第2章　信息的度量

主要内容

1 、信源的数学模型及分类

根据参数集和值域是离散集合还是连续区间，可将信源分为四类：

(1) 时间离散空间离散信源

(2) 时间离散空间连续信源

(3) 时间连续空间离散信源

(4) 时间连续空间连续信源

简单的表示：

空间离散信源→离散信源

空间连续信源→连续信源

根据信源在 t_k 时刻的输出 X_{t_k} 中各随机变量是统计关联性的，可将信源分为:

（1）有记忆信源

$$F_{X_{t_1}\cdots X_{t_N}}(x_1,x_2,\cdots,x_N)=P(X_{t_1}\le x_1,X_{t_2}\le x_2,\cdots,X_{t_N}\le x_N)$$

（2）无记忆信源 $F_{X_{t_1}\cdots X_{t_N}}(x_1,x_2,\cdots,x_N)=\prod_{i=1}^{N}F_{X_{t_i}}(x_i)$

（3）平稳信源

序列的统计特性与时间的推移无关

$$F_{X_{t_1+\tau}\cdots X_{t_N+\tau}}(x_1,x_2,\cdots,x_N)=F_{X_{t_1}\cdots X_{t_N}}(x_1,x_2,\cdots,x_N)$$

（4）非平稳信源

离散单符号无记忆信源的数学模型

用概率空间（信源空间表示）：

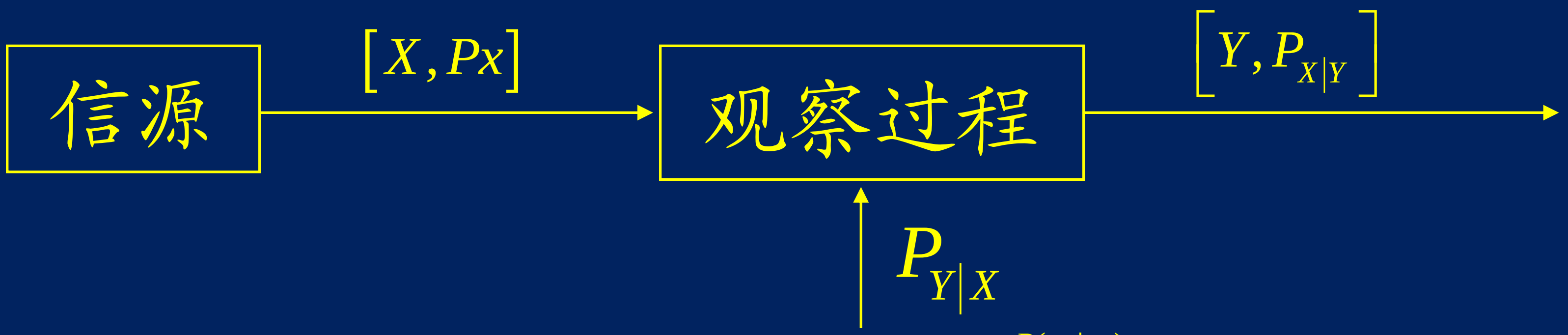
$$\begin{bmatrix} X \\ P_X \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ P(x_1) & P(x_2) & \cdots & P(x_K) \end{bmatrix}$$

