} \frac{5}{6} & \frac{1}{6} \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix} = \begin{bmatrix} \frac{4}{5} & \frac{1}{5} \end{bmatrix}$

$$I(Y=0;X=0) = I(Y=0) - I(Y=0\,|\,X=0) = -\log\frac{4}{5} - (-\log\frac{5}{6}) = 0.059bit$$

$$I(Y=1;X=0) = I(Y=1) - I(Y=1\,|\,X=0) = -\log\frac{1}{5} - (-\log\frac{1}{6}) = -0.263bit$$

$$I(Y=0;X=1) = I(Y=0) - I(Y=0\,|\,X=1) = -\log\frac{4}{5} - (-\log\frac{3}{4}) = -0.093bit$$

$$I(Y=1;X=1) = I(Y=1) - I(Y=1\,|\,X=1) = -\log\frac{1}{5} - (-\log\frac{1}{4}) = 0.322bit$$

平均互信息的定义

随机变量X和Y的平均互信息定义为事件的互信息在联合概率空间$P(XY)$中的统计平均，是两个概率空间之间的平均互信息，其定义式为：

$$
\begin{aligned}
I(X;Y) = E[I(x_i;y_j)] &= \sum_{i=1}^{m} p(x_i y_j) I(x_i;y_j) = \sum_{i=1}^{m}\sum_{j=1}^{n} p(x_i y_j)\log\frac{p(x_i|y_j)}{p(x_i)} \\
&= E[I(x_i) - I(x_i|y_j)] = E[I(x_i)] - E[I(x_i|y_j)] \\
&= H(X) - H(X|Y) \\
&= H(X) + H(Y) - H(XY) \\
&= H(Y) - H(Y|X)
\end{aligned}
$$

平均互信息的物理意义「从输出端来看」

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

$H(X|Y)$ **信道疑义度/损失熵/后验熵**：表示收到输出Y之后，对随机变量X仍然存在的不确定度，代表了信道中损失的信息；

$H(X)$ **X的先验不确定度|先验熵|无条件熵**；

$I(X;Y)$ 收到Y前后对于X的平均不确定度的减少量，即从Y获得的关于X的平均信息量。

平均互信息的物理意义「从输入端来看」

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

$H(Y|X)$ **噪声熵：表示发出X之后，对随机变量Y存在的平均不确定度**

可以理解为噪声引入的信息量（$Y=N+X$），若信道中不存在噪声那么输入输出存在确定的对应关系，发出X后必然能确定对应的Y，如果不能确定，那么这个不确定性就是由噪声引起的。

$I(Y;X)$ 收到Y获得的信息量减去由噪声引起的额外的信息量

平均互信息的物理意义「从整个通信系统来看」

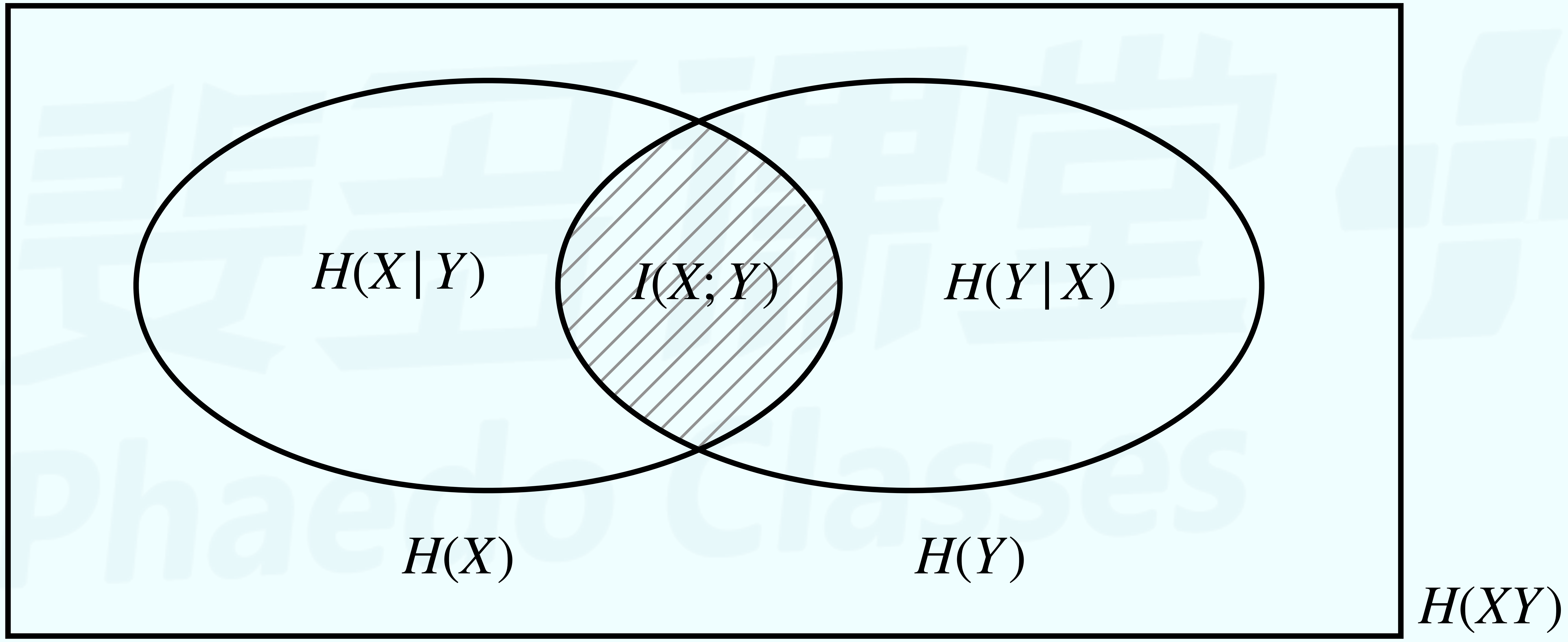
$$I(X;Y) = H(X) + H(Y) - H(XY)$$

$H(X) + H(Y)$ 假设把X和Y视为通信之前两个独立的随机变量，则此时系统的**先验不确定度则为二者平均不确定度之和$H(X)+H(Y)$;**

$H(XY)$ 通信之后，信道的传递统计特性将X与Y联系起来，具有统计关联性，此时的**后验不确定度则由联合熵$H(XY)$表示;**

$I(X;Y)$ 通信前、后整个系统的不确定度的减少量。

利用维拉图表示平均互信息



X 和 Y 的联合熵 ⟷ X 和 Y 的并集

X 和 Y 的平均互信息 ⟷ X 和 Y 的交集

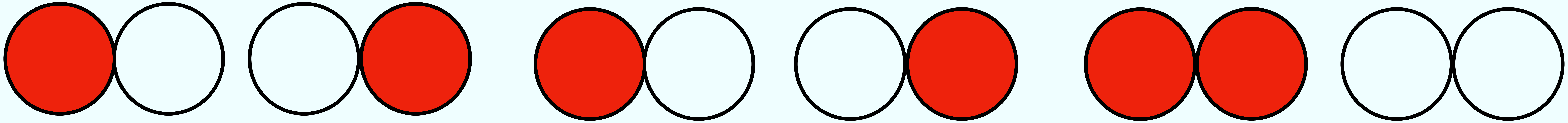
X 和 Y 的条件熵 ⟷ X 和 Y 集合间的减法

利用维拉图表示平均互信息

X 和 Y 有依赖关系时：

$H(X)$ $H(Y)$ $H(X|Y)$ $H(Y|X)$ $H(XY)$ $I(X;Y)$

X 和 Y 相互独立时：



$H(X)$ $H(Y)$ $H(X|Y)=H(X)$ $H(Y|X)=H(Y)$ $H(XY)=H(X)+H(Y)$ $I(X;Y)=0$

平均互信息的性质

1 非负性 $I(X;Y) \geq 0$

互信息可负，平均互信息非负，因为是从统计平均的角度研究的；
一般输出 Y 多少都会得到一些关于 X 的信息，当 X 与 Y 独立时，等号才成立。

2 对称性 $I(X;Y) = I(Y;X)$

从X中得到的Y的信息和从Y中得到的X的信息对等。

3 极值性 $I(X;Y) \leq H(X)$ $I(X;Y) \leq H(Y)$ $I(X;Y) \leq \min\{H(X), H(Y)\}$

从一个事件中提取另一个事件的信息量，**最多可以提取出该事件的自信息量。**

平均互信息的性质

4 凸性定理

平均互信息是输入信源概率分布和信道传递概率分布的凸函数。

对于固定信道，即 $p(y_j|x_i)$ 固定，平均互信息是输入信源概率的上凸函数。

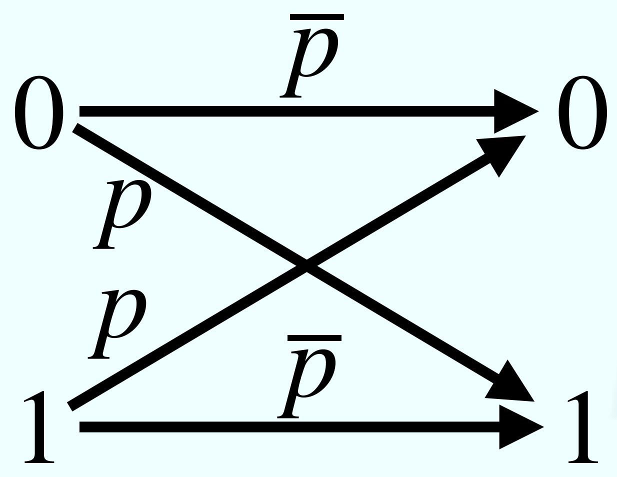
因此，存在一个最佳输入分布，使平均互信息最大，最大值由信道特性决定。

对于固定信源，即 $p(x_i)$ 固定，平均互信息是信道传递概率的下凸函数。

存在一个最差信道，此信道的干扰最大，导致输出端获得的信息量最小。

关于凸性定理的举例

二元对称信道的输入概率分布，即信源概率空间为 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \omega & 1-\omega \end{bmatrix}$



$$P = \begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix}$$

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - \sum_{i=1}^{2} p(x_i)H(Y|X_i)$$

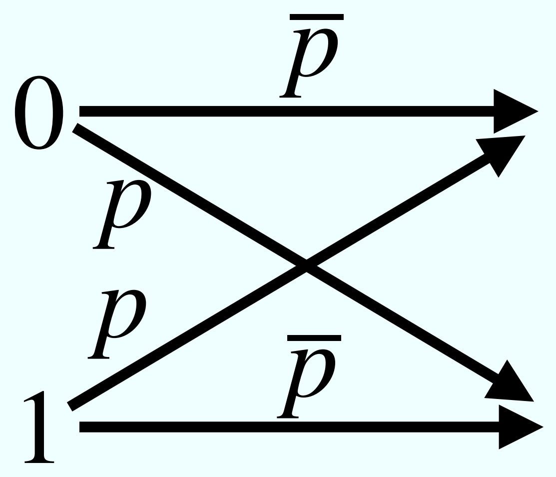
$$= H(Y) - [\omega H(\bar{p},p) + (1-\omega)H(\bar{p},p)] = H(Y) - H(\bar{p},p)$$

$$= H(Y) - H(p)$$

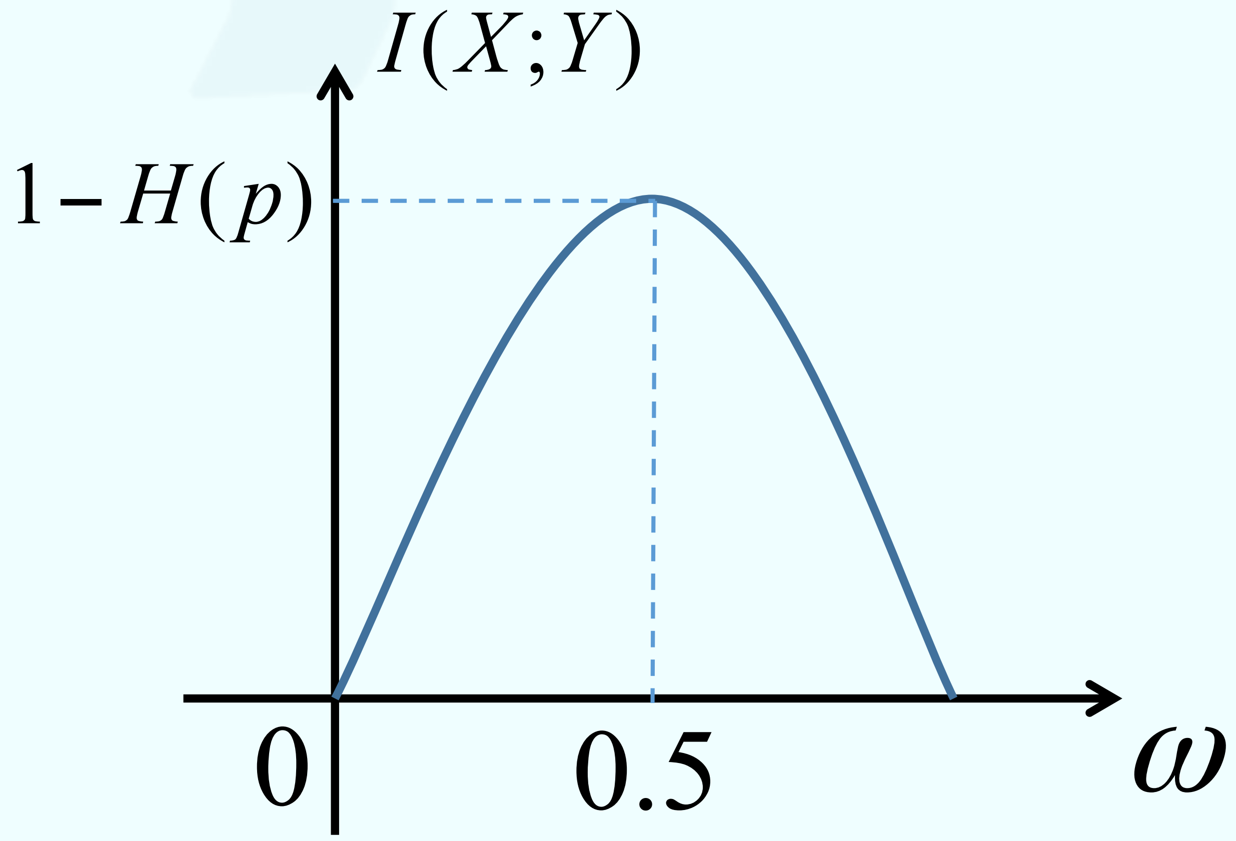
$$P_Y = P_X P_{Y|X} = [\omega \quad \bar{\omega}] \begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix} = [\omega\bar{p} + \bar{\omega}p \quad \omega p + \bar{\omega}\bar{p}] \quad \Rightarrow \quad H(Y) = H(\omega\bar{p} + \bar{\omega}p, \omega p + \bar{\omega}\bar{p})$$

