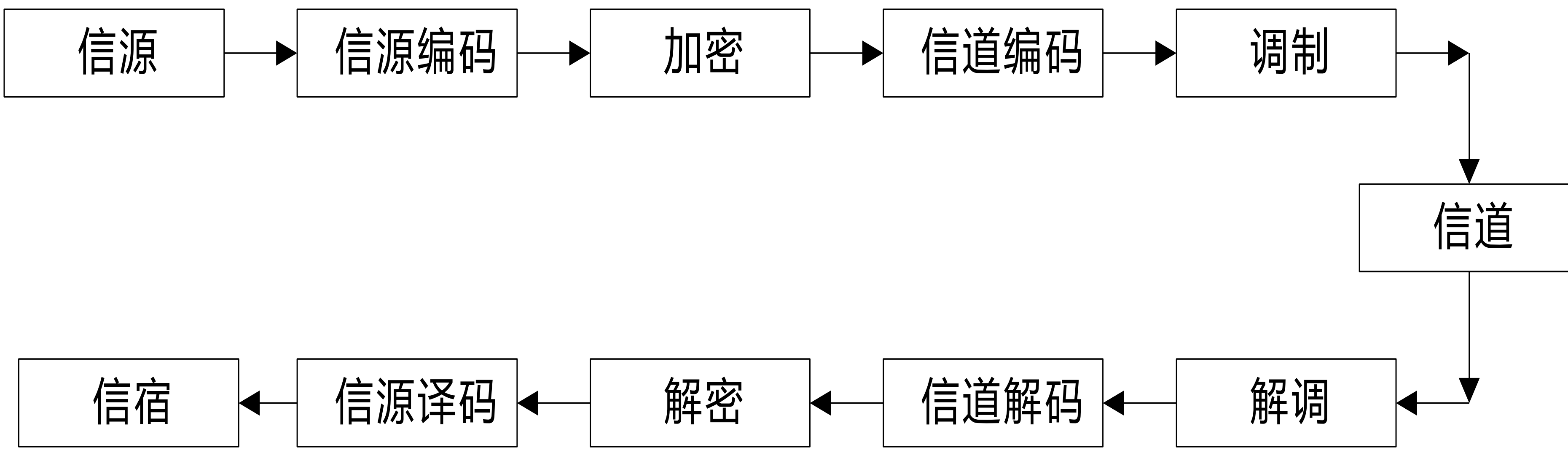
通信原理II的内容

- 信息论初步及在信源编码中的应用
- 信道及信道编码理论
- 信道编码
- 伪随机序列与扩频通信



第一讲 信源熵与互信息

信息与通信工程学院
无线信号处理与网络实验室（WSPN）
孙卓
zhuosun@bupt.edu.cn

前言

- 本章将介绍信息科学的基础理论和基本方法，基于一个通讯系统的抽象数学模型进行展开。整章可分为基础理论和编码理论两部分组成。
- 本部分以概率论为基础，数学推导较多，学习时主要把注意力集中到概念的理解上，不过分追求数学细节的推导。学习时一定要从始至终注意基本概念的理解，不断加深概念的把握。学习时注意理解各个概念的“用处”，结合其他课程理解它的意义，而不要把它当作数学课来学习，提倡独立思考，注重思考在学习中的重要性。