若满足约束条件

$$\sum_{k=1}^{K} p(x_k) = 1$$

称为**完备信源**。

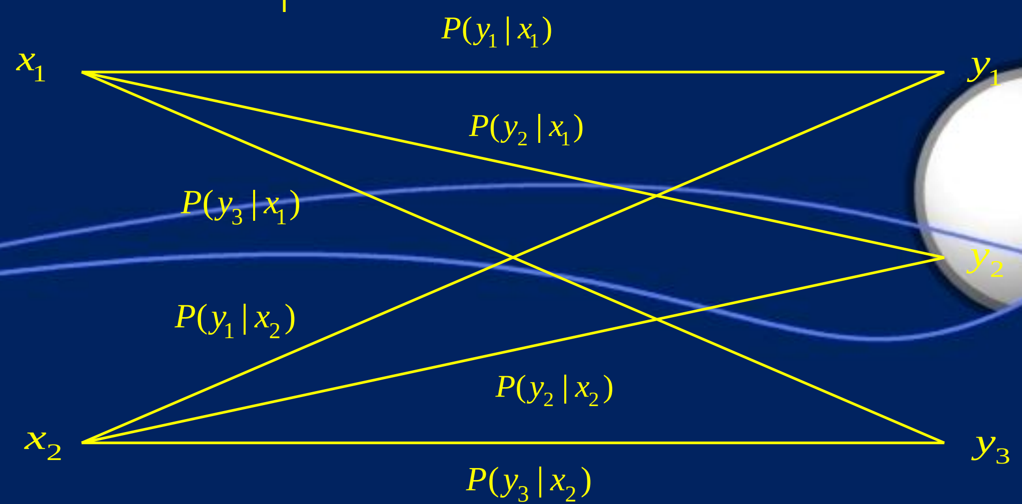
非理想观察模型（信息传递系统的抽象）



P_X ——先验概率

$P_{X|Y}$ ——后验概率

$P_{Y|X}$ ——转移概率



传递的信息 = 先验不确定性 − 后验不确定性

2、信息量及其性质

（1）自信息量及其性质

自信息量

$$I(x_k)=\log\frac{1}{P(x_k)}=-\log P(x_k)\quad k=1，2，\cdots,K$$

表示了信源符号的先验不确定性

单位：bit/符号　　取2为对数底

联合自信息量 $I(x_k,y_j)=-\log P(x_k,y_j)$

表示二元符号的先验不确定性。

单位：bit/二元符号

条件自信息量

$$I(x_k \mid y_j) = -\log P(x_k \mid y_j)$$

表示观察到 y_j 符号的条件下 x_k 还剩下的不确定性。

单位：bit/符号

$$I(y_j \mid x_k) = -\log P(y_j \mid x_k)$$

表示输入 x_k 且观察到 y_j 时干扰引入的不确定性。

单位：bit/符号

例　　甲在一8×8的方格棋盘上随意放入一个棋子，在乙看来棋子落入的位置是不确定的。

（1）在乙看来，棋子落入某方格的不确定性为多少？

（2）若甲告知乙棋子落入方格的行号，这时，在乙看来棋子落入某方格的

不确定性为多少？

解　将棋子方格从第一行开始按顺序编号，得到一个序号集合$\{z_l \mid l = 1, 2, \cdots, 64\}$，棋子落入的方格位置可以用取值于序号集合的随机变量 Z 来描述，即 $Z = \{z_l \mid l = 1, 2, \cdots, 64\}$

（1）由于棋子落入任一方格都是等可能的，则

$$P(z_l) = \frac{1}{64} \qquad l = 1, 2, \cdots, 64$$

棋子落入某方格的不确定性就是自信息量

$$I(z_l) = -\log P(z_l) = -\log \frac{1}{64} = 6 \quad \text{bit/符号}$$

（2）棋盘方格可分为8行×8列，已知行号$x_k(k=1,2,\cdots,8)$后，棋子落入某方格的不确定性就是条件自信息量 $I(z_l|x_k)$ 与条件概率 $P(z_l|x_k)$有关。由于

$$P(z_l \mid x_k)=\frac{1}{8} \qquad l=1,2,\cdots,64;k=1,2,\cdots,8$$

故

$$I(z_l \mid x_k)=-\log P(z_l \mid x_k|)=-\log\frac{1}{8}=3$$

bit/符号

由此可以看出，已知行号后，棋子位置的不确定性减小了一半，这与我们的常识是相符的。

三种自信息量的关系

全概率公式联合概率

$$P(x_k,y_j)=P(x_k)P(y_j|x_k)=P(y_j)P(x_k|y_j)$$

可导出

$$I(x_k,y_j)=I(x_k)+I(y_j|x_k)$$

$$=I(y_j)+I(x_k|y_j)$$

物理意义:

如果 (x_k,y_j)是某信源发出的符号串，x_k是观察输入符号，y_j是输出符号，则串 (x_k,y_j)的不确定性为 $I(x_k,y_j)$，输入符号 x_k的不确定性 $I(x_i)$加上此时干扰引入的不确定性 $I(y_j|x_k)$ 或联合符号 (x_k,y_j)的不确定性 $I(x_k,y_j)$，等于输出符号 y_j的不确定性 $I(y_j)$ 加上观察到 y_j后 x_k 还剩余的不确定性 (x_k,y_j) 。

当 x_k 和 y_j 统计独立时，

$$P(x_k, y_j) = P(x_k)P(y_j)$$

$$I(x_k, y_j) = I(x_k) + I(y_j)$$

$$I(x_k \mid y_j) = I(x_k)$$

$$I(y_j \mid x_k) = I(y_j)$$

推广到多维空间中去：

$$I(u_1, u_2, \cdots, u_N) = I(u_1) + I(u_2 \mid u_1) + I(u_3 \mid u_1 u_2) + \cdots + I(u_N \mid u_1 u_2 \cdots u_N)$$

对于无记忆信源

$$I(u_1, u_2, \cdots, u_N) = I(u_1) + I(u_2) + I(u_3) + \cdots + I(u_N)$$

（2）互信息量及其性质

输入 x_k ，$k=1，2，\cdots,K$ ，

输出 y_j ，$j=1，2，\cdots,J$。

从观察结果y_j中得到的有关输入符号x_k的信息称为互信息，记该信息为 $I(x_k;y_j)$。

$$I(x_k;y_j)=I(x_k)-I(x_k\mid y_j)$$

$I(x_k)$为先验不确定性，$I(x_k\mid y_j)$为后验不确定性

概率计算式：

$$\begin{aligned}I(x_k;y_j)&=[-\log P(x_k)]-[-\log P(x_k\mid y_j)]\\&=\log\frac{P(x_k\mid y_j)}{P(x_k)}\\&=\log\frac{P(x_k,y_j)}{P(x_k)P(y_j)}\end{aligned}$$

例 甲在一8×8的方格棋盘上随意放入一个棋子，在乙看来棋子落入的位置是不确定的。

（1）若甲告知乙棋子落入方格的行号，这时乙得到了多少信息量?

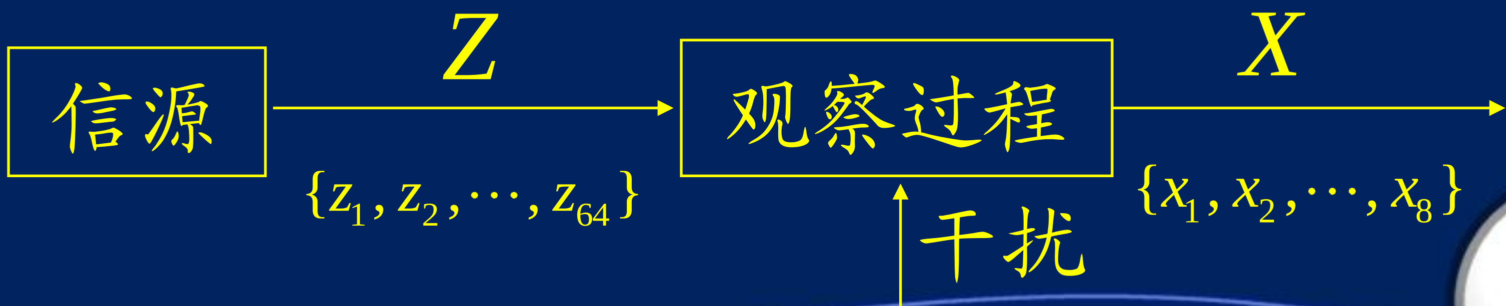
（2）若甲将棋子落入方格的行号和列号都告知乙，这时乙得到了多少信息量?