关于凸性定理的举例

二元对称信道的输入概率分布，即信源概率空间为 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \omega & 1-\omega \end{bmatrix}$



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



$I(X;Y) = H(Y) - H(p) \quad H(Y) = H(\omega\overline{p} + \overline{\omega}p, \omega p + \overline{\omega}\overline{p})$

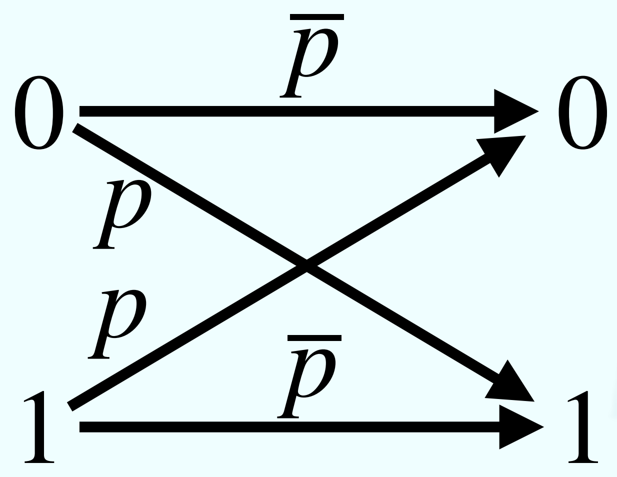
当 $\omega = 1 - \omega = \dfrac{1}{2}$ 时，$H(Y) = H(\omega\overline{p} + \overline{\omega}p) = H(\dfrac{1}{2}) = 1$

因此 $I(X;Y)_{\max} = 1 - H(p)$

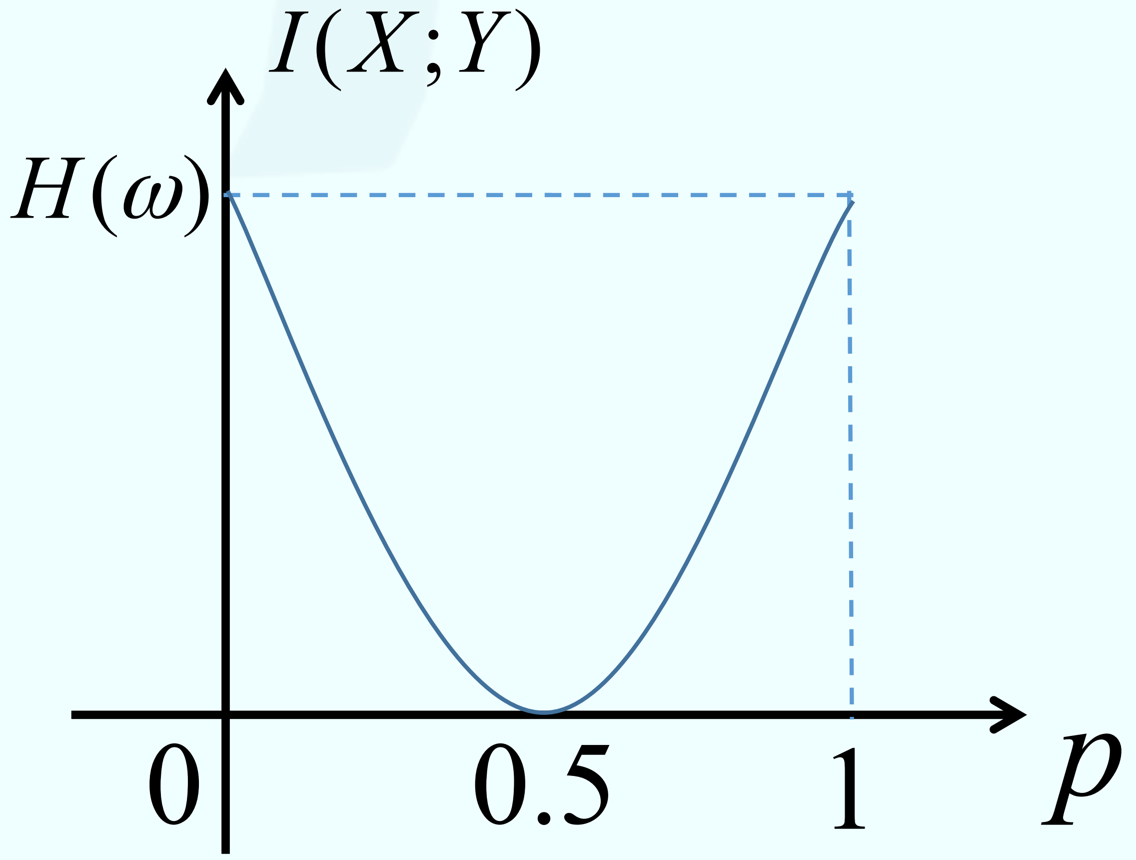
此时互信息取到最大值，其值由信道的转移概率，也就是参数 p 决定

关于凸性定理的举例

二元对称信道的输入概率分布，即信源概率空间为 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \omega & 1-\omega \end{bmatrix}$



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



$$I(X;Y) = H(\omega\overline{p} + \overline{\omega}p, \omega p + \overline{\omega}\overline{p}) - H(p,\overline{p})$$

p=0时，$I(X;Y) = H(\omega,\overline{\omega}) - H(0,1) = H(\omega,\overline{\omega})$

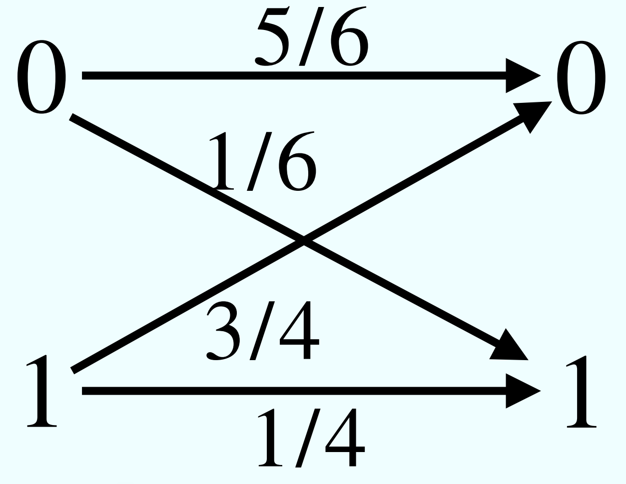
p=1时，$I(X;Y) = H(\overline{\omega},\omega) - H(1,0) = H(\overline{\omega},\omega)$

$p=\frac{1}{2}$时，$I(X;Y) = H(\frac{1}{2},\frac{1}{2}) - H(\frac{1}{2},\frac{1}{2}) = 0$

此时输入的信息全部损失在信道中，为最差的信道，与信道传输特性有关。

例题3-3 离散无记忆信源，其概率空间为 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0.6 & 0.4 \end{bmatrix}$

信源的发出符号经过一个干扰信道，信道输出端的接收符号集为Y={0,1}，信道传递概率如图所示，试求平均互信息$I(X;Y)$



解析3-3 根据信源空间与信道传递概率，易得输入矩阵与信道矩阵分别为

$$P_X = [0.6 \quad 0.4] \qquad P_{Y|X} = \begin{bmatrix} \frac{5}{6} & \frac{1}{6} \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix}$$

因此输出矩阵 $P_Y = P_X P_{Y|X} = \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \end{bmatrix} \begin{bmatrix} \frac{5}{6} & \frac{1}{6} \\ \frac{3}{4} & \frac{1}{4} \end{bmatrix} = \begin{bmatrix} \frac{4}{5} & \frac{1}{5} \end{bmatrix}$

$$I(X;Y) = H(Y) - H(Y|X) = H(\frac{4}{5},\frac{1}{5}) - \sum_{i=1}^{2} p(x_i)H(Y|x_i)$$

$$= H(\frac{4}{5},\frac{1}{5}) - \frac{3}{5}H(\frac{5}{6},\frac{1}{6}) - \frac{2}{5}H(\frac{3}{4},\frac{1}{4})$$

$$= 0.722 - 0.390 - 0.325$$

$$= 0.007 bit/symbol$$

平均互信息、平均条件互信息

联合互信息量

$y_j z_k$联合发生后提供的有关x_i的信息量

$$I(x_i; y_j z_k) = \log \frac{p(x_i \mid z_k y_j)}{p(x_i)} = I(x_i) - I(x_i \mid y_j z_k)$$

平均联合互信息量

定义：在联合集XYZ中，将互信息 $I(x;yz)$ 求统计平均

$$\sum_{i,j,k} p(x_i y_j z_k)\log\frac{1}{p(x_i)}$$

$$= \sum_{i} p(x_i)\log\frac{1}{p(x_i)} \boxed{\sum_{j,k} p(y_j z_k)} = 1$$

$$= \sum_{i} p(x_i)\log\frac{1}{p(x_i)}$$

$$I(X;YZ) = \sum_{i,j,k} p(x_i y_j z_k)\log\frac{p(x_i|y_j z_k)}{p(x_i)}$$ 定义式

$$= \boxed{\sum_{i,j,k} p(x_i y_j z_k)\log\frac{1}{p(x_i)}} - \sum_{i,j,k} p(x_i y_j z_k)\log\frac{1}{p(x_i|y_j z_k)}$$

$$= \sum_{i} p(x_i)\log\frac{1}{p(x_i)} - \sum_{i,j,k} p(x_i y_j z_k)\log\frac{1}{p(x_i|y_j z_k)}$$

