1

熵和互信息量的基本概念

目录

- 熵函数
- 熵和互信息量的定义
- Jensen不等式
- 熵的计算

1.1 熵函数

- 熵函数的基本假设

熵函数：给定正整数 n,

设事件集合$\{E_1, E_2, \ldots, E_n\}$,

相应的概率分布为$\{p_1, p_2, \ldots, p_n\}$,

$$\sum_{i=1}^{n} p_i = 1$$

定义f为事件集的熵函数

1.1 熵函数

- 熵函数的基本假设(cont.)

满足:

(1) 当 $p_i = 1/n,\ i = 1, 2, \ldots, n$时,

$f(p_1, p_2, \ldots p_n)$ 取极大值

(2) 如果事件集中增加一个不可能事件 E_{n+1}（$p_{n+1} = 0$），

相应的熵不变.

即 $f(p_1, p_2, \ldots, p_n) = f(p_1, p_2, \ldots, p_n, 0)$

(3)

$$f(p_1, \cdots, p_j, \cdots, p_k, \cdots, p_n) = f(p_1, \cdots, p_j + p_k, \cdots, 0, \cdots, p_n) + (p_j + p_k) f\left(\frac{p_j}{p_j + p_k}, \frac{p_k}{p_j + p_k}\right)$$

1.1 熵函数

- 熵函数的唯一性定理

  满足上述基本假设的熵函数的唯一解是:

$$f(p_1, p_2, \cdots, p_n) = \sum_{i=1}^{n} p_i \log \frac{1}{p_i}$$

1.1 熵函数

- 熵函数的唯一性定理(cont.)

  说明：

  (1)对数运算的基底没有明确规定：

  以2为基底的单位称为bits(binary digits)，

  以e为基底的单位称为nats(natural digits)

  (2)如果$p_i=0$，式中$p_i \log p_i^{-1}$项的值是不确定的，但是我们规定它为0

  (3)如果事件集合是无限的，和式有可能不收敛，这种情况下，规定它的值为$+\infty$

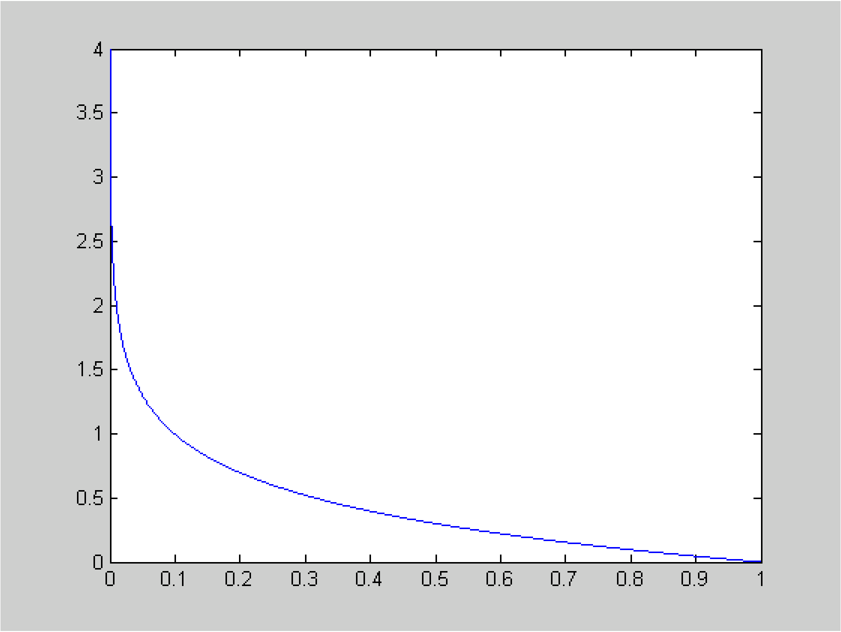
目录

- 熵函数
- 熵和互信息量的定义
- Jensen不等式
- 熵的计算

1.2 熵和互信息量的定义

- 事件的自信息量：

$$I(x) = \log \frac{1}{p(x)}$$



含义：由事件X = x提供的信息量

1.2 熵和互信息量的定义

- 熵(平均自信息量):

$$H(X)=\sum_{x}p(x)\log\frac{1}{p(x)}$$

含义：(1) 通过观测X所获得的“信息量”

(2) 关于X的“不确定性”的度量

(3) X的“随机性”的度量

(a) The amount of “**information**” provided by an observation of X
(b) Our “**uncertainty**” about X
(c) The “**randomness**” of X

英文的平均信息熵是4.03比特

中文的平均信息熵是9.65比特

1.2 熵和互信息量的定义

- 熵：

  例1.1: (a) 信源等概输出8种符号{0, 1, 2, 3, 4, 5, 6, 7}, 即 $p_i = 1/8$, i=0, 1, …7.