解 （1）采用与前例相同的符号约定，棋子落入方格的顺序号用随机变量 Z 表示，棋子落入方格的行号用随机变量 X，即

$$Z=\{z_l \mid l=1,2,\cdots,64\}$$

$$X=\{x_k \mid k=1,2,\cdots,8\}$$

这时，信息传输的模型如下图所示。



乙从甲告知的方格行号 x_k 中获得的关于方格顺序号 z_l 的信息就是互信息量 $I(z_l;x_k)$ 按定义, $I(z_l;x_k)$ 应等于获知行号 x_k 前后 z_l 的不确定性的减少量，由前例

$$I(z_l;x_k)=I(z_l)-I(z_l \mid x_k)=6-3=3\text{bit/符号}$$

（2）若用随机变量Y表示棋子落入方格的列号，即 $Y=\{y_j \mid j=1,2,\cdots,8\}$。当甲将棋子落入方格的行号和列号都告知乙时，乙就可以完全确定棋子落入方格的顺序号了。

这时，z_l 后验概率为

$$P(z_l \mid x_k y_j)=1$$

后验不确定性为

$$I(z_l \mid x_k y_j)=-\log P(z_l \mid x_k y_j)=0$$

因此互信息量为

$$I(z_l;x_k y_j)=I(z_l)-I(z_l \mid x_k y_j)=6-0=6\ \text{bit/符号}$$

互信息量的性质：

(1) 互易性； $I(x_k;y_j)=I(y_j;x_k)$

(2) 独立变量的互信息量为0；

x_k和 y_j统计独立，$I(x_k;y_j)=I(y_j;x_k)=0$

(3) 互信息量可正可负；

(4) 互信息量不可能大于符号的自信息。

$$I(x_k;y_j)=I(y_j;x_k)\le\begin{cases}I(x_k)\\I(y_j)\end{cases}$$

条件互信息量

三元联合概率空间

$$[XYZ,P_{XYZ}]=[(x_k,y_j,z_l),P(x_k,y_j,z_l)\,|\,k\in I_X,j\in I_Y,l\in I_Z]$$

在z_l出现的条件之下，x_k与y_j之间的互信息量为

$$I(x_k;y_j\,|\,z_l)=I(x_k\,|\,z_l)-I(x_k\,|\,y_jz_l)$$

$$I(x_k;y_j\,|\,z_l)=-\log P(x_k\,|\,z_l)+\log P(x_k\,|\,y_jz_l)$$

$$=\log\frac{P(x_k\,|\,y_jz_l)}{P(x_k\,|\,z_l)}$$

$$=\log\frac{P[(x_k,y_j)\,|\,z_l]}{P(x_k\,|\,z_l)P(y_j\,|\,z_l)}$$

x_k 与 (y_j, z_l) 之间的互信息量为

$$
\begin{aligned}
I(x_k; y_j z_l) &= \log \frac{P(x_k \mid y_j z_l)}{P(x_k)} \\
&= \log \frac{P(x_k \mid y_j z_l) P(x_k \mid y_j)}{P(x_k) P(x_k \mid y_j)} \\
&= \log \frac{P(x_k \mid y_j)}{P(x_k)} + \log \frac{P(x_k \mid y_j z_l)}{P(x_k \mid y_j)} \\
&= I(x_k; y_j) + I(x_k; z_l \mid y_j)
\end{aligned}
$$

上式这一性质称为**可加性**。

3、离散信源的熵及性质

(1) 离散熵

DMS: $[X, P_X] = [x_k\ , p_k \mid k = 1\ ,\ 2\ ,\cdots, K]$

自信息量 $I(x_k)$ 表示信源符号的不确定性。

信源X的不确定性是K个 $I(x_k)$ 的统计平均值，称为熵:

$$H(X) = H(p_1, p_2, \cdots, p_K) = \sum_{k=1}^{K} p_k I(x_k) = -\sum_{k} p_k \log p_k$$

熵中可忽略零概率事件

(2) 性质

对称性：$H(p_1, p_2, \cdots, p_K) = H(p_{m(1)}, p_{m(2)}, \cdots, p_{m(K)})$

可扩展性：

$$H(p_1, p_2, \cdots, p_K) = H(0\ , p_1, p_2, \cdots, p_K) = \cdots = H(p_1, \cdots, p_i, 0, p_{i+1}\ , \cdots, p_K)$$

$$= H(p_1, p_2, \cdots, p_K\ , 0)\quad i = 1\ ,\ 2\ , \cdots, K-1$$

非负性：$H(p_1, p_2, \cdots, p_K) = H(P) \geq 0$

强可加性：

$$H(p_1 q_{11}, \cdots, p_1 q_{1J}, \cdots, p_K q_{K1} \cdots, p_K q_{KJ})$$

$$= H(p_1, p_2, \cdots, p_K) + \sum_{k=1}^{K} p_k H(q_{k1}, \cdots, q_{kJ})$$

其中：$P_k=P\left(x_k\right)$ $P_X=\left\{P_1,P_2,\cdots,P_k\right\}$ $P_Y=\left\{q_1,q_2,\cdots,q_j\right\}$