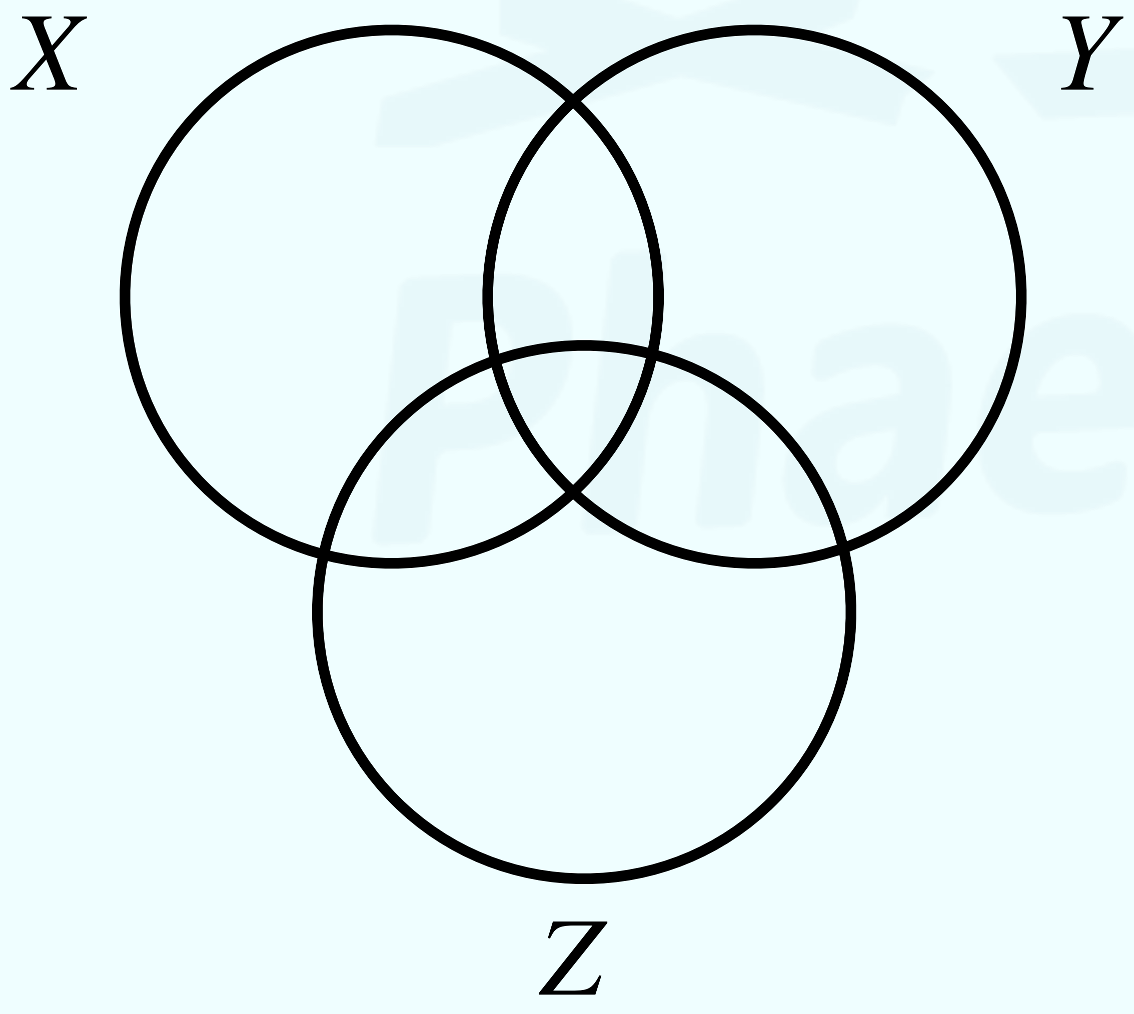
$$= E[I(x_i) - I(x_i|y_j z_k)]$$

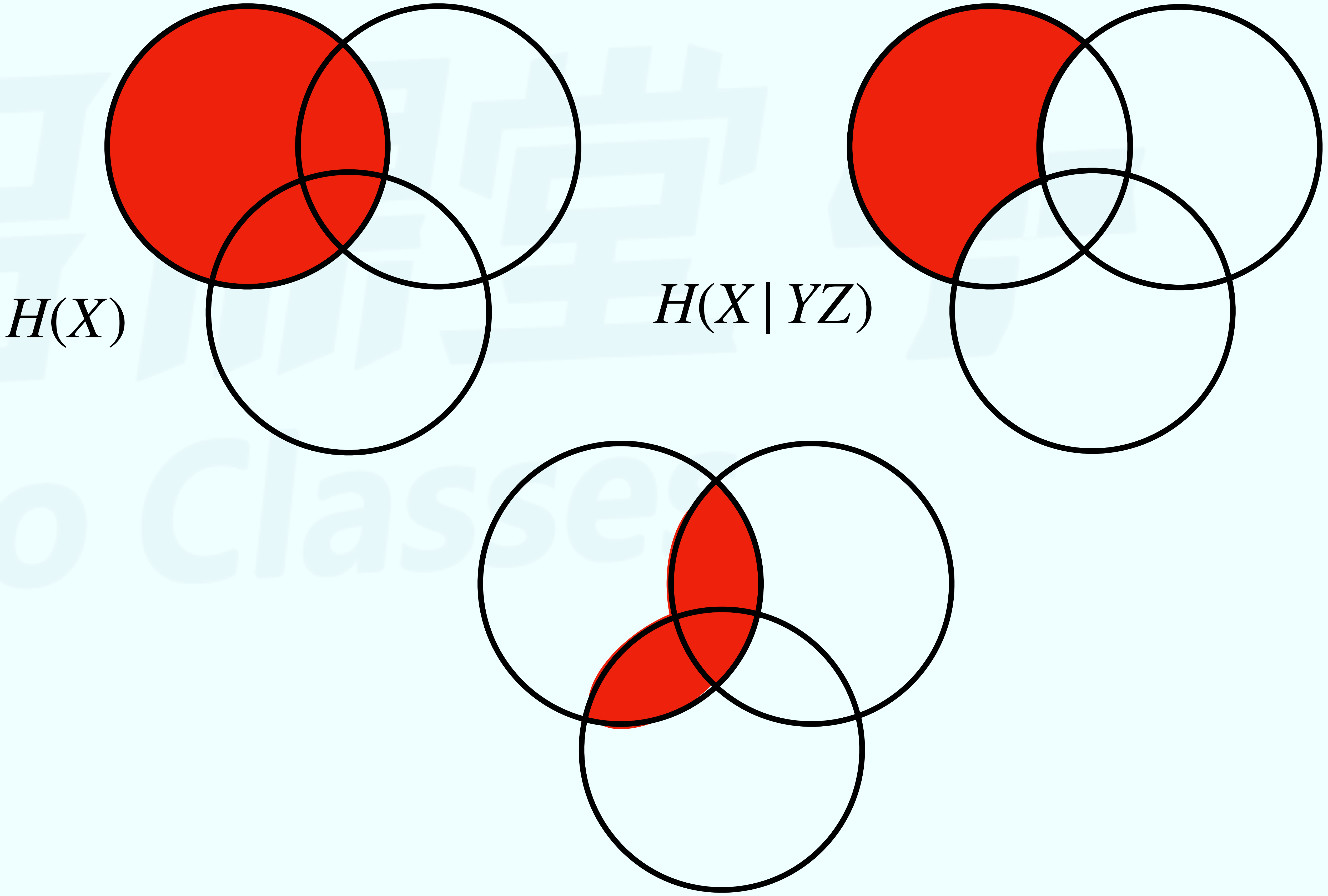
$$= H(X) - H(X|YZ)$$ 计算式

利用维拉图表示平均联合互信息

对于三变量X，Y和Z：

$$I(X;YZ) = H(X) - H(X|YZ)$$





$$I(X;YZ) = H(X) - H(X|YZ)$$

条件互信息量

定义： 在联合集XYZ中，给定事件y_j的条件下，x_i与z_k的互信息

$$I(x_i; z_k | y_j) = \log \frac{p(x_i | z_k y_j)}{p(x_i | y_j)} = I(x_i | y_j) - I(x_i | z_k y_j)$$

与互信息的区别：

加了一个特定的条件。

联合互信息量可以用条件互信息量求解

$$I(x_i; y_j z_k) = I(x_i) - I(x_i | y_j z_k) = \log \frac{p(x_i | y_j z_k)}{p(x_i)} \cdot \frac{p(x_i | y_j)}{p(x_i | y_j)}$$

$$= \log \frac{p(x_i | y_j z_k)}{p(x_i)} \cdot \frac{p(x_i | y_j z_k)}{p(x_i | y_j)}$$

$$= I(x_i; y_j) + I(x_i; z_k | y_j)$$

$y_j z_k$联合发生后提供的有关x_i的信息量，等于y_j给出有关x_i的信息量，加上y_j已知条件下z_k给出的x_i的信息量。

平均条件互信息量

定义： 在联合集XYZ中，将条件互信息 $I(x;z|y)$ 求统计平均

$$I(X;Z|Y)=\sum_{i,j,k}p(x_i,y_j,z_k)\log\frac{p(x_i|z_ky_j)}{p(x_i|y_j)}$$ **定义式**

$$\sum_{i,j,k}p(x_iy_jz_k)\log\frac{1}{p(x_i|y_j)}$$

$$=\sum_{i,j}p(x_iy_j)\log\frac{1}{p(x_i|y_j)}\sum_k p(z_k)\quad =1$$

$$=\sum_{i,j}p(x_iy_j)\log\frac{1}{p(x_i|y_j)}$$

$$=\sum_{i,j,k}p(x_iy_jz_k)\log\frac{1}{p(x_i|y_j)}-\sum_{i,j,k}p(x_iy_jz_k)\log\frac{1}{p(x_i|z_ky_j)}$$

$$=\sum_{i,j}p(x_iy_j)\log\frac{1}{p(x_i|y_j)}-\sum_{i,j,k}p(x_iy_jz_k)\log\frac{1}{p(x_i|z_ky_j)}$$

$$=E[I(x_i|y_j)]-E[I(x_i|z_ky_j)]$$

$$=H(X|Y)-H(X|YZ)$$ **计算式**

平均条件互信息量

利用平均条件互信息求平均联合互信息量

$$I(X;YZ)=\sum_{i,j,k}p(x_iy_jz_k)\log\frac{p(x_i|y_jz_k)}{p(x_i)}$$

$$=\sum_{i,j,k}p(x_i,y_j,z_k)\log\frac{p(x_i|y_jz_k)}{p(x_i|y_j)}\cdot\frac{p(x_i|y_j)}{p(x_i)}$$

$$=\sum_{i,j,k}p(x_i,y_j,z_k)\log\frac{p(x_i|y_jz_k)}{p(x_i|y_j)}+\sum_{i,j}p(x_i,y_j)\log\frac{p(x_i|y_j)}{p(x_i)}$$

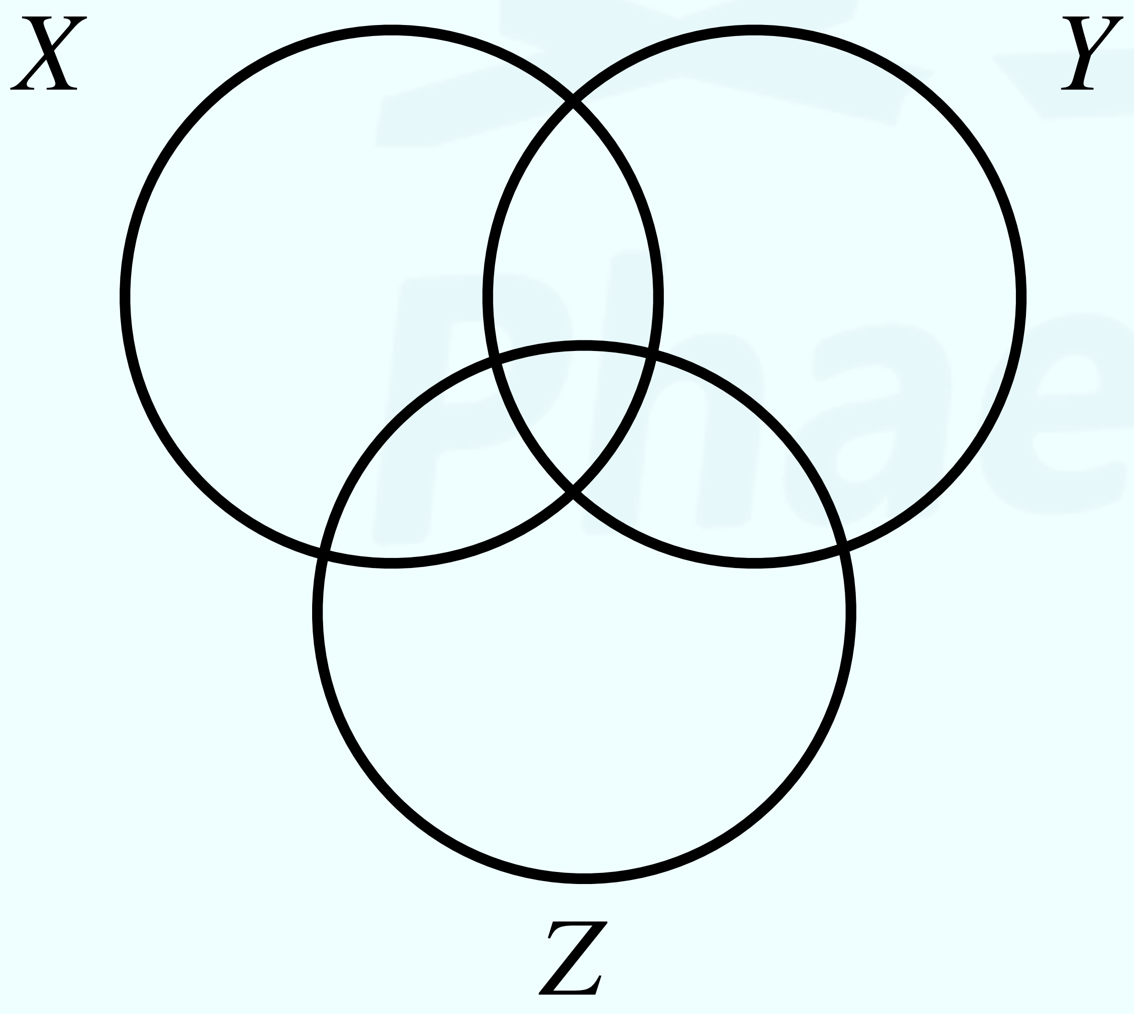
$$=E[I(x_i;z_k|y_j)]+E[I(x_i;y_j)]$$

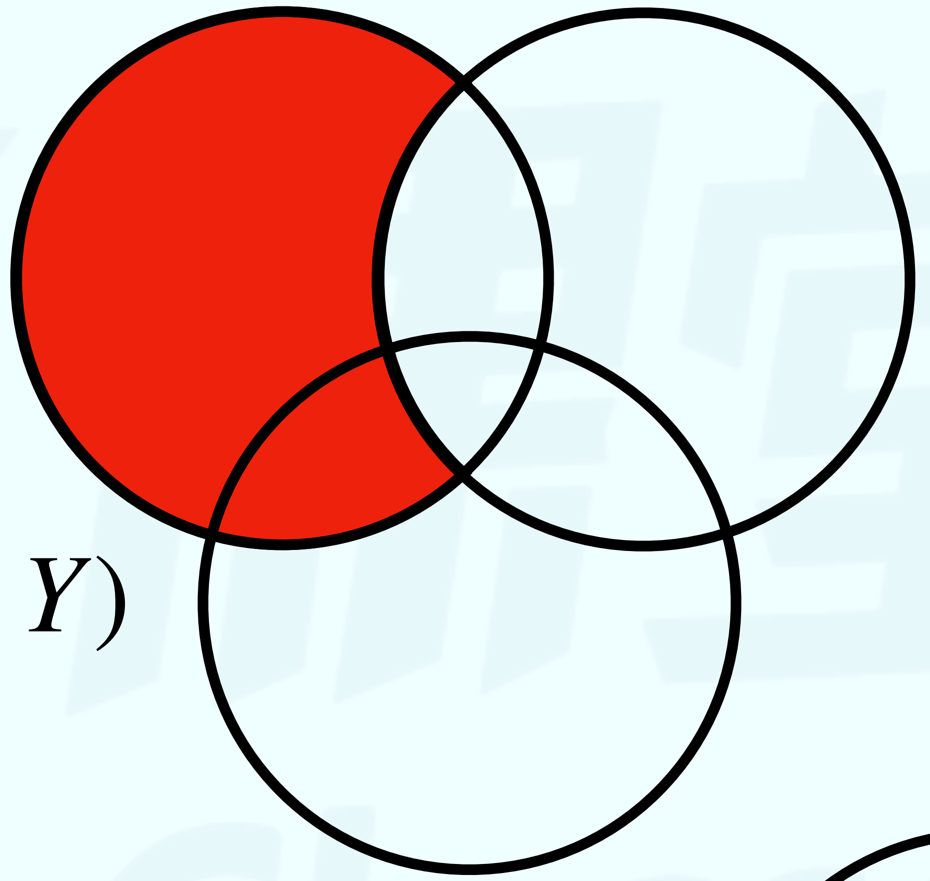
$$=I(X;Y)+I(X;Z|Y)$$ 故 $I(X;Z|Y)=I(X;YZ)-I(X;Y)$