  $H(X) = \log 8 = 3\text{bits}$

  问题：平均需要用几比特表示信源输出?

1.2 熵和互信息量的定义

- 熵：

  例1.1: (b) 信源输出8种符号{0, 1, 2, 3, 4, 5, 6, 7}的概率分别为{1/2， 1/4， 1/8， 1/16， 1/64， 1/64，1/64， 1/64}.

$$H(X)=\frac{1}{2}\log 2+\frac{1}{4}\log 4+\frac{1}{8}\log 8+\frac{1}{16}\log 16+4\times\frac{1}{64}\log 64=2bits$$

问题：平均需要用几比特表示信源输出?

1.2 熵和互信息量的定义

- 熵：

例1.2: R= {0, 1}, P{x=0} = p, P{x=1} = 1-p

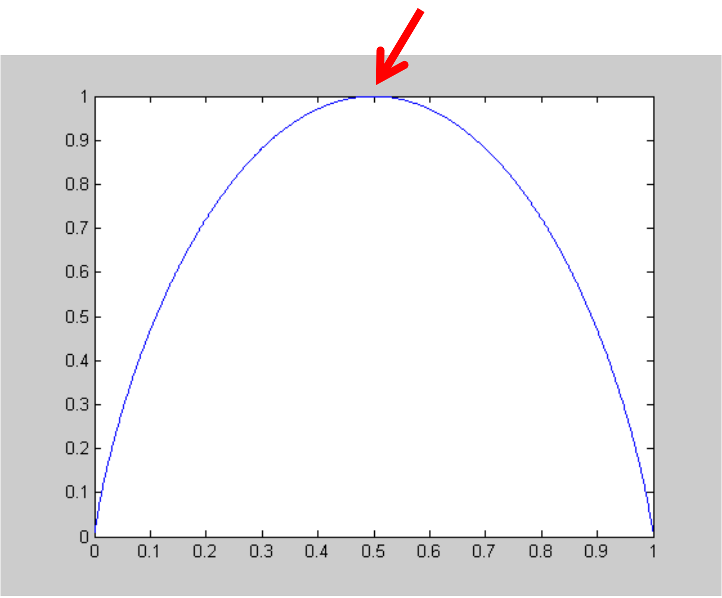
$$H(X) = -p\log p-(1-p)\log(1-p) = H(p)$$

注意区分:

H(X)

H(p)

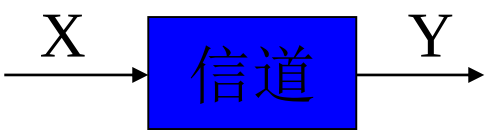
$$H(\mathbf{p}) = \sum_i p_i \log \frac{1}{p_i}$$



1.2 熵和互信息量的定义

- 条件熵：

$$H(X/Y)=\sum_{x,y}p(x,y)\log\frac{1}{p(x/y)}$$



物理含义：

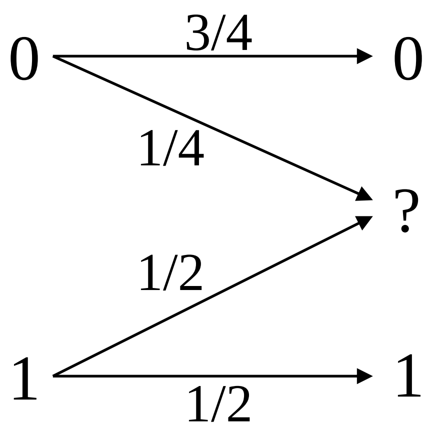
观测Y以后，仍保留的关于X的不确定量

1.2 熵和互信息量的定义

- 条件熵：

  例2.3: 二进制删除信道，

  输入概率分布$P_x(0)=2/3$, $P_x(1)=1/3$.



$$
\begin{aligned}
H_2(X) &= 0.918\text{bits} \\
H_2(X/Y=0) &= 0 \\
H_2(X/Y=1) &= 0 \\
H_2(X/Y=?) &= 1\text{bits} \\
H_2(X/Y) &= 0.333\text{bits}
\end{aligned}
$$