7.1 引言

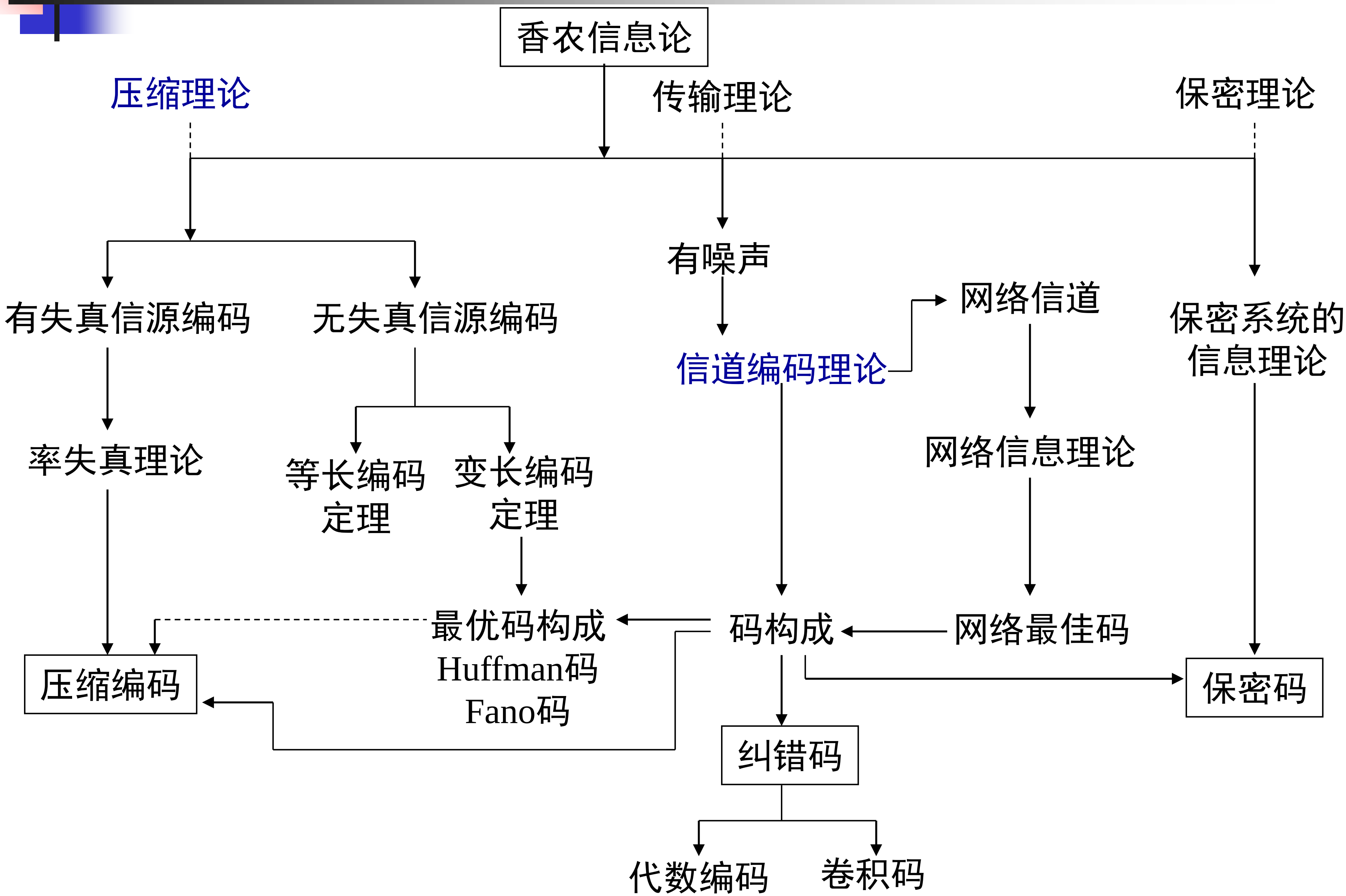
- 信息论研究内容
  - 信息论基础：香农信息论/狭义信息论，主要研究信息的测度、信道容量、信息率失真函数。代表人物 **Shannon**。
  - 一般信息论：主要研究信息的传输和处理问题，除香农基本理论外还包括噪声理论、信号滤波和预测、统计检测与估计理论、调制理论。代表人物 **Wiener**。
  - 广义信息论：凡是能够用广义通信系统模型描述的过程或系统，都能用信息基本理论来研究

参考书目

- Thomas M. Cover and Joy A. Thomas, "**Elements of Information theory**," John Wiley & Sons, Inc.

  中译本，"**信息论基础**" 阮吉寿 张华 译, 机械工业出版社, 2005.