利用维拉图表示平均条件互信息

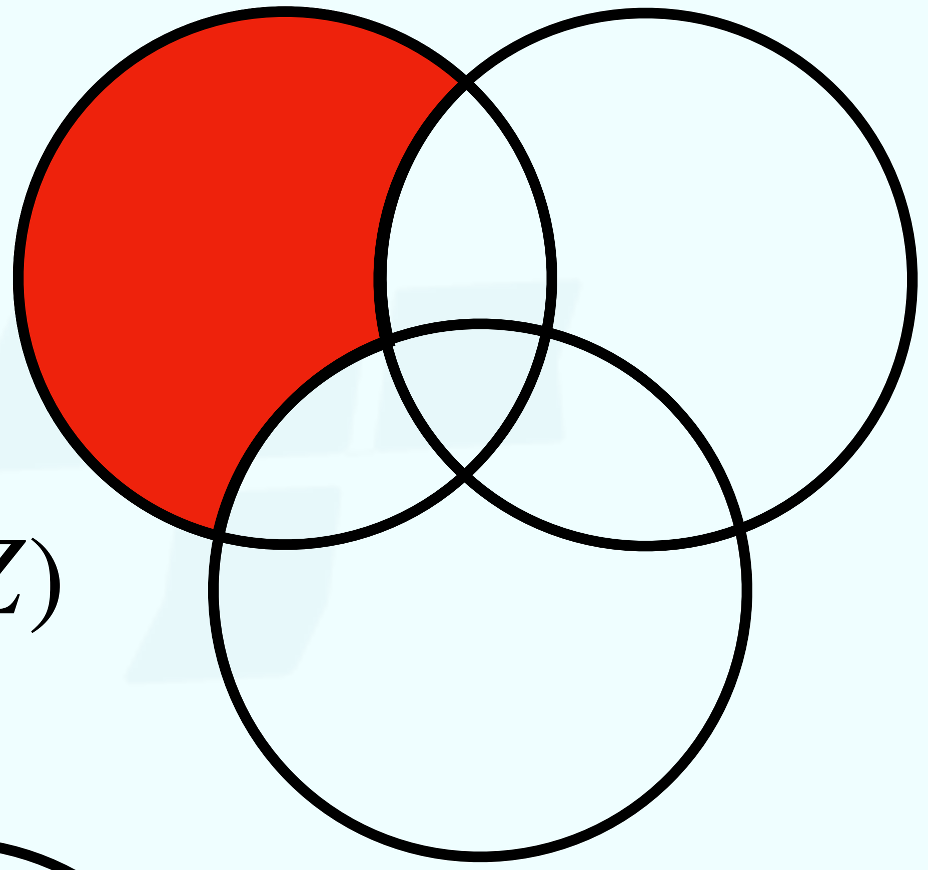
对于三变量X，Y和Z：

$$I(X;Z|Y)=H(X|Y)-H(X|YZ)$$

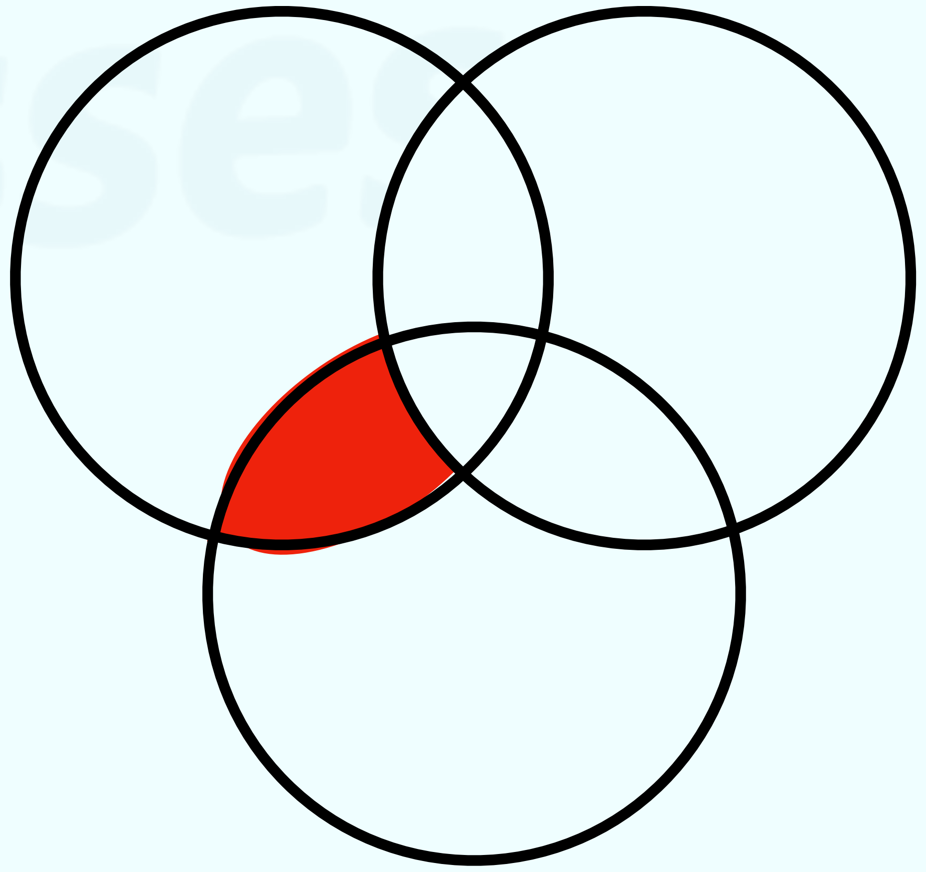




$H(X|Y)$



$H(X|YZ)$

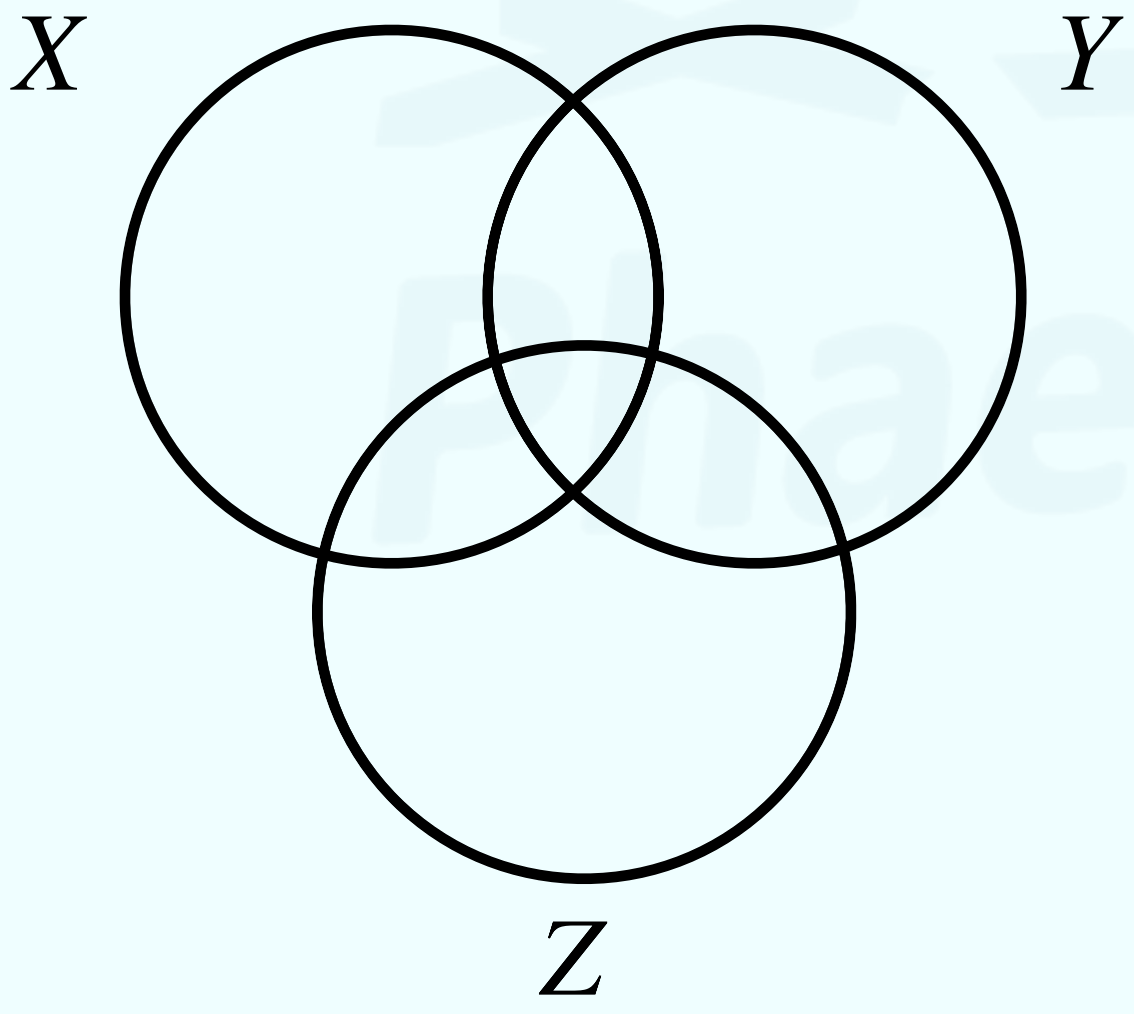


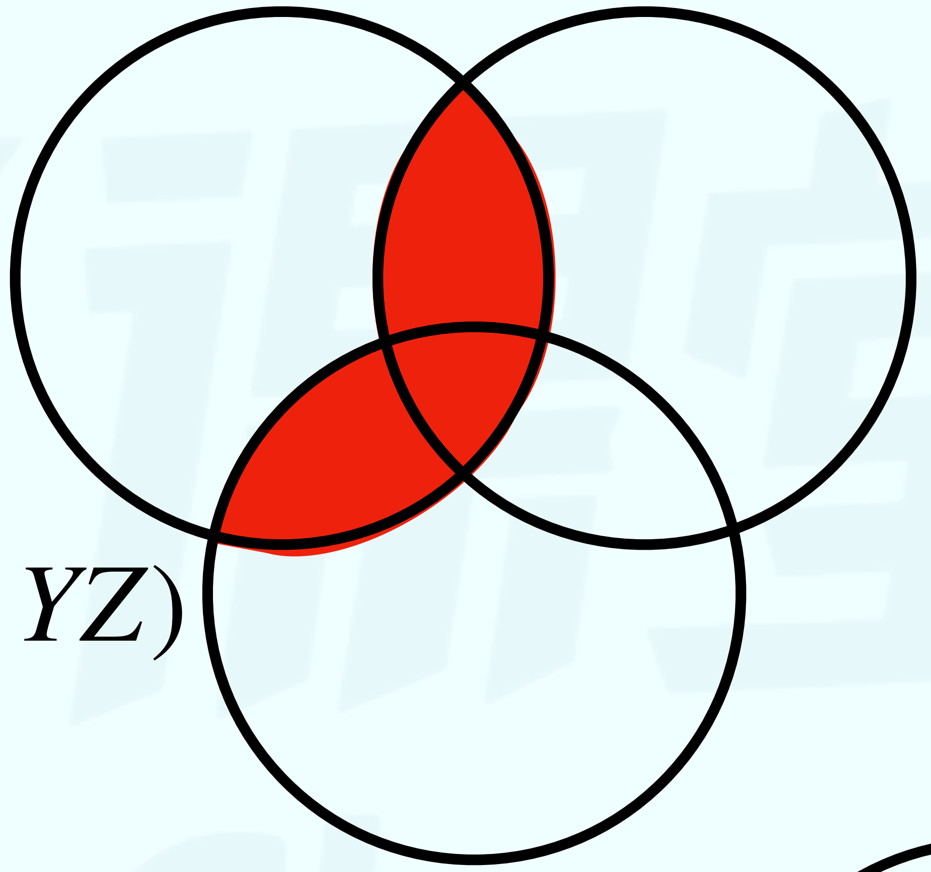
$$I(X;YZ)=H(X)-H(X|YZ)$$

利用维拉图表示平均条件互信息

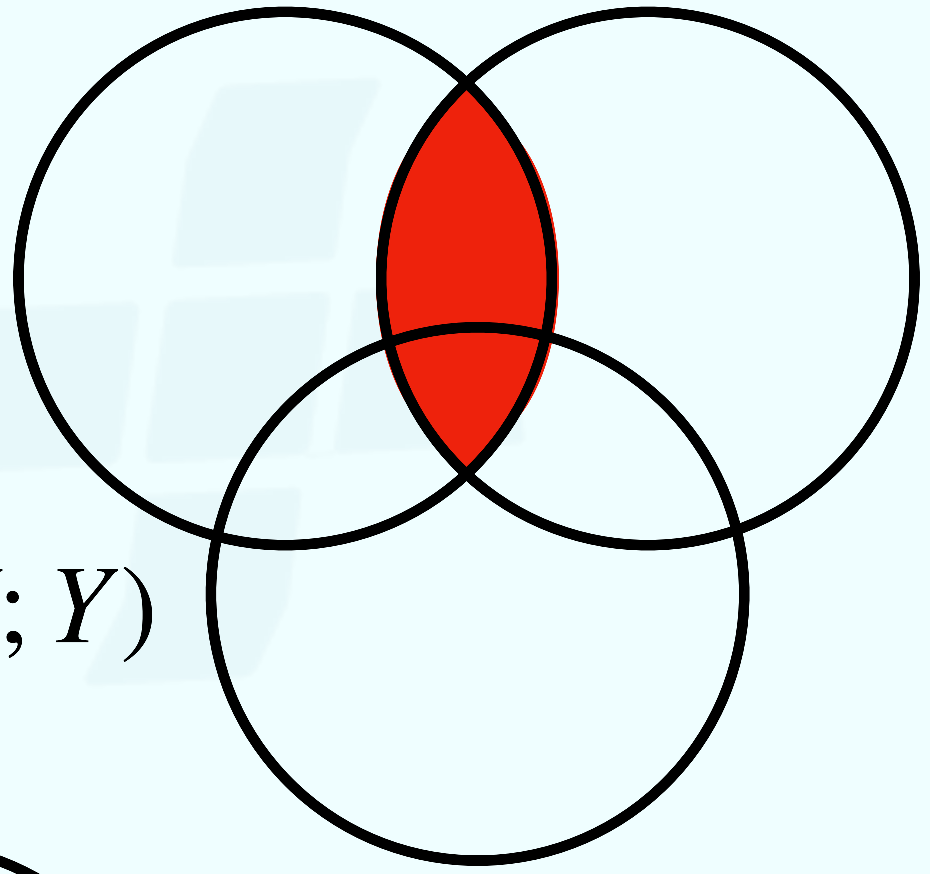
对于三变量X，Y和Z：

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

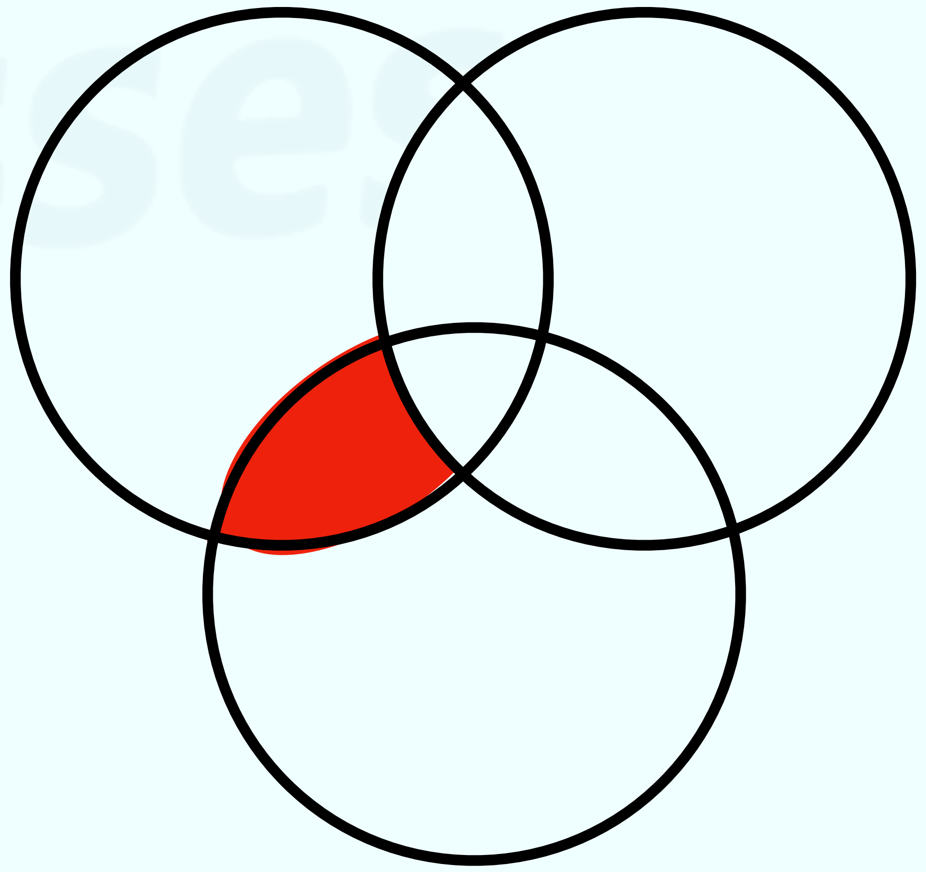




$I(X;YZ)$



$I(X;Y)$



$$I(X;YZ) = H(X) - H(X|YZ)$$

平均互信息、平均联合互信息、平均条件互信息小结

定义式

$$I(X;Y)=\sum_{i,j}p(x_iy_j)\log\frac{p(x_i|y_j)}{p(x_i)}$$

$$I(X;YZ)=\sum_{i,j,k}p(x_iy_jz_k)\log\frac{p(x_i|y_jz_k)}{p(x_i)}$$

$$I(X;Z|Y)=\sum_{i,j,k}p(x_iy_jz_k)\log\frac{p(x_i|z_ky_j)}{p(x_i|y_j)}$$

计算式

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

$$I(X;YZ)=H(X)-H(X|YZ)$$

$$I(X;Z|Y)=H(X|Y)-H(X|ZY)$$

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

例题3-4 对于任意三个离散随机变量X,Y,Z，试求证下列各式(不能用维拉图)：

(1) $H(XYZ)=H(XZ)+H(Y|X)-I(Z;Y|X)$

(2) $H(XYZ)-H(XY)\le H(XZ)-H(X)$

(3) $I(X;Y|Z)\ge 0$，当且仅当(X,Z,Y)是马氏链时等号成立。

解析3-4 (1) 根据熵函数的强可加性, $H(XYZ)=H(XZ)+H(Y|XZ)$；

根据平均条件互信息的计算公式$I(Z;Y|X)=I(Y;Z|X)=H(Y|X)-H(Y|XZ)$；

可得$H(Y|XZ)=H(Y|X)-I(Z;Y|X)$；

因此$H(XYZ)=H(XZ)+H(Y|X)-I(Z;Y|X)$，得证；

(2) 根据熵函数的强可加性, $H(XYZ)=H(XY)+H(Z|XY)$；

因此 $H(XYZ)-H(XY)=H(Z|XY)$ ；

根据条件越少，不确定度、即熵越大，有 $H(Z|XY)\le H(Z|X)$；

第二次使用熵函数的强可加性, $H(XZ)=H(X)+H(Z|X)$，即$H(Z|X)=H(XZ)-H(X)$；

因此$H(XYZ)-H(XY)=H(Z|XY)\le H(Z|X)=H(XZ)-H(X)$，得证。

例题3-4 对于任意三个离散随机变量X,Y,Z，试求证下列各式(不能用维拉图)：

(1) $H(XYZ)=H(XZ)+H(Y|X)-I(Z;Y|X)$

(2) $H(XYZ)-H(XY)\le H(XZ)-H(X)$

(3) $I(X;Y|Z)\ge 0$，当且仅当(X,Z,Y)是马氏链时等号成立。

解析3-4 (3) 根据平均条件互信息的计算公式$I(X;Y|Z)=H(X|Z)-H(X|YZ)$；

根据条件越多，不确定度、即熵越小，有$H(X|Z)\ge H(X|YZ)$；

因此有$I(X;Y|Z)\ge 0$，接下来我们判断等号成立的条件：

$I(X;Y|Z)=0$时，$I(X;Y|Z)=H(X|Z)-H(X|YZ)=\sum\limits_{i,j,k}p(x_iy_jz_k)\log\dfrac{p(x_i|y_jz_k)}{p(x_i|z_k)}=0$