$$X=\left\{x_1,x_2,\cdots,x_k\right\}\ \ Y=\left\{y_1,y_2,\cdots,y_j\right\}$$

本性质还可表示为

$$H\left(XY\right)=H(X)+H(Y|X)$$

可加性：当X，Y统计独立时，

$$H(p_1q_1,\cdots,p_1q_J,\cdots,p_Kq_1\cdots,p_Kq_J)$$

$$=H(p_1,p_2,\cdots,p_K)+H(q_1,q_2\cdots,q_J)$$

由强可加性，考虑到 $q_{1j}=q_{2j}=\cdots=q_{Kj}=q_j$ ， $j=1,\cdots,J$

即得到可加性的结果，也可表示为：

$$H\left(XY\right)=H(X)+H(Y|X)$$

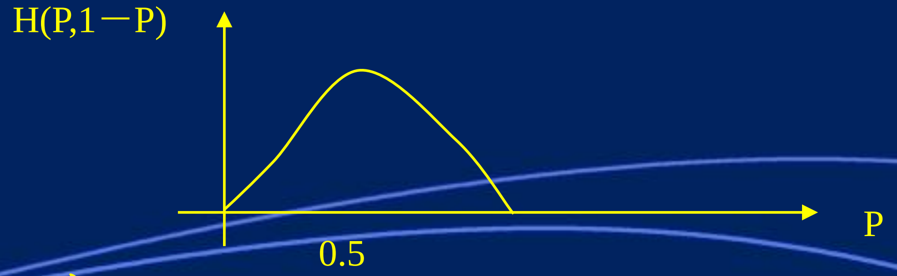
递增性:

$$H(p_1, p_2, p_3 \cdots, p_K)$$

$$= H(p_1 + p_2, p_3 \cdots, p_K) + (p_1 + p_2) H\left(\frac{p_1}{p_1 + p_2}, \frac{p_2}{p_1 + p_2}\right)$$

凸状性: $H(p_1, p_2, \cdots, p_K)$ 是上凸函数。

二元信源H（P，1－P）



凸函数有极大值。

极值性:

$$H(p_1, p_2, \cdots, p_K) \le H\left(\frac{1}{K}, \cdots, \frac{1}{K}\right) = \log K$$

两个重要不等式：

信息论不等式　　　$\ln Z \leqslant Z-1$

其中Z为任意大于0的实数，Z＝1时，等号成立。

香农不等式

$$-\sum_{k=1}^{K} P_k \log P_k \le -\sum_{k=1}^{K} P_k \log q_k$$

$$\sum_{k=1}^{K} P_k = 1,\ \sum_{k=1}^{K} q_k = 1$$

4、联合熵和条件熵

对联合自信息量和条件自信息量进行统计平均，可分别得出联合熵和条件熵。

(1) 联合熵

两个随机变量：

$$H(XY)=\sum_{i,j}P(x_k,y_j)I(x_k,y_j)=\sum_{i,j}P(x_k,y_j)\log P(x_k,y_j)$$

多个随机变量：

$$H(X_1X_2\cdots X_N)=-\sum_{i_1,i_2,\cdots,i_N}P(x_{i_1},x_{i_2},\cdots,x_{i_N})\log P(x_{i_1},x_{i_2},\cdots,x_{i_N})$$

(2) 条件熵

$$H(X|Y) = \sum_k \sum_j P(x_k, y_j) I(x_k|y_j) = \sum_{i,j} P(x_k, y_j) \log P(x_k|y_j)$$

$$H(X|Y) = -\sum_{k,j} P(x_k, y_j) \log P(x_k|y_j)$$

$$= -\sum_k \sum_j P(y_j) P(x_k|y_j) \log P(x_k|y_j)$$

$$= \sum_j P(y_j) H(X|Y = y_j)$$

式中

$$H(X|Y = y_j) = -\sum_{k=1}^{K} P(x_k|y_j) \log P(x_k|y_j)$$

(3) 各类熵之间的关系

前面已有:

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

当与统计独立时,

$$\boldsymbol{H(XY)=H(X)+H(Y)}$$

推广到多变量

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

5、平均互信息量及其性质

随机变量X与Y之间的平均互信息量是统计平均意义下的先验不确定性与后验不确定性之差，也是互信息量的统计平均。

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X\mid Y) \\
&= -\sum_{k} P(x_k)\log P(x_k) + \sum_{k,j} P(x_k, y_j)\log P(x_k \mid y_j) \\
&= -\sum_{k,j} P(x_k, y_j)\log P(x_k) + \sum_{k,j} P(x_k, y_j)\log P(x_k \mid y_j) \\
&= \sum_{k,j} P(x_k, y_j)\log\left[\frac{P(x_k \mid y_j)}{P(x_k)}\right] \\
&= \sum_{k,j} P(x_k, y_j) I(x_k; y_j)
\end{aligned}
$$