香农信息论的科学体系


香农信息论
压缩理论
传输理论
保密理论
有噪声
有失真信源编码
无失真信源编码
网络信道
信道编码理论
保密系统的信息理论
率失真理论
等长编码定理
变长编码定理
网络信息理论
最优码构成
Huffman码
Fano码
码构成
网络最佳码
压缩编码
保密码
纠错码
代数编码
卷积码

本章学习内容

- 信源分类及其统计特性
- 信息熵，信源冗余度
- 互信息，各种熵与互信息的关系
- 信源编码定理
- 率失真函数
- 无失真/限失真信源编码
  - Huffman编码
  - 预测编码，变换编码

7.2 信源的数学模型及分类

信源消息对收信者而言在收到之前是不确定的，是随机的，所以可以用随机变量、随机矢量或随机过程来描述信源输出的消息，或者说用一个样本空间及其概率测度来描述信源。

eg： 足彩投注——欧冠比赛，赔率

玩法1：输赢结果

玩法2：进球数

$$Y \in \{\text{赢，输，平}\}$$

$$Z\ \square \left(\text{巴萨进球数，皇马进球数}\right) \subset Z^{+} \times Z^{+}$$

信源的数学模型及分类

- **信源分类**
  - 根据信源取值空间
    - 连续/模拟信源：取值空间为连续集合, eg: 温度，语音
    - 离散/数字信源：取值空间为离散集合, eg: 文字，比特流
  - 根据信源统计特性，离散信源分为
    - 无记忆信源：各时刻取值相互独立, eg: 掷骰子
    - 有记忆信源：各时刻取值相互有关联, eg: 语音
  - 根据信源统计特性与时间的推移是否有关
    - 平稳信源，eg: 高斯白噪声
    - 非平稳信源，eg: 天气温度

单消息(符号)信源

- 对信源的描述可以转化为用概率空间描述信源输出的消息。
- 对单符号离散信源，用离散型随机变量X的取值集合$\{x_1, x_2, \ldots x_i, \ \ldots x_n\}$及其取值概率为$p(x_i)$共同描述：

$$(X, P(X)) \quad 或 \quad \begin{pmatrix} X \\ P(X) \end{pmatrix} = \begin{pmatrix} x_1, & x_2, & \cdots, & x_n \\ P(x_1), & P(x_2), & \cdots, & P(x_n) \end{pmatrix}$$

- 对单符号连续信源，用连续性随机变量X与概率密度$p(x)$来描述：

$$\begin{pmatrix} X \\ p(x) \end{pmatrix} = \begin{pmatrix} x \in (a, b) \\ p(x) \end{pmatrix}$$

消息(符号)序列信源

- **消息(符号)序列信源**
  - 对信源的描述采用随机序列/过程来描述信源输出的消息；
  - 根据取值类型可分为离散型、连续型；
  - 通过采样，模拟信号可离散化；
- **离散符号序列信源的统计特性描述**
  - 离散消息序列信源由L个离散符号构成，输出的消息序列是L维随机变量 $\boldsymbol{X}=\left(X_1\cdots X_l\cdots X_L\right)$，对应的$L$维联合概率为 $P\left(\boldsymbol{x}\right)=P\left(x_1\cdots x_l\cdots x_L\right)$
  - 统计描述可用消息序列的取值集合 X^L 及其对应的概率 $P\left(\boldsymbol{x}\right)$ 共同描述，即：

$$\left(X^L,P\left(\boldsymbol{x}\right)\right) \quad \text{或} \quad \begin{pmatrix} X^L \\ P\left(X\right) \end{pmatrix}=\begin{pmatrix} a_1, & a_2, & \cdots, & a_{n^L} \\ P\left(a_1\right), & P\left(a_2\right), & \cdots, & P\left(a_{n^L}\right) \end{pmatrix}$$