此时$\dfrac{p(x_i|y_jz_k)}{p(x_i|z_k)}=1$，则$\dfrac{p(x_i|y_jz_k)}{p(x_i|z_k)}=\dfrac{p(x_iy_jz_k)p(z_k)}{p(y_jz_k)p(x_iz_k)}=\dfrac{p(x_Iy_jz_k)}{p(x_iz_k)}\cdot\dfrac{p(z_k)}{p(y_jz_k)}=\dfrac{p(y_j|x_iz_k)}{p(y_j|z_k)}=1$

此时$p(y_j|x_iz_k)=p(y_j|z_k)$，即Y仅与Z有关，与X无关；

也就是(X,Z,Y)满足马尔可夫链。

信道容量及其计算

信道容量及其计算

小节1 信道容量的定义

小节2 信道容量的计算

信息传输率

信息传输率*R* 信道中平均每个符号所传输的信息量

平均互信息*I*(*X*;*Y*) 接收到符号Y后平均获得的关于X的信息量

两者在意义上完全等价，因此**信道的信息传输率本质上就是平均互信息**，即

$$R = I(X;Y) = H(X) - H(X|Y) \quad \text{单位：} bit/symbol$$

信息传输速率 R_t 指的是信道每秒钟平均传输的信息量，即 $R_t = \frac{1}{t} I(X;Y)$ ，单位：bit/sec

信道容量的定义

对于固定信道，平均互信息是输入信源概率的上凸函数，那么对于固定信道，存在一个最佳输入分布，使信息传输率$R=I(X;Y)$最大，这个最大值由信道特性决定。

那么在最佳输入分布时，最大信息传输率$R=I(X;Y)$定义为**信道容量，即** $C=\max\limits_{p(x)}\{I(X;Y)\}$

信道容量的单位是*bit*/*symbol*，**信息传输率达到信道容量时输入分布称为最佳输入分布**。

信道在单位时间内平均传输的最大信息量$C_t=\frac{1}{t}\max\limits_{p(x)}\{I(X;Y)\}$，单位：*bit*/*sec*。

关于信道容量的几点说明

1. 信道容量与信源的概率分布无关，是完全描述信道统计特性的参量
2. 信道容量是信道能够传输的最大信息率，信道输入为最佳概率分布时，信道的信息传输率刚好达到了该信道容量。

信道容量及其计算

小节1 信道容量的定义

小节2 信道容量的计算

几种特殊离散信道的容量

1 无噪无损信道（无噪一一对应信道）

① 输入输出一一对应

② 给定任意 x_i ,必然存在 y_j 使得 $p(y_j|x_i) = p(x_i|y_j) = 1$

③ $H(X|Y) = H(Y|X) = 0, \quad I(X;Y) = H(X) = H(Y)$

$$\begin{array}{ccc} a_1 & \xrightarrow{1} & b_1 \\ a_2 & \xrightarrow{1} & b_2 \\ a_3 & \xrightarrow{1} & b_3 \\ \vdots & & \vdots \\ a_r & \xrightarrow{1} & b_r \end{array}$$

$$P(a_i|b_j) = P(b_j|a_i) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \Longleftrightarrow P = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

$r = s$
单位矩阵

此时信源的最佳输入分布为等概分布

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

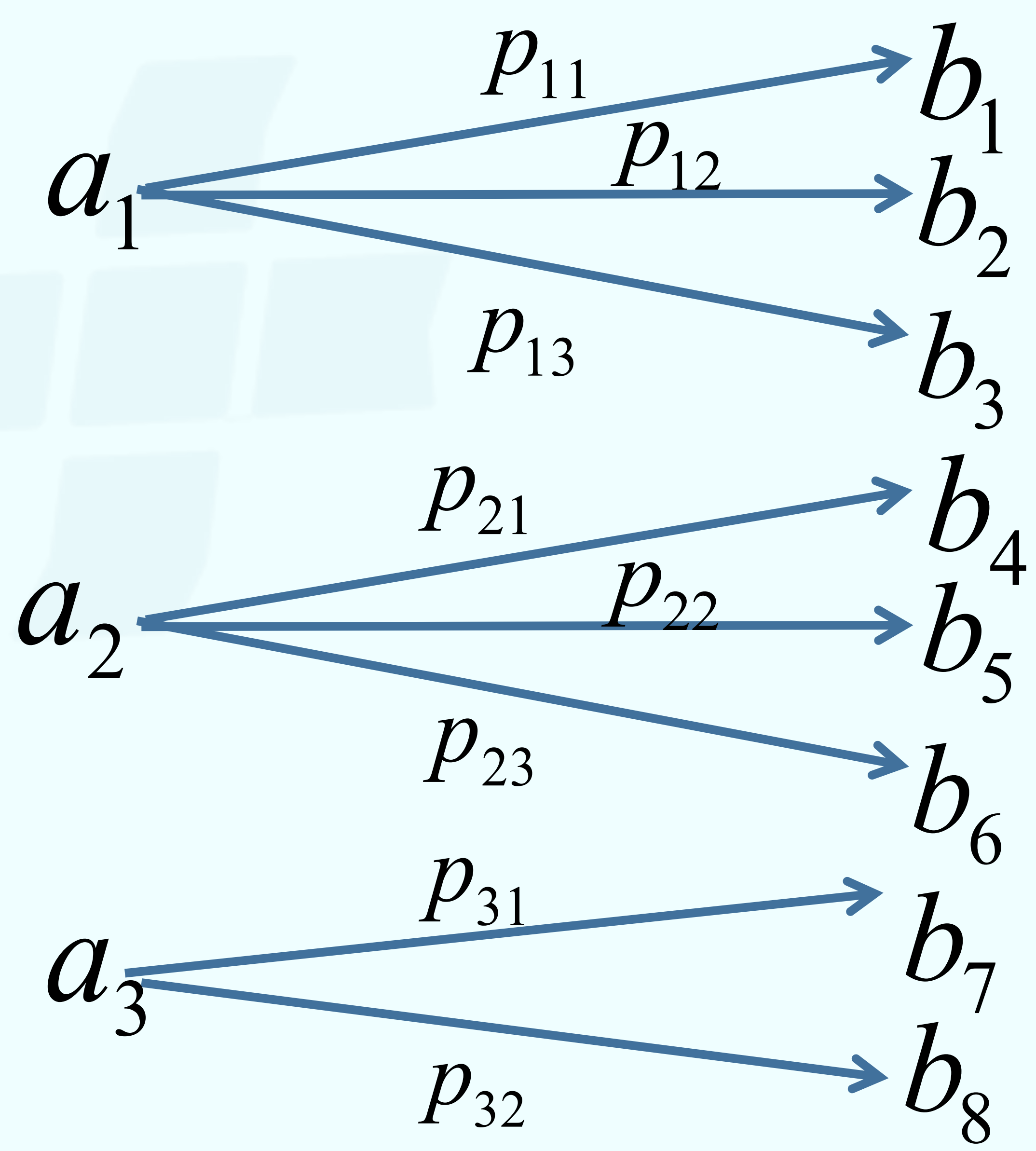
几种特殊离散信道的容量

2 有噪无损信道

① 一个输入对应多个互不相交的输出

② 给定任意输出y_j,必然存在输入x_i使得 $p(x_i|y_j)=1$ ，反之不一定

③ 损失熵 $H(X|Y)=0,\quad I(X;Y)=H(X)$



$$p=\begin{bmatrix} p_{11} & p_{12} & p_{13} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_{21} & p_{22} & p_{23} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{31} & p_{32} \end{bmatrix}$$