性质：

互易性，即 $I(X;Y)=I(Y;X)$ 。

非负性，即 $I(X;Y)\geq 0$。

由 $I(X;Y)=H(X)-H(X|Y)\geq 0$ 得

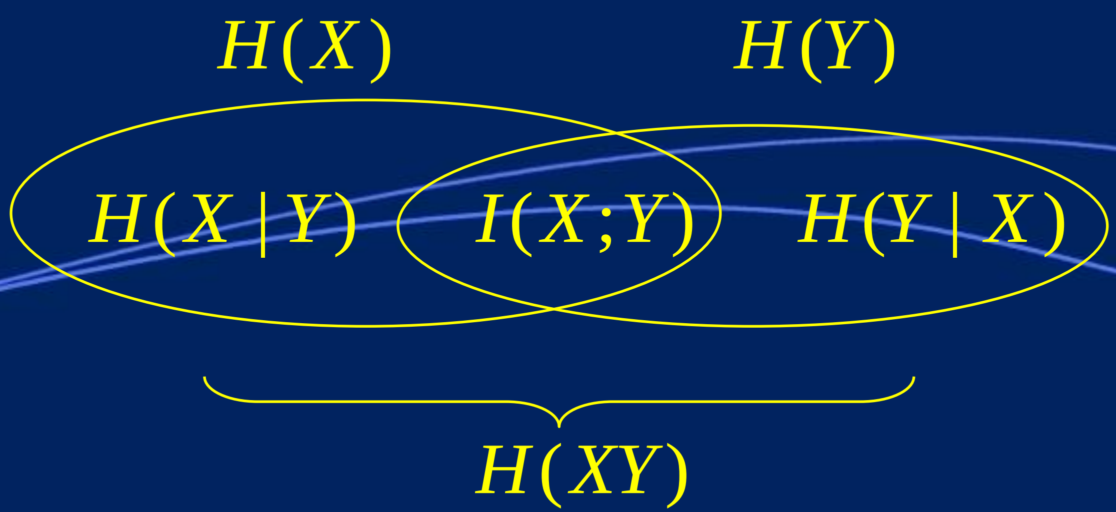
$H(X|Y)\leq H(X)$即条件熵不大于无条件熵。

推广

$$H(X|Y_1Y_2\cdots Y_N)\leq H(X|Y_1Y_2\cdots Y_{N-1})\leq\cdots\leq H(X)$$

有界性:

$$I(X;Y)=I(Y;X)\le\begin{cases}H(X)=I(X)\\H(Y)=I(Y)\end{cases}$$



6、扩展信源

N次扩展信源：

考虑离散无记忆信源X，将X的任意N个相邻时刻的输出随机变量 $X^N = X_1 X_2 \cdots X_N$ 看成是一个新的信源，称为X的N次扩展信源。

因为 X_n，$n = 1, 2, \cdots, N$，独立同分布

$$H(X^N) = H(X_1) + H(X_2) + \cdots + H(X_N) = NH(X)$$

Bit/N元符号

7、离散有记忆信源的熵

离散有记忆信源X，N阶平稳的，不确定性的描述方法有：

联合熵：$H(X^N) = H(X_1 X_2 \cdots X_N)$

单位：bit/N长符号串

条件熵：$H(X_N | X_1 X_2 \cdots X_{N-1}) = H(X^N | X^{N-1})$

单位：bit/符号

平均符号熵：

$$H_N(X) = \frac{1}{N} H(X^N) = \frac{1}{N} H(X_1 X_2 \cdots X_N)$$

单位：bit/符号

对于一般的离散有记忆信源，极限熵定义为：

$$H_\infty(X)=\lim_{N\to\infty}H_N(X)=\lim_{N\to\infty}\frac{1}{N}H(X_1X_2\cdots X_N)$$ bit/符号

可以证明，同样 $H_\infty(X)=\lim\limits_{N\to\infty}H(X_N|X^{N-1})$

规定 $H_0(X)$ 表示最大熵（等概）

$H_1(X)$ 表示不考虑统计关联的单符号熵

有如下关系：

$$0\le H_N(X)\le H_{N-1}(X)\le\cdots\le H_1(X)\le H_0(X)<\infty$$

8、马尔可夫信源的信息熵

1） 马尔可夫链

随机过程$\{X_n, n \in T\}$ X1X2X3......Xi的值取自同一符号集{x1， x2， ...， xq}

若记忆长度为n，即用条件概率表示随机变量之间的关系

$$P(x_{kn} | x_{k_1} \ x_{k_2} \cdots x_{k_{n-1}}) \text{，} k_1, k_2, \cdots, k_n = 1, 2, \cdots, q$$

记S为组成的状态空间集 $x_{k_1} x_{k_2} \cdots x_{k_{n-1}}$

若对所有正整数$\in T$ ，条件概率均满足

$$P\{X_n = S_{i_n} | X_{n-1} = S_{i_{n-1}}, X_{n-2} = S_{i_{n-2}}, \cdots X_1 = S_{i_1}\}$$

$$= P\{X_n = S_{i_n} | X_{n-1} = S_{i_{n-1}}\}$$

则此随机过程就是一个马尔可夫链。

马尔科夫链的初始分布概率记为$\{p_i, i \in S\}$

$p_i = p\{X_0 = i\} \geq 0$ （$i \in S$ ）且满足，$\sum_{i \in S} p_i = 1$

马尔科夫链的k步转移概率：

$$P_{ij}^{(k)}(m) = P\left\{S_{m+k} = j \middle| S_m = i\right\}$$

齐次马尔科夫链：

$p_{ij}^{(k)}(m) = p_{ij}^k$ 表示此转移概率与m无关。

遍历的马尔科夫链：

对于齐次马尔科夫链 $\{X_n, n \in T\}$ ，对一切i，j均存在不依赖于i的极限 $\lim_{n \to \infty} p_{ij}^{(n)} = p_j \geq 0$

且满足 $p_j = \sum_{i=0}^{\infty} p_i p_{ij}$， $\sum_j p_j = 1$ 即遍历。

马尔可夫链的有限维分布:

$$P(S_{t_1}=i_1,\cdots S_{t_n}=i_n)$$

$$=\sum_{i\in s} p_i P(S_{t_1}=i_1\,|\,S_0=i,)P(S_{t_2}=i_2\,|\,P(S_{t_1}=i_1)\cdots P(S_{t_n}=i_n\,|\,S_{t_{n-1}}=i_{n-1})$$