注：$H(X/Y=y_i)=\sum_x p(x/y=y_i)\log\dfrac{1}{p(x/y=y_i)}$

> ➢如果收到Y=0或Y=1，就确知发端X的取值；
> ➢如果收到Y=?对X的不确定性反而增加了；
> ➢H(X/Y)=0.333bits，平均来讲，观测Y后降低了对X的不确定性！

➢已知先验、转移概率

$$p_x(0)=2/3$$

$$p_x(1)=1/3$$

$$[p(y/x)]=\begin{bmatrix}3/4 & 1/4 & 0\\ 0 & 1/2 & 1/2\end{bmatrix}$$

➢得到联合概率分布

$$[p(x,y)]=\begin{bmatrix}1/2 & 1/6 & 0\\ 0 & 1/6 & 1/6\end{bmatrix}$$

➢得到概率分布

$$p_y(0)=\frac{1}{2},\ p_y(?)=\frac{1}{3},\ p_y(1)=\frac{1}{6}$$

➢得到后验概率分布

$$[p(x/y)]=\begin{bmatrix}1 & 1/2 & 0\\ 0 & 1/2 & 1\end{bmatrix}$$

➢得到

$$\begin{cases}H(X/Y=0)=1\log 1+0\log 0=0\\ H(X/Y=?)=\frac{1}{2}\log 2+\frac{1}{2}\log 2=1bit\\ H(X/Y=1)=1\log 1+0\log 0=0\end{cases}$$

➢得到

$$H(X/Y)=\sum_y p(y)H(X/Y=y)=0.333bits$$

1.2 熵和互信息量的定义

- （平均）互信息量

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

$$=\sum_{x}p(x)\log\frac{1}{p(x)}-\sum_{x,y}p(x,y)\log\frac{1}{p(x/y)}$$

$$=\sum_{x,y}p(x,y)\log\frac{1}{p(x)}-\sum_{x,y}p(x,y)\log\frac{1}{p(x/y)}$$

$$=\sum_{x,y}p(x,y)\log\frac{p(x/y)}{p(x)}$$

$$=\sum_{x,y}p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$

1.2 熵和互信息量的定义

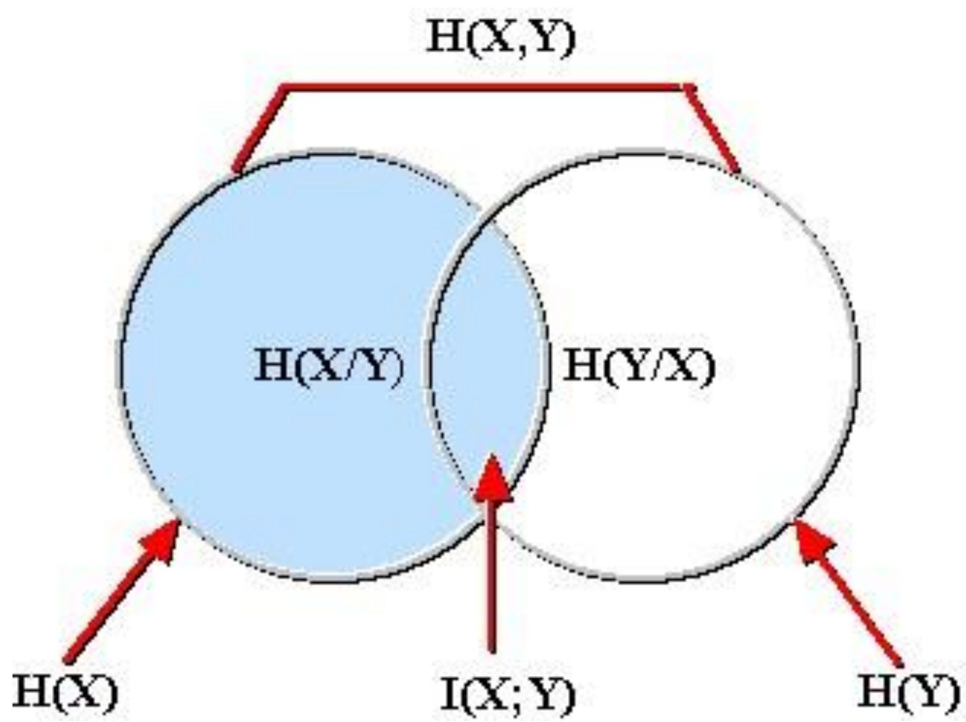
- 互信息量(cont.)