此时信源的最佳输入分布为等概分布

$$C=\max_{P(X)}\{I(X;Y)\}=\max_{P(X)}H(X)=\log r$$ （取决于输入符号的个数）

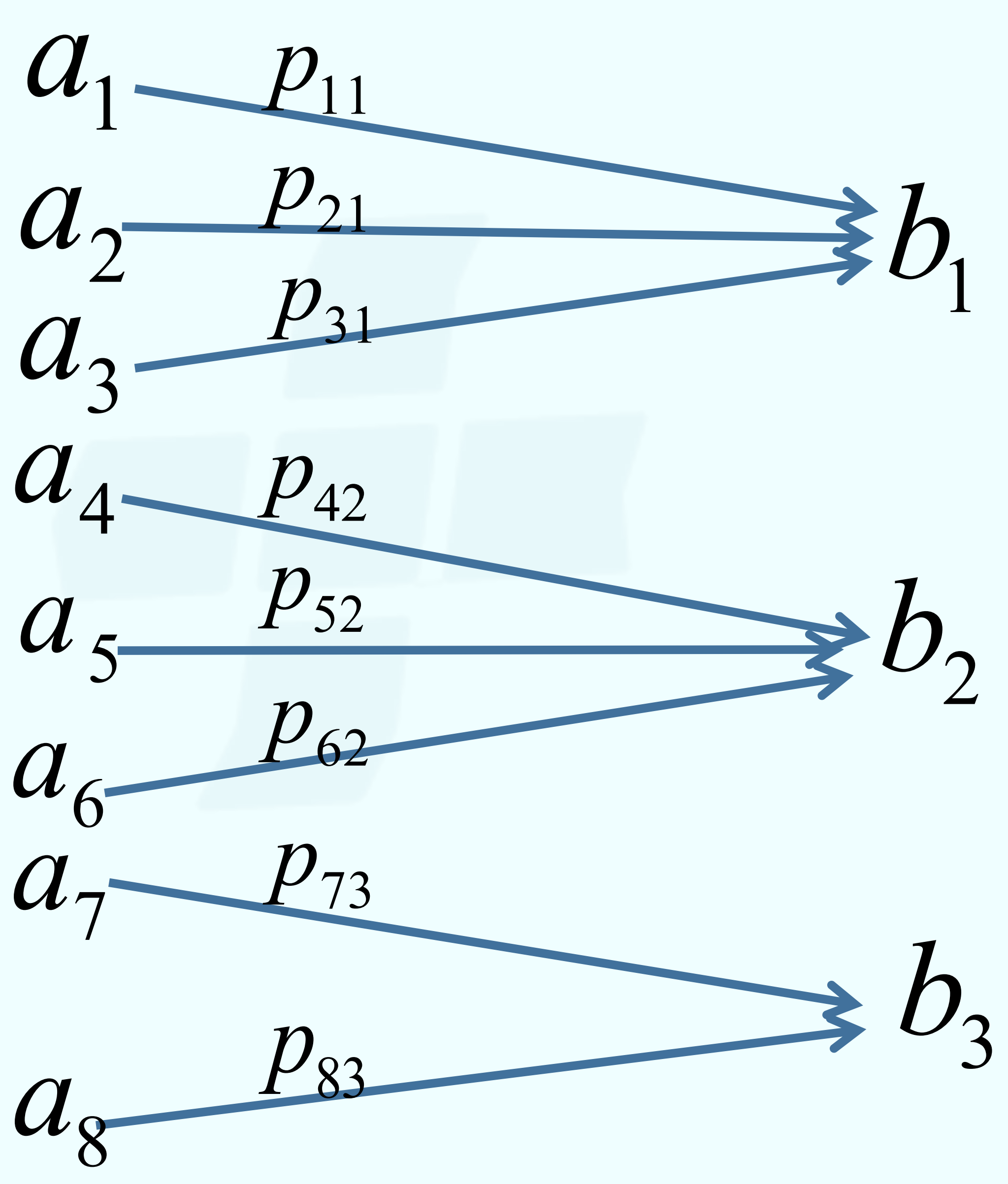
几种特殊离散信道的容量

3 无噪有损信道（确定信道）

① 多个输入对应一个输出

② 给定任意输入x_i ，必然存在一个输出 y_j 使得$p(y_j|x_i)=1$

③ 噪声熵 $H(Y|X)=0,\quad I(X;Y)=H(Y)$



$$p=\begin{bmatrix} p_{11} & 0 & 0 \\ p_{21} & 0 & 0 \\ p_{31} & 0 & 0 \\ 0 & p_{42} & 0 \\ 0 & p_{52} & 0 \\ 0 & p_{62} & 0 \\ 0 & 0 & p_{73} \\ 0 & 0 & p_{83} \end{bmatrix}=\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

$$C=\max_{P(X)}\{I(X;Y)\}=\max_{P(X)}H(Y)=\log s$$

（取决于输出符号的个数）

离散（准）对称信道的容量

1 离散输入对称信道（行对称信道）

信道矩阵**每一行**都是相同元素的排列

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

2 离散输出对称信道（列对称信道）

给信道矩阵**每一列**都是相同元素的排列

$$P=\begin{bmatrix}0.4 & 0.6\\ 0.6 & 0.4\\ 0.5 & 0.5\end{bmatrix}$$

3 对称信道

每行都是相同元素的排列

每列都是相同元素的排列

若行数小于列数，则列元素集合为行元素子集

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

离散（准）对称信道的容量

I 强对称信道（均匀信道）

① 行数等于列数

② 对于每个输入符号，正确传递概率相等，错误传递概率p均匀分配到r-1个符号

③ 为$r\times r$阶对称矩阵

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

II 准对称信道

① 准对称信道不是对称信道；

② 可以按列（按输出）划分出一些对称的子矩阵，即对输出集Y进行划分；

③ 划分子集只有一个时，就是对称信道；

$$p=\begin{bmatrix} \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{13}{60} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{3} & \frac{13}{60} \\ \frac{1}{5} & \frac{1}{3} & \frac{1}{4} & \frac{13}{60} \end{bmatrix}$$

离散（准）对称信道的容量

- **重要引理**

对于对称信道，当信道输入概率分布为等概分布时，输出概率分布必为等概分布。此引理对离散列对称信道亦成立。

对称信道的信道容量

当信道输入概率分布为等概分布时，输出也为等概分布，此时达到信道容量。

$$C = \max_{P(X)}[H(Y) - H(Y|X)] = \log s - H(p_1', p_2', \cdots, p_s')$$

其中 $p_1', p_2', \cdots, p_s'$ 为信道矩阵中**任意一行**的元素

例题3-5 若设某离散对称信道的信道矩阵如下，试求该信道的信道容量与对应最佳概率分布。

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

解析3-5

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

$$=\log 3-H(\frac{1}{2},\frac{1}{3},\frac{1}{6})$$

$$=\log 3-(-\frac{1}{2}\log\frac{1}{2}-\frac{1}{3}\log\frac{1}{3}-\frac{1}{6}\log\frac{1}{6})$$

$$\approx 0.126bit/symbol$$

此时对应的最佳输入分布为 $P_x=\begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{bmatrix}$

例题3-6 已知二元对称信道的传递概率矩阵，试求：

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

（1） $P(0) = \frac{3}{4}, P(1) = \frac{1}{4}$ 时的 $H(X), H(X|Y), H(Y|X), I(X;Y)$;

（2） 求该信道的信道容量以及达到信道容量时的输入概率分布。

解析3-6 （1） $P_X = [\frac{3}{4} \quad \frac{1}{4}]$，则输入熵 $H(X) = -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} = 0.81 bit/symbol$

输出概率分布 $P_Y = P_X P_{Y|X} = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \end{bmatrix} \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix} = \begin{bmatrix} \frac{7}{12} & \frac{5}{12} \end{bmatrix}$

则噪声熵为 $H(Y|X) = \sum_{i=1}^{2} p(x_i)H(Y|x_i) = \frac{3}{4}H(\frac{2}{3},\frac{1}{3}) + \frac{1}{4}H(\frac{2}{3},\frac{1}{3}) = H(\frac{2}{3},\frac{1}{3}) = 0.918 bit/symbol$

$$P_{X|Y} = \begin{bmatrix} p(x_1|y_1) & p(x_2|y_1) \\ p(x_1|y_2) & p(x_2|y_2) \end{bmatrix} = \begin{bmatrix} \frac{p(x_1y_1)}{p(y_1)} & \frac{p(x_2y_1)}{p(y_1)} \\ \frac{p(x_1y_2)}{p(y_2)} & \frac{p(x_2y_2)}{p(y_2)} \end{bmatrix}$$

例题3-6 已知二元对称信道的传递概率矩阵，试求：

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

（1） $P(0)=\frac{3}{4},P(1)=\frac{1}{4}$ 时的 $H(X),H(X|Y),H(Y|X),I(X;Y)$;

（2） 求该信道的信道容量以及达到信道容量时的输入概率分布。

解析3-6 （1） $P_{X|Y}=\begin{bmatrix}p(x_1|y_1) & p(x_2|y_1)\\ p(x_1|y_2) & p(x_2|y_2)\end{bmatrix}=\begin{bmatrix}\frac{p(x_1y_1)}{p(y_1)} & \frac{p(x_2y_1)}{p(y_1)}\\ \frac{p(x_1y_2)}{p(y_2)} & \frac{p(x_2y_2)}{p(y_2)}\end{bmatrix}$

$$=\begin{bmatrix}\frac{p(y_1|x_1)p(x_1)}{p(y_1)} & \frac{p(y_1|x_2)p(x_2)}{p(y_1)}\\ \frac{p(y_2|x_1)p(x_1)}{p(y_2)} & \frac{p(y_2|x_2)p(x_2)}{p(y_2)}\end{bmatrix}=\begin{bmatrix}\frac{6}{7} & \frac{1}{7}\\ \frac{3}{5} & \frac{2}{5}\end{bmatrix}$$

则损失熵为 $H(X|Y)=\sum_{j=1}^{2}p(y_j)H(X|y_j)=\frac{7}{12}H(\frac{6}{7},\frac{1}{7})+\frac{5}{12}H(\frac{3}{5},\frac{2}{5})=0.749bit/symbol$

因此平均互信息 $I(X;Y)=H(X)-H(X|Y)=0.061bit/symbol$

例题3-6 已知二元对称信道的传递概率矩阵，试求：

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

（1） $P(0) = \frac{3}{4}, P(1) = \frac{1}{4}$ 时的 $H(X), H(X|Y), H(Y|X), I(X;Y)$;

（2） 求该信道的信道容量以及达到信道容量时的输入概率分布。

解析3-6 (2) 此信道为对称信道，故信道容量

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

$$= \log 2 - H(\frac{1}{3}, \frac{2}{3})$$

$$\approx 0.082 bit/symbol$$

此时对应的最佳输入分布为等概率分布，即 $P_X = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$

离散（准）对称信道的容量

均匀信道的信道容量

均匀信道为对称信道，因此，其信道容量

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

$$C = \log s - H(p_1', p_2', \cdots, p_s') = \log r - H(\bar{p}, \frac{p}{r-1}, \frac{p}{r-1}, \cdots, \frac{p}{r-1})$$

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

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

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

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

离散（准）对称信道的容量

准对称信道的信道容量

当信道**输入概率分布为等概分布**时，准对称信道达到信道容量。

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

设信道矩阵可以划分成n个子矩阵，则N_k为第k个子矩阵中行元素之和，M_k为第k个子矩阵中列元素之和。

离散（准）对称信道的容量计算通法

准对称信道和对称信道达到信道容量时对应的最佳输入分布均为等概分布。

在此，给出计算（准）对称信道信道容量的统一方法，减轻记忆公式的负担。

对于有r个符号的信源，其最佳输入分布为 $p(x_i)=\dfrac{1}{r}$；

根据输入输出与信道矩阵的关系，$P_Y=P_XP_{Y|X}$；

则信道容量为 $C=H(Y)-H(Y|X)$

一般信道容量需要运用条件极值进行求解，我们在此不做要求，如有需要可以查阅有关资料。

例题3-7 求下列两个信道的信道容量，并加以比较

$$\begin{bmatrix} \overline{p}-\varepsilon & p-\varepsilon & 2\varepsilon \\ p-\varepsilon & \overline{p}-\varepsilon & 2\varepsilon \end{bmatrix} \qquad \begin{bmatrix} \overline{p}-\varepsilon & p-\varepsilon & 2\varepsilon & 0 \\ p-\varepsilon & \overline{p}-\varepsilon & 0 & 2\varepsilon \end{bmatrix}$$