离散序列信源

- **离散无记忆序列信源**
  - 序列中前后符号相互统计独立，概率密度为：

$$P(x_1\cdots x_l\cdots x_L)=\prod_{l=1}^{L}P(x_l)\overset{(a)}{=}P^L$$

(a): 无记忆且平稳（符号统计特性与时间无关）

- **离散有记忆序列信源**
  - 序列中前后符号存在相关性，统计特性描述复杂；
  - 消息序列中任一符号仅与前k个符号存在直接统计关联: k阶马尔可夫链信源；
  - 马尔可夫链信源统计特性由**起始状态概率**与**条件转移概率矩阵**确定；
  - 齐次性：条件转移概率与时间无关；
  - 遍历性：转移步数足够长时，序列状态与起始状态无关。

7.3 信息熵

- 信息的基本特征：**不确定性**。因此信息应该是概率P的函数。
- 信息的两个特点
  - 随概率P的递减性:概率越大，信息量越少
  - 可加性：两个独立消息的总信息量应该是两个消息量的和。
- 满足这两个条件的表示信息量的函数只有一种可能：对数函数的负数

单消息离散信源的信息度量

- **自信息量**
  - 出现某个消息时的信息量

$$I\left[P(x_i)\right]=-\log_2 p(x_i)$$

- **条件自信息量**
  - 知道消息y_j的情况下，消息x_i新带来的信息量

$$I\left[P(x_i/y_j)\right]=-\log_2 p(x_i/y_j)$$

- **联合自信息量**
  - 两个消息x_i , y_j一共带来的信息量

$$I\left[P(x_iy_j)\right]=-\log_2 p(x_iy_j)$$

- **三者关系**

$$I\left[P(x_iy_j)\right]=I\left[P(x_i)\right]+I\left[P(y_j\mid x_i)\right]=I\left[P(y_i)\right]+I\left[P(x_j\mid y_i)\right]$$

信息熵$H(X)$

- 单消息离散信源的平均信息量
  - 自信息量定义的是一个具体消息的信息量，而信源输出的消息有多种可能性，所有可能输出消息的自信息量的平均定义为信息熵（平均自信息量）

$$H(x) = E[-\log p(x_i)] = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$$

  - 也可理解为信源的不确定性的平均度量
- 信息量和熵的单位
  - 对数以2为底时，单位为比特 (bit)
  - 对数以e为底时，单位为奈特 (nat)
  - 对数以10为底时，单位为笛特 (Det)
  - 1 bit = 0.693 Nat = 0.301 Det

信息熵$H(X)$

例：

PM2.5空气质量，有两个信源

$$\begin{bmatrix} X_1 \\ p(x) \end{bmatrix} = \begin{bmatrix} a_{1,} & a_2 \\ 1/4, & 3/4 \end{bmatrix} \quad \begin{bmatrix} X_2 \\ p(x) \end{bmatrix} = \begin{bmatrix} a_{1,} & a_2 \\ 1/2, & 1/2 \end{bmatrix}$$

分别有：

$$H(X_1) = \frac{1}{4}\log 4 + \frac{3}{4}\log\frac{4}{3} = 0.809$$

$$H(X_2) = \frac{1}{2}\log 2 + \frac{1}{2}\log 2 = 1$$

说明第二个信源的平均不确定性更大一些

信源熵

- 条件熵(知道一个符号/信源下，另一符号/信源带来的信息熵)
  - 条件自信息的期望

$$H(X/Y)=E\{I[P(x_i/y_j)]\}=E\{-\log P(x_i/y_j)\}=-\sum_{j=1}^{m}\sum_{i=1}^{n}p(x_iy_j)\log_2 p(x_i/y_j)$$

  - 可理解为：已知Y时为消除X的不确定性还需的信息量

- 联合熵(两个符号/信源带来的总信息熵)
  - 联合自信息的期望

$$H(X,Y)=E\{I[P(x_iy_j)]\}=E\{-\log P(x_iy_j)\}=-\sum_{j=1}^{m}\sum_{i=1}^{n}p(x_iy_j)\log_2 p(x_iy_j)$$