从该式可知马尔科夫链的有限维分布函数由初始分布和一步转移概率决定。

齐次马尔科夫链有C-K方程:

$$p_{ij}^{(m+r)}=\sum_{k\in S} p_{ik}^{(m)}p_{kj}^{(r)}$$

其中m、r、k为任意正整数。

C-K方程的证明：

$$p_{ij}^{(m+r)} = p\{X_{n+m+r} = S_j \mid X_n = S_i\}$$

$$= \frac{p\{X_{n+m+r} = S_j, X_n = S_i\}}{p\{X_n = S_i\}}$$

$$= \sum_{k\in S} \frac{p\{X_{n+m+r} = S_j, X_{n+m} = S_k, X_n = S_i\}}{p\{X_{n+m} = S_k, X_n = S_i\}} \bullet \frac{p\{X_{n+m} = S_k, X_n = S_i\}}{p\{X_n = S_i\}}$$

$$= \sum_{k\in S} p\{X_{n+m+r} = S_j \mid X_{n+m} = S_k, X_n = S_i\} \bullet p\{X_{n+m} = S_k \mid X_n = S_i\}$$

根据马尔可夫链的特性以及齐次性，可得

$$= \sum_{k\in S} p\{X_{n+m+r} = S_j \mid X_{n+m} = S_k\} \bullet p\{X_{n+m} = S_k \mid X_n = S_i\}$$

$$= \sum_{k\in S} p_{ik}^{(m)} p_{kj}^{(r)}$$

2) 马尔可夫信源

设信源所处的状态序列为

$$u_1, u_2, \cdots, u_i, \cdots \quad \in \{S_1, S_2, \cdots, S_J\}$$

在每一个状态下可能输出的符号序列为

$$x_1, x_2, \cdots, x_i, \cdots \quad \in \{a_1, a_2, \cdots, a_q\}$$

马尔可夫信源满足的条件：

(1) 某一时刻信源符号的输出只与当时的信源状态有关，而与以前的状态无关，即

$$p(x_l = a_k \mid u_l = S_j, x_{l-1} = a_k, u_{l-1} = S_i, \cdots) = p(x_l = a_k \mid u_l = S_j)$$

其中，$a_k \in (a_1, a_2, \cdots, a_q)$,S_i ,S_j $\in (S_1, S_2, \cdots, S_J)$

当具有时齐性时，有

$$p(x_l = a_k \mid u_l = S_j) = p(a_k \mid S_j) \quad 以及 \quad \sum_{a_k \in A} p(a_k \mid S_j) = 1$$

(2) 信源状态只由当前输出符号和前一时刻信源状态唯一确定，即

$$p(x_l = S_i \mid x_l = a_k, u_{l-1} = S_j) = \begin{Bmatrix} 1 \\ 0 \end{Bmatrix}$$

马尔可夫信源可用状态图来表示。

m阶有记忆的马尔可夫信源:

信源空间:

$$\begin{bmatrix} a_1 & ,a_2 & \cdots & , & a_q \\ p(a_{\mathrm{k}_{\mathrm{m}+1}} \mid a_{k_1} a_{k_2} \cdots a_{k_m}) \end{bmatrix}$$

状态空间:

$$\begin{bmatrix} S_1 & S_2 & \cdots & S_{q^m} \\ & p(S_j \mid S_i) & & \end{bmatrix}$$

3） 马尔可夫信源的信息熵

m阶马尔可夫信源，齐次、遍历，信源熵的极限值为：

$$H_\infty = H_\infty(X) = \lim_{N\to\infty}\frac{1}{N}H(X_1X_2\ \cdots\ X_{N\text{-}1}X_N)$$

$$= \lim_{N\to\infty}H(X_N \mid X_1X_2\ \cdots\ X_{N\text{-}1}) \quad = H(X_{m+1} \mid X_1X_2\ \cdots X_m) = H_{m+1}$$

证明：对于齐次、遍历的马尔可夫链，其状态可由唯一确定，可有

$$p(a_{k_{m+1}} \mid a_{k_m},\cdots,a_{k_1}) = p(a_{k_{m+1}} \mid S_j)$$

上式两端同时取对数，并对 $(a_{k1},a_{k2},\cdots,a_{km},a_{km+1})$ 和 S_j 求统计平均再取负，可得

$$左端 = -\sum_{k_1,k_2,\cdots k_{m+1};S_j} p(a_{k_{m+1}},\cdots,a_{k_1};S_j)\bullet\log p(a_{k_{m+1}}\mid a_{k_m},\cdots,a_{k_1})$$

$$= -\sum_{k_1,k_2,\cdots k_{m+1}} p(a_{k_{m+1}},\cdots,a_{k_1})\bullet\log p(a_{k_{m+1}}\mid a_{k_m},\cdots,a_{k_1})$$

$$= H(a_{k_{m+1}}\mid a_{k_m},\cdots,a_{k_1}) = H_{m+1}$$

$$右端 = -\sum_{k_1,k_2,\cdots k_{m+1};S_j} p(a_{k_{m+1}},\cdots,a_{k_1};S_j)\bullet\log p(a_{k_{m+1}}\mid S_j)$$

$$= -\sum_{k_1,k_2,\cdots k_{m+1};S_j} p(a_{k_m},\cdots,a_{k_1};S_j)p(a_{k_{m+1}}\mid S_j)\bullet\log p(a_{k_{m+1}}\mid S_j)$$