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

其中联合熵：$H(X,Y)=\sum_{x,y}p(x,y)\log\frac{1}{p(x,y)}$

1.2 熵和互信息量的定义

- 互信息量



Venn图

目录

- 熵函数
- 熵和互信息量的定义
- Jensen不等式
- 熵的计算

1.3 Jensen不等式

- 引理1.1: $\log x \le x - 1,$

  当且仅当$x = 1$时, 等号成立.

1.3 Jensen不等式

- 引理1.2: 如果X是在 $x>0$ 区间严格凸的，

  且 $E[X]$存在，

  则 $E[\log X] \le \log E[X]$

凸集：设α、$\beta \in R$，有$\theta\alpha+(1-\theta)\beta \in R$，

则R为凸集，其中$0\le\theta\le1$.

1.3 Jensen不等式

- Jensen不等式:

  如果f (x) 是X凸集上定义的上凸函数,

  且 E[X]存在,

  则E [f(X)] ≤ f (E[X])

上凸函数: $\theta f(\alpha)+(1-\theta)f(\beta) \le f[\theta\alpha+(1-\theta)\beta]$

1.3 Jensen不等式

- 例1.4: 证明互信息量 I(X ;Y) ≥ 0.
- 证明:

$$-I(X;Y) = -\sum_{x,y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$

$$= \sum_{x,y} p(x,y)\log\frac{p(x)p(y)}{p(x,y)}$$

$$\leq \log\sum_{x,y} p(x,y)\frac{p(x)p(y)}{p(x,y)}$$

$$= \log\sum_{x,y} p(x)p(y) = 0$$

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

1.3 Jensen不等式

- 例1.5: 证明 $H(X) \geq H(X/Y)$.

1.3 Jensen不等式

- 例1.6: 离散信源X共有r种取值，证明 $\log r \geq H(X) \geq 0$.

证明： 因为$0 \leq p_i \leq 1$，H(X)中的每一项$p_i \log p_i^{-1} \geq 0$，

因此 $H(X) \geq 0$

并且当且仅当p=0或1时，$p\log p^{-1}=0$

也就是当且仅当一个$p_i=1$，其余都为0， $H(X)=0$

现在根据Jensen不等式，因为logx是严格上凸的

$$H(X)=\sum_{i=1}^{r} p_i \log \frac{1}{p_i} \leq \log \sum_{i=1}^{r} p_i \frac{1}{p_i} = \log r$$

并且当且仅当对于所有i，$p_i=1/r$时，等式成立.

目录

- 熵函数
- 熵和互信息量的定义
- Jensen不等式
- 熵的计算

1.4 熵的计算

- 离散N维扩展信源的熵

  信源连续发送N个符号:

  $\mathbf{x} = (x_1, x_2, \ldots, x_N)$

  $p(\mathbf{x}) = p(x_1, x_2, \ldots, x_N)$

  $= p(x_1)\, p(x_N, \ldots, x_2/x_1)$

  $= p(x_1)\, p(x_2/x_1)\, p(x_3/x_2,x_1) \ldots p(x_N/x_{N-1},\ldots, x_1)$

1.4 熵的计算

- 离散N维扩展信源的熵(cont.)

熵 $H(\mathbf{X})=\sum_{\mathbf{x}} p(\mathbf{x})\log\frac{1}{p(\mathbf{x})}$

$$=H(X_1)+H(X_2/X_1)+\cdots+H(X_n/X_{n-1},\cdots X_1)$$
$$+\cdots+H(X_N/X_{N-1},\cdots X_1)$$

其中 $H(X_n/X_{n-1},\cdots,X_1)$

$$=\sum_{\mathbf{x}} p(x_n,x_{n-1},\cdots,x_1)\log\frac{1}{p(x_n/x_{n-1},\cdots,x_1)}$$

1.4 熵的计算

- 离散N维扩展信源的熵(cont.)

推论: 如果信源是离散的、无记忆的、平稳的，

则 $H(\mathbf{X}) = N\,H(X_1) = N\,H(X_N) = N\,H(X)$

1.4 熵的计算

- 例：某离散信源由0,1,2,3四种符号组成，其分布概率为

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ 3/8 & 1/4 & 1/4 & 1/8 \end{pmatrix}$$

求消息201 020 130 213 001 203 210 100 321 010 023 102 002 010 312 032 100 120 210的信息量.