- 信息熵、条件熵与联合熵之间关系

$$H(X,Y)=H(X)+H(Y|X)=H(Y)+H(X|Y)$$

可理解为:两个符号先后到达时，两个符号的总信息熵=一个符号的信息熵+知道此符号条件下另一个符号带来的信息熵

例

例：离散信源*X*取值于{*A*,*B*}。信源每次输出一个符号，前后符号之间有统计相关性。前一次输出*X'*和当前输出*X*之间的转移概率*P(X|X')*为：*P(A|A)=0.8*, *P(B|A)=0.2*, *P(A|B)=0.6*, *P(B|B)=0.4*

(a) 求信源输出A或B的概率*P(A)*, *P(B)*

(b) 分别求前次输出为A或B时的熵*H(X|A)*和*H(X|B)*,并求*H(X|X')*

(c) 若信源的输出符号统计独立，且A、B出现概率相同，求*H(X|X')*

解：（a）由全概率公式,

$$P(X=A)=P(X'=A)P(X=A|X'=A)+P(X'=B)P(X=A|X'=B)$$

且有 $P(X=A)+P(X=B)=1$

解方程组可得 $P(X=A)=3/4, P(X=B)=1/4$

(b) $H(X|A)=-\left[P(A|A)\log P(A|A)+P(B|A)\log P(B|A)\right]=\log_2 5-\frac{8}{5}\approx 0.7$

$$H(X|B)=-\left[P(A|B)\log P(A|B)+P(B|B)\log P(B|B)\right]=\log_2 5-\frac{1}{5}\left[3\log_2 3+2\right]\approx 0.94$$

$$H(X|X')=P(X'=A)H(X|X'=A)+P(X'=B)H(X|X'=B)\approx\frac{0.94}{4}+\frac{3*0.7}{4}=0.76$$

（c）**独立信源的条件熵** $H(X|X')$ **等于无条件熵** $H(X)$, **等概信源的熵最大** $H(X)=1$

信源熵的特性

- 非负性
  - 信息熵是自信息的数学期望，自信息是非负值
- 对称性

$$H(X)=H[p(x_1),p(x_2),\ldots,p(x_n)]=-\sum_{i=1}^{n}p(x_i)\log_2 p(x_i)$$

  - 任意调换$p(x_i)$, $p(x_j)$的顺序上式仍成立，说明熵只与信源的总体结构有关，不在乎个别消息的概率，与消息的取值无关
- 扩展性

$$\lim_{\varepsilon\to 0}H_{n+1}[p(x_1),p(x_2),\ldots,p(x_n)-\varepsilon,p(x_{n+1})=\varepsilon](X)=H_n[p(x_1),p(x_2),\ldots,p(x_n)]$$

  - 一个随机变量取值有n种，另一种有$n+1$种，如后者中有一种取值的概率趋于零，且其他概率均与前者在总体上相等，则两者的熵值相同
  - 物理意义：信源的消息数可以很多，但如果某些离散消息对应的概率很小，可以忽略其对熵的贡献

信源熵的特性

- 可加性

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

- 极值性

$$H(p_1,p_2,\ldots,p_q)\le H(1/q,1/q,\ldots,1/q)=\log q$$

上式表明，对于具有q个符号的离散信源，只有在q个信源符号等可能出现的情况下，信源熵才能达到最大值，这也表明等概分布的信源的平均不确定性最大，这是一个很重要得结论，称为**最大离散熵定理**