$$= -\sum_{k_{m+1}}\sum_{S_j} p(S_j)p(a_{k_{m+1}}\mid S_j)\bullet\log p(a_{k_{m+1}}\mid S_j)$$

$$= \sum_{S_j} p(S_j)H(X\mid S_j)$$

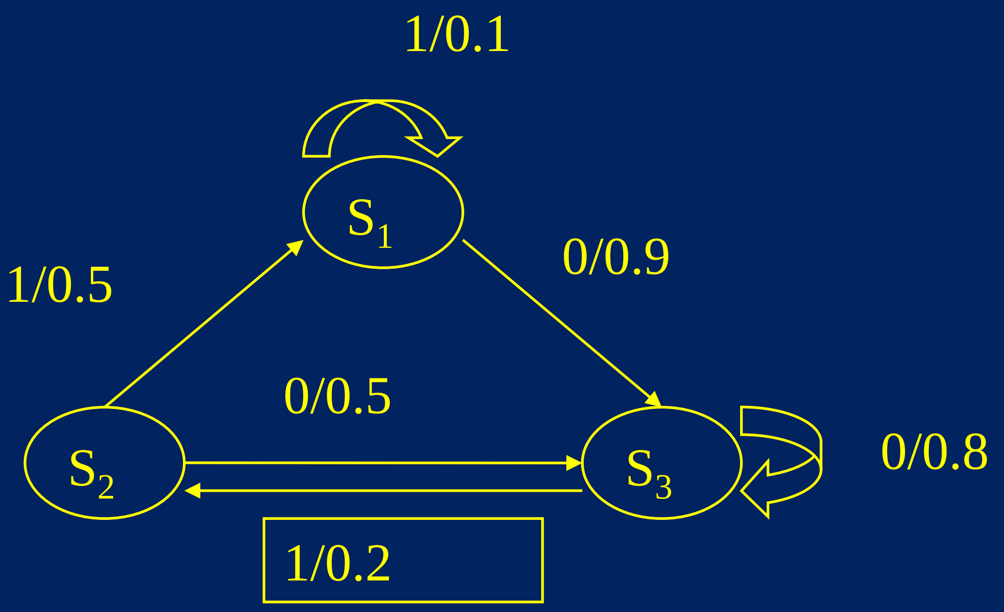
所以有

$$H_{m+1}=\sum_{S_j} p(S_j)H(X|S_j)=-\sum_{k_{m+1}}\sum_{S_j} p(S_j)p(a_{k_{m+1}}|S_j)\bullet\log p(a_{k_{m+1}}|S_j)$$

且

$$p(S_j)=\sum_{S_j\in S} p(S_j)P(S_i|S_j),\qquad \sum_{S_i\in S} p(S_i)=1$$

例：如下页图所示为三状态马尔可夫信源，计算信息熵



解：转移概率矩阵为

$$P = \begin{array}{c} S_1 \\ S_2 \\ S_3 \end{array} \begin{bmatrix} 0.1 & 0 & 0.9 \\ 0.5 & 0 & 0.5 \\ 0 & 0.2 & 0.8 \end{bmatrix}$$

(1) 求平稳分布P(S₁)、P(S₂)、P(S₃)

$$\begin{cases} P(S_1)=0.1P(S_1)+0.5P(S_2) \\ P(S_2)=\quad 0.2P(S_3) \\ P(S_3)=0.9P(S_1)+0.5P(S_2)+0.8P(S_3) \\ P(S_1)+P(S_2)+P(S_3)=1 \end{cases}$$

解得: $\begin{cases} P(S_1)=\dfrac{5}{59} \\ P(S_2)=\dfrac{9}{59} \\ P(S_3)=\dfrac{45}{59} \end{cases}$

(2) 计算S_j状态下每输出一个符号的平均信息量

$$H(X|S_1)=0.1\log\frac{1}{0.1}+0.9\log\frac{1}{0.9}=0.468996 \text{ 比特/符号}$$

$$H(X|S_2)=0.5\log\frac{1}{0.5}+0.5\log\frac{1}{0.5}=1\text{比特/符号}$$

$$H(X|S_3)=0.2\log\frac{1}{0.2}+0.8\log\frac{1}{0.8}=0.721928\text{比特/符号}$$

(3) 求信息熵

$$H_\infty = \sum_{j=1}^{3} p(S_j) H(X \mid S_j)$$

$$= 0.742910 \quad \text{比特/符号}$$

9、离散信源的信息（速）率和信息含量效率

信息率：平均一个符号所携带的信息量，记为R。反映了信息发出信息的能力

$$R = I(X) = H_\infty(X) \text{ bit/符号}$$

信息速率：单位时间内发出的平均信息量，记为 R_t

$$R_t = \frac{R}{t_s} = \frac{H_\infty(X)}{t_s} \text{ bit/秒}$$

t_s ——发信号的周期

信息含量效率：实际的实在信息与最大的实在信息之比，记为 η

$$\eta = \frac{I(X)}{I_{\max}(X)} = \frac{H_{\infty}(X)}{H_{\max}(X)} \qquad 0 \le \eta \le 1$$

当且仅当x是*DMS*且等概率分布（$P_X = P_0$）时，$\eta = 1$。

冗余度：

$$\gamma = 1 - \eta = 1 - \frac{H_{\infty}(X)}{H_{\max}(X)} = \frac{H_{\max}(X) - H_{\infty}(X)}{H_{\max}(X)}$$