1.4 熵的计算

- 解：此消息总长57个符号，其中0出现23次，1出现14次，2出现13次，3出现7次.

  此消息的信息量

$$I=-\sum_{i=0}^{3}n_i\log P(x_i)=-23\log_2\frac{3}{8}-14\log_2\frac{1}{4}-13\log_2\frac{1}{4}-7\log_2\frac{1}{8}$$

$$=32.55+28+26+21=108.55bit$$

$$H(X)=-\sum_{i=0}^{3}P(x_i)\log P(x_i)=-\frac{3}{8}\log_2\frac{3}{8}-\frac{1}{4}\log_2\frac{1}{4}-\frac{1}{4}\log_2\frac{1}{4}-\frac{1}{8}\log_2\frac{1}{8}$$

$$=0.53+0.5+0.5+0.375=1.91bit$$

$$1.91\times57=108.87bit$$

1.4 熵的计算

- 离散N维扩展Markov信源的熵

  初始状态 S_0

  连续发送N个符号 $X_1, X_2, \ldots, X_N$

  信源的熵：$$H(X_1, X_2, \cdots, X_N / S_0)$$

$$= \sum_{\mathbf{x}} p(\mathbf{x}/s_0) \log \frac{1}{p(\mathbf{x}/s_0)}$$

1.4 熵的计算

- 离散N维扩展Markov信源的熵(cont.)

其中
$$
\begin{aligned}
p(\mathbf{x}/s_0) &= p(x_N, x_{N-1}, \cdots, x_1/s_0) \\
&= p(x_N/x_{N-1}, \cdots, x_1, s_0)\,p(x_{N-1}/x_{N-2}, \cdots, x_1, s_0)\cdots \\
&\quad \cdots p(x_2/x_1, s_0)\,p(x_1/s_0) \\
&= \prod_{r=1}^{N} p(x_r/x_{r-1}, \cdots, x_1, s_0) \\
&= \prod_{r=1}^{N} p(x_r/s_{r-1})
\end{aligned}
$$

1.4 熵的计算

- 离散N维扩展Markov信源的熵

  对于 m阶Markov信源

$$p(\mathbf{x}/s_0) = p(x_N, x_{N-1}, \cdots, x_1/s_0)$$

$$= p(x_N/x_{N-1}, \cdots, x_{N-m})\,p(x_{N-1}/x_{N-2}, \cdots, x_{N-m-1})\cdots$$

$$\cdots p(x_2/x_1, s_0)\,p(x_1/s_0)$$

$$= \prod_{r=1}^{N} p(x_r/x_{r-1}, x_{r-2}, \cdots, x_{r-m})$$

1.4 熵的计算

- 离散N维扩展Markov信源的熵

$$H(X_1, X_2, \cdots, X_N / S_0) = \sum_{\mathbf{x}} p(\mathbf{x} / s_0) \log \frac{1}{\prod_{r=1}^{N} p(x_r / x_{r-1}, \cdots, x_{r-m})}$$

$$= \sum_{r=1}^{N} \left[\sum_{\mathbf{x}} p(\mathbf{x} / s_0) \log \frac{1}{p(x_r / x_{r-1}, \cdots, x_{r-m})} \right]$$

$$= \sum_{r=1}^{N} \left[\sum_{x_r, x_{r-1}, \cdots, x_{r-m}} p(x_r, x_{r-1}, \cdots, x_{r-m} / s_0) \log \frac{1}{p(x_r / x_{r-1}, \cdots, x_{r-m})} \right]$$

1.4 熵的计算

- 离散N维扩展Markov信源的熵

$$
=\sum_{r=1}^{N}\left[\begin{array}{l}\displaystyle\sum_{x_{r-1},x_{r-2},\cdots,x_{r-m}} p(x_{r-1},x_{r-2},\cdots,x_{r-m}/s_0)\times \\ \displaystyle\left(\sum_{x_r} p(x_r/x_{r-1},\cdots,x_{r-m})\log\frac{1}{p(x_r/x_{r-1},\cdots,x_{r-m})}\right)\end{array}\right]
$$

$$
=\sum_{r=1}^{N}\left[\sum_{s_{r-1}} p(s_{r-1}/s_0)H(X_r/S_{r-1})\right]
$$

1.4 熵的计算

- 离散N维扩展Markov信源的熵

当r→∞时，每个符号的熵：

$$H(X)=\sum_{s_{r-1}} p(s_{r-1}/s_0)H(X_r/S_{r-1})$$

$$=\sum_{s} p(s)H(X/S)$$

含义: (1) $p(s_{r-1}/s_0) \to p(s_{r-1})$，与初始状态$s_0$无关

(2) $s_{r-1} \to s$，达到平稳

1.4 熵的计算

- 离散N维扩展Markov信源的熵

例1.7: 一阶Markov信源发送三种符号: 0, 1, 2.