例：对于一个二元信源

H(X)=H(1/2,1/2)=log2=1bit

各类熵的关系

- Shannon不等式：条件熵不大于信息熵

$$H(Y|X) \le H(Y)$$

理解：一个消息没有任何前兆时带来的信息大于有前兆时的信息

仅当X,Y独立时，等式成立。

也称为**熵的不增原理**：在信息处理过程中，条件越多，熵就越小。

证明：（用到相对熵$D(p\|q)$总为非负的性质）

$$H(Y)-H(Y|X) = -\sum_y \log p(y) + \sum_x \sum_y p(x)p(y|x)\log p(y|x)$$

$$= \sum_x p(x) \sum_y p(y|x) \log \frac{p(y|x)}{p(y)}$$

$$= \sum_x p(x) D(p(y|x) \| p(y))$$

各类熵的关系（续）

- 联合熵不大于各个信息熵的和，即

$$H(X_1X_2\cdots X_n)\leq\sum_{i=1}^{N}H(X_i)$$

证明：

$$H(X_1X_2\cdots X_n)=H(X_1)+H(X_2|X_1)+\cdots+H(X_n|X_1\cdots X_{n-1})$$

$$\leq\sum_{i=1}^{N}H(X_i)$$

回顾-信息的度量

- 信息的性质
  - 不确定性
  - 概率的减函数
  - 独立事件的信息相加
- 信息的度量：概率的负对数
  - 自信息，条件自信息，联合自信息

    $$I[P(x_i)] = -\log_2 p(x_i)$$

  - 信息熵：集合的概率统计平均

$$H(x) = E[-\log p(x_i)] = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i)$$

  - 条件熵 $H(X/Y) = E\{I[P(x_i/y_j)]\} = E\{-\log P(x_i/y_j)\}$

    $$= -\sum_{j=1}^{m}\sum_{i=1}^{n} p(x_i y_j)\log_2 p(x_i/y_j)$$

  - 联合熵 $H(X,Y) = E\{I[P(x_i y_j)]\} = E\{-\log P(x_i y_j)\}$

    $$= -\sum_{j=1}^{m}\sum_{i=1}^{n} p(x_i y_j)\log_2 p(x_i y_j)$$

- 熵的性质
  - 非负性
  - 最大熵
    - 离散—均匀分布
    - 连续—正态分布
  - 链式法则

$$H(X_1X_2\cdots X_n) = H(X_1) + H(X_2|X_1) + \cdots + H(X_n|X_1\cdots X_{n-1})$$

$$\le \sum_{i=1}^{N} H(X_i)$$

  - Shannon不等式：熵的不增原理

$$H(Y|X) \le H(Y)$$

  - 联合熵不大于各信息熵之和

信源冗余度

- L维离散平稳有记忆信源输出的消息序列的熵为：

$$H(\boldsymbol{X})=H(X_1,X_2,\cdots,X_L)=\sum_{i=1}^{L}H(X_i\,|X_{i-1},\cdots,X_1)$$

平均符号熵： $H_L(\boldsymbol{X})=\dfrac{1}{L}H(X_1,X_2,\cdots,X_L)$

- L 趋于无穷时：极限熵

$$H_\infty=\lim_{L\to\infty}H_L(X)=\lim_{L\to\infty}H(X_L\,|\,X_1X_2\ldots X_{L-1})$$

且有 $0\le H_\infty(\boldsymbol{X})\le H(X_2|X_1)\le H(X_1)\le H(X_0)=\log_2 N$

其中，$H(X_2|X_1)$为一维记忆长度时的信息熵；$H(X_1)$为无记忆不等概信息熵；$H(X_0)$为无记忆等概信源最大熵；$H_\infty(\boldsymbol{X})$为无限记忆长度信息序列平均每符号具有的信息熵。

- 理论上信道只需传送$H_\infty(\boldsymbol{X})$的信息量，收端利用信源统计关联的记忆特性就可恢复信源的全部信息；换言之，如不利用信源的统计特性，信道要多传送$H_0(\boldsymbol{X})-H_\infty(\boldsymbol{X})$的信息量

信源冗余度 (Cont’d)

- 信源效率：信源输出符号之间存在相关性与相互独立两种情况下每符号信息量比值。

$$\eta = \frac{H_\infty(\boldsymbol{X})}{H_0(\boldsymbol{X})}$$

- 信源冗余度：信源的相对冗余度。

$$R = 1 - \eta$$

- 冗余度越大，采用信源编码、数据压缩的必要性和潜力越大：传输的有效性。

例：英文字母出现概率统计表（表7.3.3)

7.4 互信息量

事件x_i是否发生具有不确定性，用$I(x_i)$度量。

接收到符号y_j后，事件x_i是否发生仍保留有一定的不确定性，用$I(x_i \mid y_j)$度量。

观察事件前后，这两者之差就是通信过程中所获得的信息量，称为事件x_i和事件y_j之间的互信息量，用$I(x_i\,;\,y_j)$表示：

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

- 互信息量：收到 y_j 获得了多少 x_i 的信息？
  - 物理意义：$I(x_i)$可以理解为x_i所含的信息，$I(x_i/y_j)$是在已知y_j的条件下x_i还能带来的信息量，二者之差即是由于已经知道y_j使得x_i减少的信息量，即从y_j得到关于x_i的信息量度