10、连续随机变量下的熵和平均互信息量

(1) 连续随机变量下的熵

X: 连续随机变量，在区间$[a,b]$内取值，概率密度函数 $f_X(x)$的数学模型为:

$$[X,P_X]=[x,f_X(x)\mid x\in[a,b]]$$

$$f_X(x)\geq 0\ ,\ \int_a^b f_X(x)dx=1$$

将连续问题离散化:

将X的值域$[a,b]$等分为K个子区间:

$$\Delta_1,\Delta_2,\cdots,\Delta_K\quad,\quad \Delta_k=\Delta=\frac{b-a}{K}\ ,\ k=1,2,\cdots,K$$

第k个子区间内的概率 p_k 为

$$p_k = \int_{a+(k-1)\Delta}^{a+k\Delta} f_X(x)dx = f_X(x_k)\Delta \quad ，a+(k-1)\Delta \le x_k \le a+k\Delta$$

这样得到一个离散随机变量 X_Δ，其概率空间为

$$\begin{bmatrix} X_\Delta \\ P_{X_\Delta} \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ f_X(x_1)\Delta & f_X(x_2)\Delta & \cdots & f_X(x_K)\Delta \end{bmatrix}$$

且概率空间是完备的:

$$\sum_{k=1}^{K} f_X(x_k)\Delta = \sum_{k=1}^{K}\left[\int_{a+(k-1)\Delta}^{a+k\Delta} f_X(x)dx\right] = \int_a^b f_X(x)dx = 1$$

信源是完备的。

根据离散熵公式，有

$$H(X_\Delta) = -\sum_{k=1}^{K} [f_X(x_k)\Delta] \log [f_X(x_k)\Delta]$$

$$= -\left[\sum_{k=1}^{K} f_X(x_k) \log f_X(x_k)\right] \cdot \Delta - (\log \Delta) \sum_{k=1}^{K} f_X(x_k) \cdot \Delta$$

$$= -\left[\sum_{k=1}^{K} f_X(x_k) \log f_X(x_k)\right] \cdot \Delta - \log \Delta$$

令 $K \to \infty$ ， $\Delta \to 0$ ， $H(X_\Delta) \longrightarrow H(X)$

$$H(X) = \lim_{\substack{K \to \infty \\ (\Delta \to 0)}} H(X_\Delta) = -\int_a^b f_X(x) \log f_X(x) dx - \lim_{\substack{K \to \infty \\ (\Delta \to 0)}} \log \Delta$$

$$= -\int_a^b f_X(x) \log f_X(x) dx + \infty$$

因此，连续熵为无穷大。

连续随机变量只能定义相对熵，称为微分熵，是上式的第一项。记为：

$$h(X) = -\int_a^b f_X(x)\log f_X(x)dx$$

一般的定义式为

$$h(X) = -\int_{-\infty}^{\infty} f_X(x)\log f_X(x)dx = -\int_R f_X(x)\log f_X(x)dx$$

微分熵可用于比较两个连续随机变量不确定性的相对大小。

2） 连续随机变量下的联合熵、条件熵以及平均互信息量

连续随机变量X和Y：

联合微分熵 $h(XY) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log f_{XY}(x,y)dxdy$

条件微分熵 $h(X|Y) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log f_{X|Y}(x|y)dxdy$

联合熵与条件熵的关系与离散熵相似

$$h(XY) = h(X) + h(Y|X) = h(Y) + h(X|Y)$$

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

$$h(XY) \le h(X) + h(Y)$$

等号成立的充要条件是X与Y统计独立。

平均互信息量 $I(X;Y)$:（类比离散情形）

$$I(X;Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log\frac{f_{XY}(x,y)}{f_X(x)f_Y(y)}dxdy$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\log\frac{f_{X|Y}(x|y)}{f_X(x)}dxdy$$

类似关系

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

$$I(X;Y)\geq 0$$

3） 微分熵的极大化问题

幅值受限（即取值区间受限，等价于峰值功率受限）

定理：设X的取值受限于有限区间 $[a,b]$ 则X服从均匀分布时，其熵达到最大。

此时，最大微分熵为

$$h(X)=-\int_a^b \frac{1}{b-a}\log\frac{1}{b-a}dx=\log(b-a)$$

方差受限（即平均功率受限）

X的均值为μ，方差为σ^2，

根据方差的定义$\int_{-\infty}^{\infty}(x-\mu)^2 f_X(x)dx=\sigma^2$

$$\int_{-\infty}^{\infty}(x-\mu)^2 f_X(x)dx$$

$$=\int_{-\infty}^{\infty}(x^2-2x\mu+\mu^2)f_X(x)dx$$

$$=\int_{-\infty}^{\infty}x^2 f_X(x)dx-\mu^2$$

$$\int_{-\infty}^{\infty}x^2 f_X(x)dx=\sigma^2+\mu^2=P$$

定理：X的均值为，方差受限为，则服从高斯分布时，其熵达到最大。

最大熵为 $h(X)=\log\sqrt{2\pi eP}$

4） 连续信源的熵功率

以奈特为单位时，最大熵为：

$$h(X)=\log\sqrt{2\pi eP}\ \rightarrow P=\frac{1}{2\pi e}e^{2h(X)}$$

若X的平均功率仍受限为P时，但不是高斯分布，则 $h(X)\le\ln\sqrt{2\pi eP}$（非高斯熵与功率的关系），即

$$P\ge\frac{1}{2\pi e}e^{2h(X)}$$

定义

$$\overline{P}=\frac{1}{2\pi e}e^{2h(X)},\qquad \overline{P}\le P$$

当且仅当X服从高斯分布时，等号成立，熵功率等于平均功率

连续信源的剩余度 $=P-\overline{P}$