发送符号的概率仅依赖于上次所发送的符号.

转移概率 $$[p(x/s)] = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix}$$

1.4 熵的计算

- 离散N维扩展Markov信源的熵

例1.7: (1)求出平稳状态概率分布p(s)

(2)求出每个状态下的熵:

$$H(X/S)=\sum_{x} p(x/s)\log\frac{1}{p(x/s)}$$

(3)求出每个符号的熵:

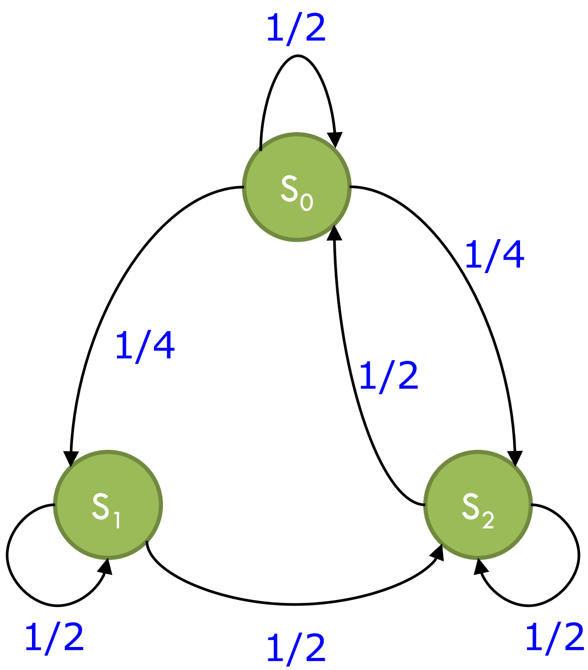
$$H(X)=\sum_{s} p(s)H(X/S)$$

提示: $(p_{n+1})=(p_n)\left[p(x/s)\right]$

平稳后 $p_n = p_{n+1} = p$

1.4 熵的计算

- 例1.7求解



由 $[P_0 \quad P_1 \quad P_2] = [P_0 \quad P_1 \quad P_2] \cdot \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{bmatrix}$

求得 $P_0 = 2/5,\ P_1 = 1/5,\ P_2 = 2/5$

$$\begin{cases} H(X/S=0) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{4}\log 4 = \frac{3}{2} bit \\ H(X/S=1) = \frac{1}{2}\log 2 + \frac{1}{2}\log 2 = 1 bit \\ H(X/S=2) = \frac{1}{2}\log 2 + \frac{1}{2}\log 2 = 1 bit \end{cases}$$

$$H(X) = \sum_s p(s)H(X/S) = \frac{6}{5} bit$$

1.4 熵的计算

- 离散N维扩展Markov信源的熵

例1.8: 假设信源Y的取值与上述Markov信源完全对应，它是离散无记忆的，

$$P_y(y) = P_x(x) = \sum_s p(s)p(x/s)$$

证明 $H(X) \le H(Y) \le \log r$

r为信源符号取值个数.

1.4 熵的计算

- 离散N维扩展Markov信源的熵

例1.8: $H(X)-H(Y)=\sum_{s}p(s)H(X/s)-\sum_{y}p(y)\log\frac{1}{p(y)}$

$$=\sum_{x,s}p(s)p(x/s)\log\frac{1}{p(x/s)}-\sum_{x,s}p(s)p(x/s)\log\frac{1}{\sum_{s'}p(s')p(x/s')}$$

$$=\sum_{x,s}p(s)p(x/s)\log\frac{\sum_{s'}p(s')p(x/s')}{p(x/s)}$$

$$\le\log\left(\sum_{x,s}p(s)\sum_{s'}p(s')p(x/s')\right)$$

$$=\log\left(\sum_{s}p(s)\sum_{s'}p(s')\right)=0$$

作业

第一章习题：1.1，1.3，1.8

其它题选作！