例题3-7 求下列两个信道的信道容量，并加以比较

$$\begin{bmatrix} \bar{p}-\varepsilon & p-\varepsilon & 2\varepsilon \\ p-\varepsilon & \bar{p}-\varepsilon & 2\varepsilon \end{bmatrix} \qquad \begin{bmatrix} \bar{p}-\varepsilon & p-\varepsilon & 2\varepsilon & 0 \\ p-\varepsilon & \bar{p}-\varepsilon & 0 & 2\varepsilon \end{bmatrix}$$

解析3-7 $C=\max\limits_{P(X)} I(X;Y)=\max\limits_{P(X)}[H(Y)-H(Y\mid X)]$

（1）输入为等概率分布时，准对称信道达到信道容量，此时 $P_X=\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\end{bmatrix}$

此时的输出概率分布 $P_Y=P_XP_{Y|X}=\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\end{bmatrix}\begin{bmatrix} \bar{p}-\varepsilon & p-\varepsilon & 2\varepsilon \\ p-\varepsilon & \bar{p}-\varepsilon & 2\varepsilon \end{bmatrix}=\begin{bmatrix}\frac{1}{2}-\varepsilon & \frac{1}{2}-\varepsilon & 2\varepsilon\end{bmatrix}$

因此输出熵为 $H(Y)=H(\frac{1}{2}-\varepsilon,\frac{1}{2}-\varepsilon,2\varepsilon)$

例题3-7 求下列两个信道的信道容量，并加以比较

$$\begin{bmatrix} \overline{p}-\varepsilon & p-\varepsilon & 2\varepsilon \\ p-\varepsilon & \overline{p}-\varepsilon & 2\varepsilon \end{bmatrix} \qquad \begin{bmatrix} \overline{p}-\varepsilon & p-\varepsilon & 2\varepsilon & 0 \\ p-\varepsilon & \overline{p}-\varepsilon & 0 & 2\varepsilon \end{bmatrix}$$

解析3-7 $C = \max\limits_{P(X)} I(X;Y) = \max\limits_{P(X)}[H(Y) - H(Y|X)]$

（1）输入为等概率分布时，准对称信道达到信道容量，此时 $P_X = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$

因此输出熵为 $H(Y) = H(\frac{1}{2}-\varepsilon, \frac{1}{2}-\varepsilon, 2\varepsilon)$

求得 $H(Y|X) = \sum\limits_{i=1}^{2} p(x_i)H(Y|x_i) = \frac{1}{2}H(\overline{p}-\varepsilon, p-\varepsilon, 2\varepsilon) \times 2 = H(\overline{p}-\varepsilon, p-\varepsilon, 2\varepsilon)$

信道容量 $C_1 = H(Y) - H(Y|X) = H(\frac{1}{2}-\varepsilon, \frac{1}{2}-\varepsilon, 2\varepsilon) - H(\overline{p}-\varepsilon, p-\varepsilon, 2\varepsilon)$

$$= -(\frac{1}{2}-\varepsilon)\log(\frac{1}{2}-\varepsilon) \times 2 - 2\varepsilon\log 2\varepsilon - [-(\overline{p}-\varepsilon)\log(\overline{p}-\varepsilon) - (p-\varepsilon)\log(p-\varepsilon) - 2\varepsilon\log 2\varepsilon]$$

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

例题3-7 求下列两个信道的信道容量，并加以比较

$$\begin{bmatrix} \bar{p}-\varepsilon & p-\varepsilon & 2\varepsilon \\ p-\varepsilon & \bar{p}-\varepsilon & 2\varepsilon \end{bmatrix} \qquad \begin{bmatrix} \bar{p}-\varepsilon & p-\varepsilon & 2\varepsilon & 0 \\ p-\varepsilon & \bar{p}-\varepsilon & 0 & 2\varepsilon \end{bmatrix}$$

解析3-7 $C = \max\limits_{P(X)} I(X;Y) = \max\limits_{P(X)}[H(Y) - H(Y|X)]$

（2）输入为等概率分布时，达到信道容量，此时 $P_X = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$

此时的输出概率分布 $P_Y = P_X P_{Y|X} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \bar{p}-\varepsilon & p-\varepsilon & 2\varepsilon & 0 \\ p-\varepsilon & \bar{p}-\varepsilon & 0 & 2\varepsilon \end{bmatrix} = [\frac{1}{2}-\varepsilon, \frac{1}{2}-\varepsilon, \varepsilon, \varepsilon]$

因此输出熵为 $H(Y) = H(\frac{1}{2}-\varepsilon, \frac{1}{2}-\varepsilon, \varepsilon, \varepsilon)$

求得 $H(Y|X) = \sum_{i=1}^{2} p(x_i) H(Y|x_i) = \frac{1}{2} H(\bar{p}-\varepsilon, p-\varepsilon, 2\varepsilon) \times 2 = H(\bar{p}-\varepsilon, p-\varepsilon, 2\varepsilon)$

信道容量 $C_2 = H(Y) - H(Y|X) = H(\frac{1}{2}-\varepsilon, \frac{1}{2}-\varepsilon, \varepsilon, \varepsilon) - H(\bar{p}-\varepsilon, p-\varepsilon, 2\varepsilon)$

$$= -(-\frac{1}{2}-\varepsilon)\log(\frac{1}{2}-\varepsilon) \times 2 - 2\varepsilon\log\varepsilon - [-(\bar{p}-\varepsilon)\log(\bar{p}-\varepsilon) - (p-\varepsilon)\log(p-\varepsilon) - 2\varepsilon\log 2\varepsilon]$$

$$= (1-2\varepsilon)\log(\frac{2}{1-2\varepsilon}) + (p-\varepsilon)\log(p-\varepsilon) + (\bar{p}-\varepsilon)\log(\bar{p}-\varepsilon) + 2\varepsilon = C_1 + 2\varepsilon$$

离散无记忆扩展信道

小节1 数学模型

小节2 平均互信息

离散无记忆扩展信道

小节1 数学模型

小节2 平均互信息

离散无记忆扩展信道的数学模型

输入序列

$$\vec{\alpha}_i = (a_{i1}a_{i2}\cdots a_{iN})$$
$$a_i \in \{a_1a_2\cdots a_r\}$$

$$\vec{\alpha}_1 = (a_1a_1\cdots a_1)$$
$$\cdots\cdots$$
$$\vec{\alpha}_{r^N} = (a_ra_r\cdots a_r)$$

$$\longrightarrow \boxed{P(\vec{\beta}_h \mid \vec{\alpha}_k)} \longrightarrow$$

输出序列

$$\vec{\beta}_j = (b_{j1}b_{j2}\cdots b_{jN})$$
$$b_j \in \{b_1b_2\cdots b_s\}$$

$$\vec{\beta}_1 = (b_1b_1\cdots b_1)$$
$$\cdots\cdots$$
$$\vec{\beta}_{s^N} = (b_sb_s\cdots b_s)$$

一般离散信道输入和输出是一个随机变量序列；

无记忆扩展：每一个随机变量均取值于同一输入或输出符号集。

离散无记忆扩展信道的传递概率矩阵

类似的，我们可以得到离散无记忆扩展信道的传递概率矩阵为

$$[P^N]=\begin{bmatrix} p(\beta_1|\alpha_1) & p(\beta_2|\alpha_1) & \cdots & p(\beta_{s^N}|\alpha_1) \\ p(\beta_1|\alpha_2) & p(\beta_2|\alpha_2) & \cdots & p(\beta_{s^N}|\alpha_2) \\ \vdots & \vdots & \ddots & \vdots \\ p(\beta_1|\alpha_{r^N}) & p(\beta_2|\alpha_{r^N}) & \cdots & p(\beta_{s^N}|\alpha_{r^N}) \end{bmatrix}$$

其中，传递概率满足 $p_{ij}=p(\beta_j|\alpha_i)\geq 0$ $\sum_{j=1}^{s^N} p_{ij}=1$，即矩阵中每行元素概率之和为1；

信道矩阵中，各传递概率一般通过测量的方法给定，对于离散无记忆信道，传递概率值为：

$$p(\beta_j|\alpha_i)=p(b_{j1}b_{j2}\cdots b_{jN}|a_{i1}a_{i2}\cdots a_{iN})=\prod_{k=1}^{N} p(b_{jk}|a_{ik})$$

离散无记忆扩展信道

小节1 数学模型

小节2 平均互信息

离散无记忆扩展信道的平均互信息

无记忆信道N次扩展信道的平均互信息定义为

$$I(\vec{X};\vec{Y}) = H(\vec{Y}) - H(\vec{Y}\,|\,\vec{X}) = \sum_{X^N Y^N} p(\alpha_k \beta_h)\log\frac{p(\beta_h\,|\,\alpha_k)}{p(\beta_h)}$$

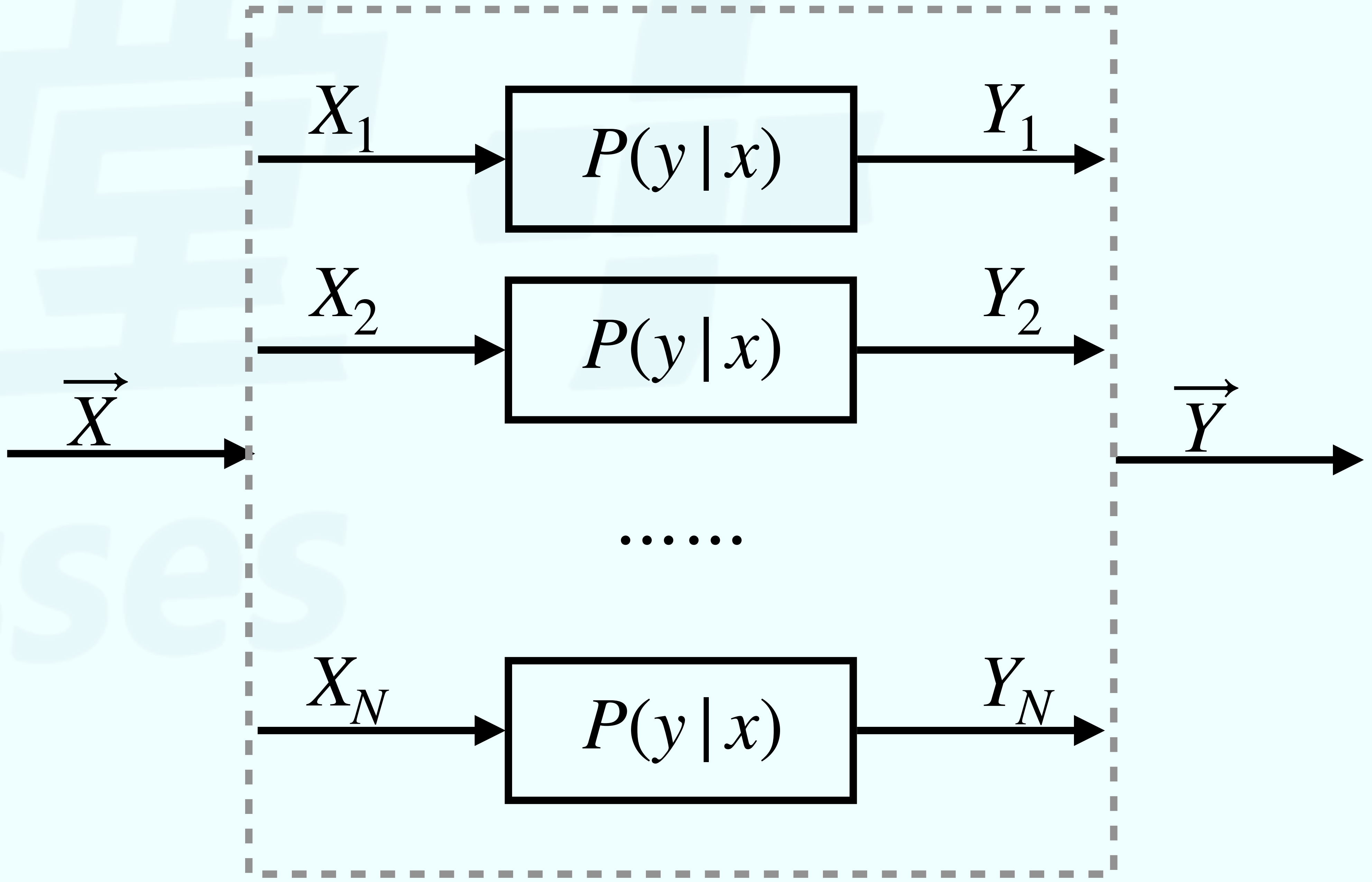
$$= \sum_{X^N Y^N} p(\alpha_k)p(\beta_h\,|\,\alpha_k)\log\frac{p(\beta_h\,|\,\alpha_k)}{p(\beta_h)}$$

$$= \sum_{X^N Y^N} p(\alpha_k)\prod_{i=1}^{N} p(b_{hi}\,|\,a_{ki})\log\frac{\prod_{i=1}^{N} p(b_{hi}\,|\,a_{ki})}{p(\beta_h)} \qquad k = 1,2,\ldots,r^N \quad h = 1,2,\ldots,s^N$$

讨论

$$I(\vec{X};\vec{Y}) = H(\vec{Y}) - H(\vec{Y}\,|\,\vec{X}) = \sum_{X^N Y^N} p(\alpha_k \beta_h)\log\frac{p(\beta_h\,|\,\alpha_k)}{p(\beta_h)}$$

信源无记忆，信道无记忆 $\quad I(\vec{X};\vec{Y}) = \sum_{k=1}^{N} I(X_k;Y_k)$



此时，信道可以等价于N个独立信道并联

证明从略。

讨论

$$I(\vec{X};\vec{Y}) = H(\vec{Y}) - H(\vec{Y}\,|\,\vec{X}) = \sum_{X^N Y^N} p(\alpha_k\beta_h)\log\frac{p(\beta_h|\alpha_k)}{p(\beta_h)}$$