注：$I(x_i\,;\,y_j)$ 和$I(x_iy_j)$的区别为：前者是事件$x_i\in X$和事件$y_j\in Y$之间的互信息量，后者是二维空间XY上元素x_iy_j的自信息量(联合事件的自信息)。

互信息量

根据概率互换公式$p(x_i\ y_j) = p(y_j \mid x_i)p(x_i)=p(x_i \mid y_j)p(y_j)$ 互信息量$I(x_i;y_j)$有多种表达形式:

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

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

将事件互信息量的概念推广至多维空间:

在三维$X\ Y\ Z$联合集中，有:

$$I(x_i;y_j\ z_k)=I(x_i;y_j)+I(x_i;z_k \mid y_j)$$

类似，在N维$U_1\ U_2\ \ldots\ U_N$联合空间,有（**互信息量链式法则**）：

$$I(u_1;u_2u_3\cdots u_N)=\sum_{i=2}^{N}I(u_1;u_i|u_{i-1}\cdots u_1)$$

平均互信息

- **平均互信息：** $I\left(X;Y\right)=\sum_{i}\sum_{j}p(x_i y_j)\log\frac{p(x_i y_j)}{p(x_i)p(y_j)}$
- **性质**

1. **(非负性)** $I\left(X;Y\right)\geq 0$

$$I\left(X;Y\middle|Z\right)\geq 0$$

2. (互易性) $I(X\,;\,Y)=I(Y\,;\,X)$

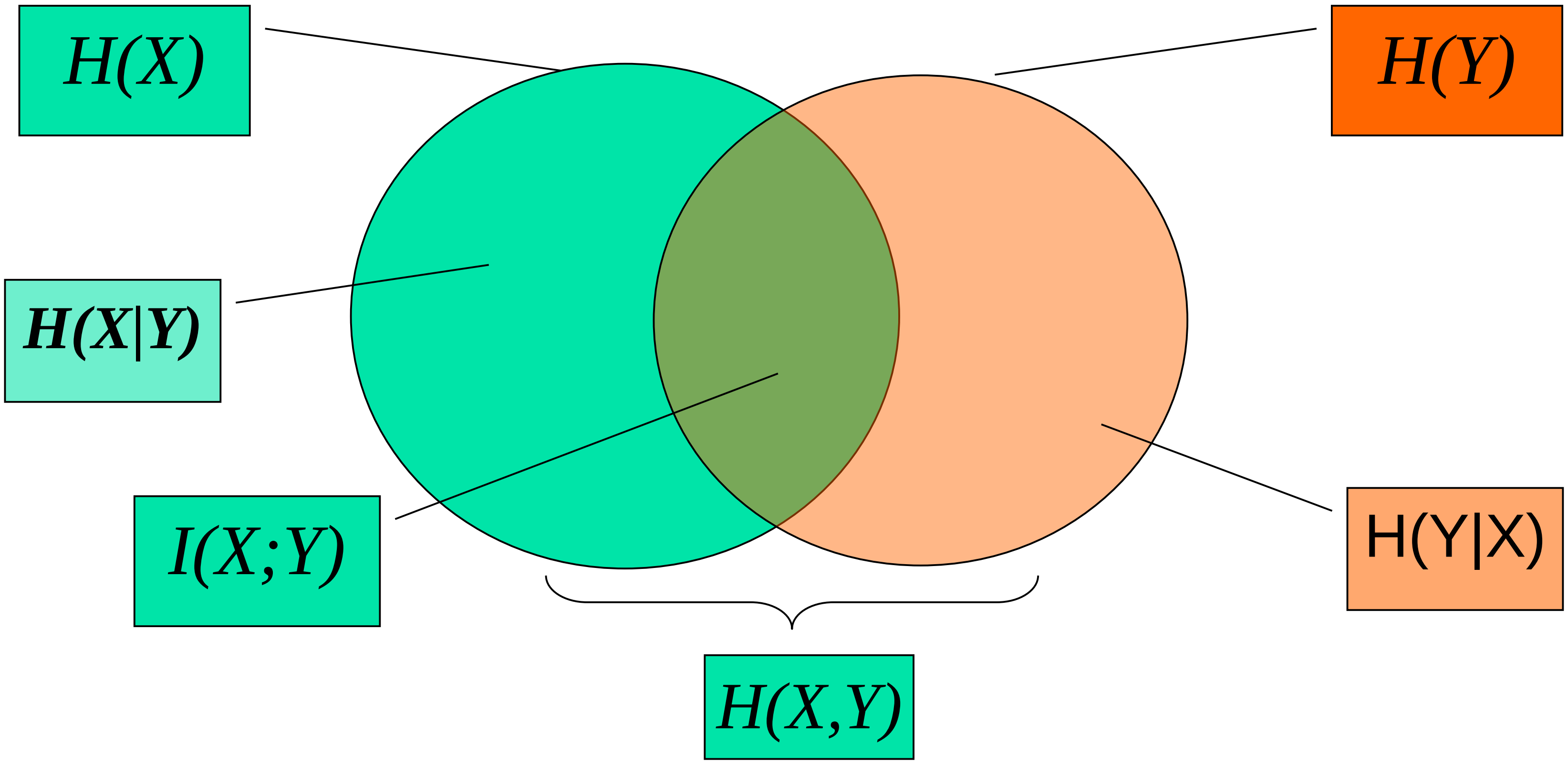
3. $I\left(X;Y\right)\leq H\left(X\right)$