信道无记忆 $$I(\vec{X};\vec{Y}) \le \sum_{k=1}^{N} I(X_k;Y_k)$$

信源无记忆，信道有记忆 $$I(\vec{X};\vec{Y}) \ge \sum_{k=1}^{N} I(X_k;Y_k)$$

信源的输入独立同分布 $$I(\vec{X};\vec{Y}) = \sum_{k=1}^{N} I(X_k;Y_k) = NI(X;Y)$$

证明从略，需要用到詹森不等式，可以直接作为结论记忆。

数据处理定理

小节1 级联信道

小节2 数据处理定理

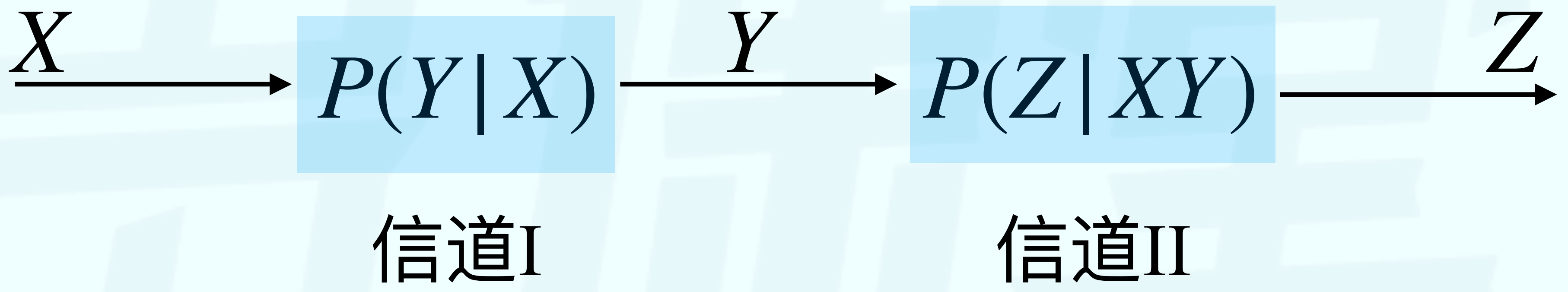
数据处理定理

小节1 级联信道

小节2 数据处理定理

级联信道及其平均互信息

级联信道的数学模型

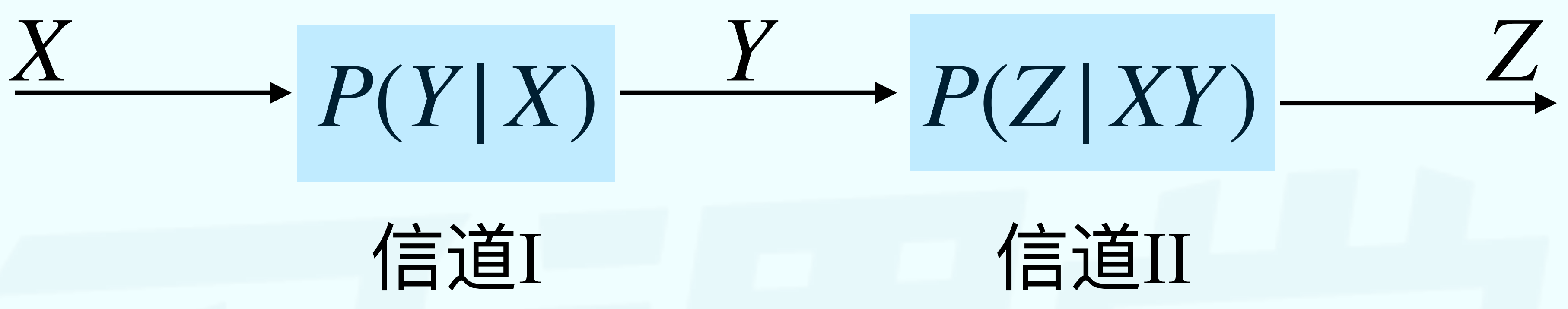


级联信道的平均互信息

收到Z后，收到的关于联合随机变量XY的平均信息量

$$I(XY;Z)=H(XY)-H(XY|Z)$$

级联信道及其平均互信息



定理一 $I(X;Z) \le I(XY;Z) \quad I(Y;Z) \le I(XY;Z)$

当且仅当 $p(z|xy) = p(z|y)$, $p(z|xy) = p(z|x)$ 时，两不等式分别取等号。

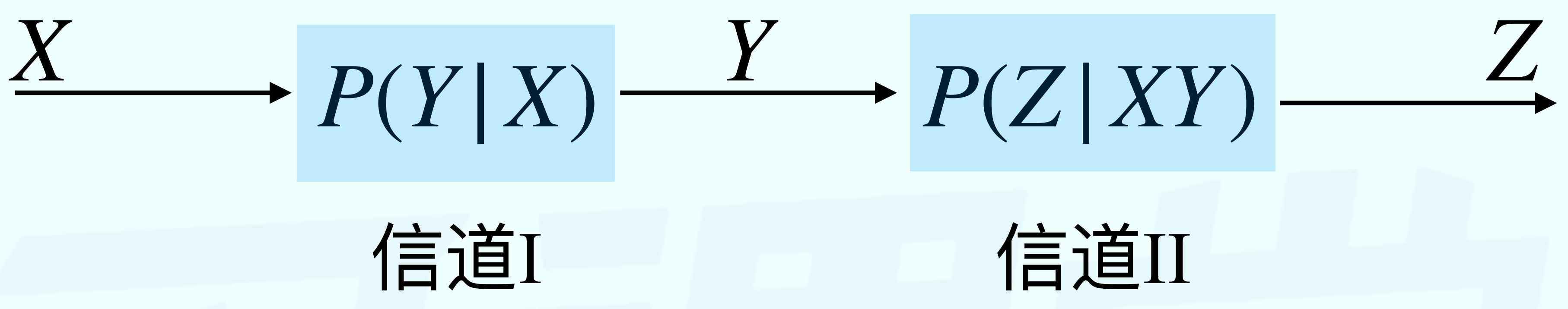
此定理表明：从Z中获得的关于X或Y的平均互信息，一般不会超过Z从联合随机变量XY中获得的平均互信息$I(XY;Z)$。

按照传递顺序，我们重点研究 $I(Y;Z) \le I(XY;Z)$

当且仅当 $p(z|xy) = p(z|y)$ 时，不等式取等号，此时，当随机变量Y确定后，随机变量Z只取决于随机变量Y，与先前的随机变量X无关，此时XYZ 构成马尔可夫链。

一般来说，串联信道中，随机变量序列XYZ 可构成马氏链，即Z与X没有直接的依赖关系。

级联信道及其平均互信息



定理二 若三随机变量*XYZ*满足马尔可夫链

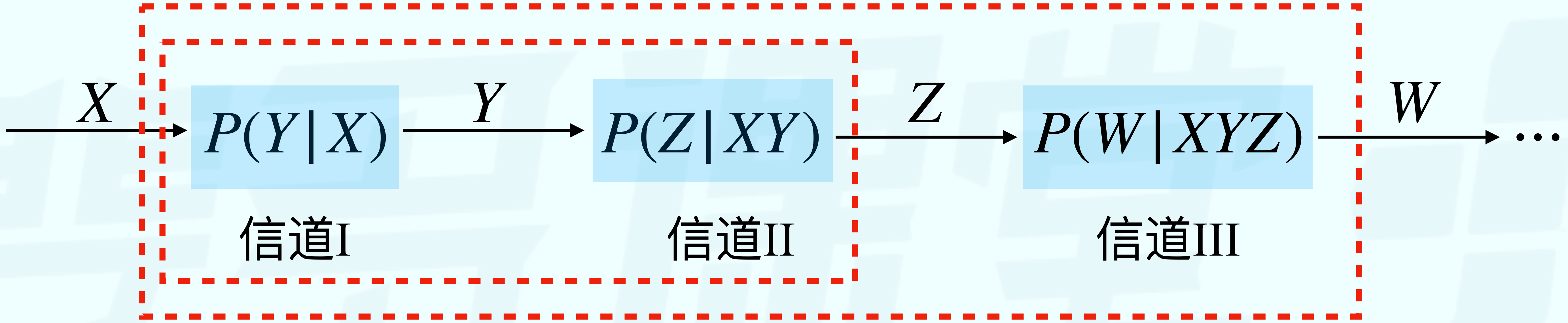
$$I(X;Z) \le I(X;Y) \qquad I(X;Z) \le I(Y;Z)$$

信道上传输的总信息量必小于等于两个子信道各自传输的信息量；

此时级联信道的矩阵等于两个子信道矩阵相乘，即：

$$P = P_1 P_2 \qquad P_{Z|X} = P_{Y|X} P_{Z|Y}$$

级联信道的信道容量



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

$$C_{I,II} = \max I(X;Z) \qquad C_{I,II,III} = \max I(X;W)$$

……

数据处理定理

小节1 级联信道

小节2 数据处理定理

数据处理定理「信息不增性原理」

$$X \longrightarrow \boxed{P(Y|X)} \xrightarrow{\;Y\;} \boxed{P(Z|XY)} \xrightarrow{\;Z\;}$$

信道I　　　　信道II

XYZ 满足马尔可夫链　$I(X;Z) \le I(X;Y)$　　$I(X;Z) \le I(Y;Z)$

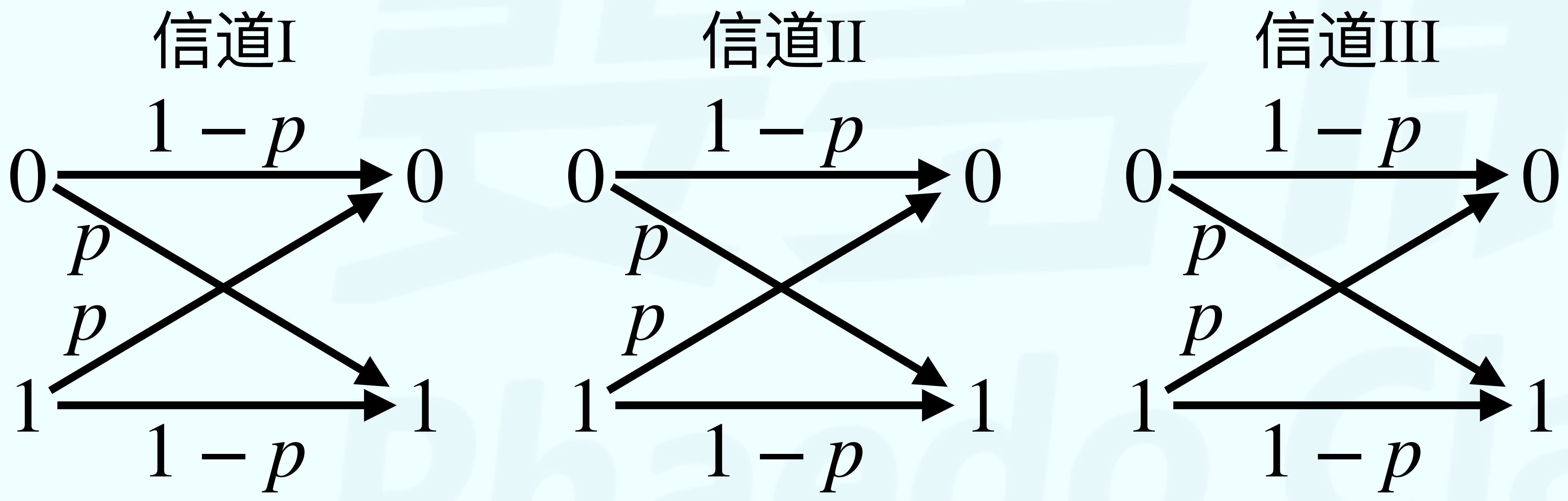
- 如果第二个信道是数据处理系统，则此定理表明：通过数据处理后，一般**只会增加信息的损失，最多保持原来获得的信息**，不可能比原来获得的信息有所增加。此为数据处理定理的基本内容。
- 在任何信息传输系统中，最后所获得的信息**至多是信源所提供的信息**。一旦在某一过程丢失一些信息，以后的系统不管如何处理，如不触及到丢失信息过程的输入端，就不能再恢复已丢失的信息。

数据处理定理示例

以二元对称信道的级联为例：

$$\xrightarrow{X} P(Y|X) \xrightarrow{Y} P(Z|XY) \xrightarrow{Z} P(W|XYZ) \xrightarrow{W} \cdots$$

$$P_{y|x} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$



$$P_{z|x} = P_{z|y}P_{y|x} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \cdot \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

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

通过第一级信道时：$I(X;Y) = 1 - H(p)$

通过第二级信道时：$I(X;Z) = 1 - H[2p(1-p)]$

通过第三级信道时：$I(X;W) = 1 - H[3p(1-p)^2 + p^2]$

本例说明：信道串联后增加信息损失，且串联级数越多，损失越大。

数据处理定理的说明

数据处理定理意味着对数据更为粗略的观测，任何无源（不涉及原始信源）的处理过程，一般来说总是会丢失信息的，最多保持原信息量不变。

如果想从测量数据中获得更多关于X的信息量，就必须进行有源处理。

- 一种方法是对X进行多次测量，扩大直接获得关于X的信息量。
- 通过付出再次测量的代价，使获得信息量增加了。测量次数越多，关于X的不确定性越小，获得的信息量越多。

在任何信息传输系统中，最后所获得的信息至多是信源所提供的信息。一旦在某一过程丢失一些信息，在后续的系统环节不管如何处理，只要不触及到丢失信息过程的输入端（即没有有源处理），就不能再恢复已丢失的信息。