$$I\left(X;Y\right)\leq H\left(Y\right)$$

对于无扰信道，$I(X\,;\,Y)=H(X)$。

对于强噪信道，$I(X\,;\,Y)=0$。

各种熵之间的关系



平均互信息量

例： BSC信道的信源分布为$\begin{bmatrix} X \\ p(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \delta & 1-\delta \end{bmatrix}$，信道转移概率矩阵为$P = \begin{bmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{bmatrix}$。试计算平均互信息量 $I(X;Y)=H(Y)-H(Y|X)$.

解：由$p(y_j) = \sum_i p(x_i)p(y_j|x_i)$可计算接收信号$Y$的概率分布为

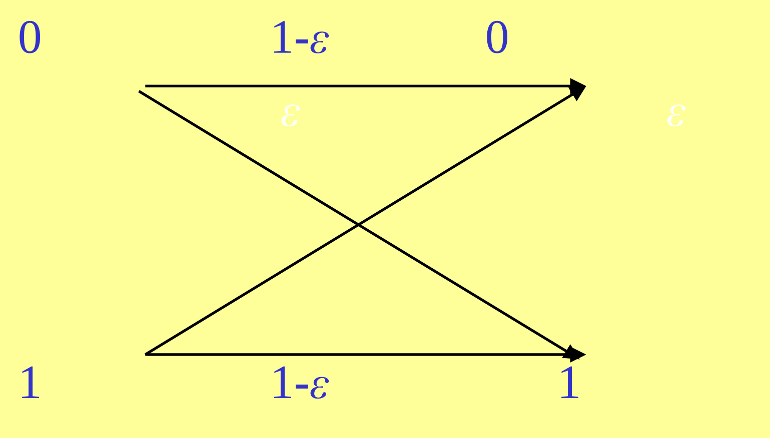
$$p(Y=0) = \varepsilon(1-\delta) + \delta(1-\varepsilon)$$

$$p(Y=1) = \varepsilon\delta + (1\text{-}\delta)(1\text{-}\varepsilon)$$

然后计算熵与条件熵

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

$$H(Y|X) = -\sum_i \sum_j p(x_i)p(y_j|x_i)\log p(y_j|x_i)$$



二进制对称信道

则平均互信息量为：

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