离散信源
及其信息测度

信息论与编码原理不挂科 第二讲

离散信源
及其信息测度

信息论与编码原理不挂科 第二讲

6大模块

12道题目

信息论与编码原理不挂科 第二讲

离散信源及其信息测度

信源的数学模型及分类

小节1 信源的定义及分类

小节2 信源的数学模型

信源的数学模型及分类

小节1 信源的定义及分类

小节2 信源的数学模型

信源的定义及基本分类

信源是产生**消息**（符号）、**消息序列**（符号序列）以及**时间连续的消息**的来源

按照信号取值集合与取值时刻集合的连续或离散分类

信号取值集合	信号取值时刻集合	信源种类
离散	离散	数字/离散
连续	离散	连续
连续	连续	模拟/波形

信源的数学模型及分类

小节1 信源的定义及分类

小节2 信源的数学模型

信源的数学模型

信源是产生**消息（符号）**、**消息序列（符号序列）**以及**时间连续的消息**的来源

随机变量　　随机矢量　　随机过程

1. 用随机变量描述信源输出的消息——输出为单符号

离散随机变量　输出消息有限、可数，且每次只输出一个消息。

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} X = x_1 & X = x_2 & \cdots & X = x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix}, \quad 0 \le p(x_i) \le 1, \quad \sum_{i=1}^{n} p(x_i) = 1$$

信源给定，概率空间就随之确定，因此可用概率空间表征离散信源的特性。

连续随机变量　输出消息为不可数的无限值，即是连续的。

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} (a,b) \\ p(x) \end{bmatrix}, \quad \int_a^b p(x)dx = 1$$

信源的数学模型

信源是产生消息（符号）、**消息序列**（符号序列）以及时间连续的消息的来源

随机变量　　随机矢量　　随机过程

2. 用随机矢量（序列）描述信源输出的消息——输出为一系列符号

例1 汉字一句20个字的话，每个字都为一个随机变量，整体可组成一个随机序列，构成不同的句子；

例2 电视屏幕每帧画面都由一系列像素点组成，每个像素点都有若干灰度等级，可组成一个随机序列，构成不同图像。

$\vec{X} = (X_1 X_2 \cdots X_N)$ →

- 离散无记忆信源的N次扩展信源
- 离散平稳信源
- 离散非平稳有记忆信源——马尔可夫信源

平稳性

记忆性

信源的数学模型

信源是产生消息（符号）、**消息序列**（符号序列）以及**时间连续的消息**的来源

随机变量　随机矢量　随机过程

2. 用随机矢量（序列）描述信源输出的消息——输出为一系列符号

离散平稳信源

信源输出的随机序列中每个随机变量 X_i 都是取值离散的离散型随机变量；且随机变量的各维概率分布都与时间起点无关。

连续平稳信源

信源输出的随机序列中每个随机变量 X_i 都是取值连续的连续型随机变量；且随机变量的各维概率密度都与时间起点无关。

信源的数学模型

信源是产生消息（符号）、**消息序列**（符号序列）以及时间连续的消息的来源

随机变量　　随机矢量　　随机过程

2. 用随机矢量（序列）描述信源输出的消息——输出为一系列符号

离散无记忆信源

信源先后发出的符号之间统计独立　　$P(\vec{X}) = P(X_1X_2\cdots X_N) = P(X_1)P(X_2)\cdots P(X_N)$

特例：离散无记忆信源的N次扩展信源　将信源输出的序列看作一组一组发出，每组长度为N。

例 二进制信源：输出符号“0”，“1”

单符号信源：	$\alpha_1 = 0,\ \alpha_2 = 1$	2^1
二次扩展信源：	$\alpha_1 = 00,\ \alpha_2 = 01,\ \alpha_3 = 10,\ \alpha_4 = 11$	2^2
三次扩展信源：	$\alpha_1 = 000, \alpha_2 = 001, \alpha_3 = 010, \alpha_4 = 011,$ $\alpha_5 = 100, \alpha_6 = 101, \alpha_7 = 110, \alpha_8 = 111$	2^3

信源的数学模型

信源是产生消息（符号）、**消息序列**（符号序列）以及时间连续的消息的来源

随机变量　　随机矢量　　随机过程

2. 用随机矢量（序列）描述信源输出的消息——输出为一系列符号

离散无记忆信源

信源先后发出的符号之间统计独立　　$P(\vec{X}) = P(X_1X_2\cdots X_N) = P(X_1)P(X_2)\cdots P(X_N)$

特例：离散无记忆信源的N次扩展信源　将信源输出的序列看作一组一组发出，每组长度为N。

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & \cdots & a_i & \cdots & a_q \\ p_1 & \cdots & p_i & \cdots & p_q \end{bmatrix} \Rightarrow \begin{bmatrix} X^N \\ P(\vec{X}) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \cdots & \alpha_i & \cdots & \alpha_{q^N} \\ p_1 & \cdots & p_i & \cdots & p_{q^N} \end{bmatrix}$$

每个 α_i 可视为由信源 X 中 N 个 a_i 组成的序列 $\alpha_i = (a_{i_1}a_{i_2}\cdots a_{i_N})$

α_i 的概率 $p(\alpha_i)$ 对应信源 X 中 N 个 a_i 组成的序列的概率

信源的数学模型

信源是产生消息（符号）、**消息序列**（符号序列）以及时间连续的消息的来源

随机变量　　随机矢量　　随机过程

2. 用随机矢量（序列）描述信源输出的消息——输出为一系列符号

离散无记忆信源

信源先后发出的符号之间统计独立　　$P(\vec{X}) = P(X_1X_2\cdots X_N) = P(X_1)P(X_2)\cdots P(X_N)$

特例：离散无记忆信源的N次扩展信源　将信源输出的序列看作一组一组发出，每组长度为N。

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & \cdots & a_i & \cdots & a_q \\ p_1 & \cdots & p_i & \cdots & p_q \end{bmatrix} \Rightarrow \begin{bmatrix} X^N \\ P(\vec{X}) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \cdots & \alpha_i & \cdots & \alpha_{q^N} \\ p_1 & \cdots & p_i & \cdots & p_{q^N} \end{bmatrix}$$

扩展信源中输出符号序列的数目 q^N

q : 符号集中的符号个数，N : 扩展序列长度　　$p(\alpha_i) = p(a_{i_1})p(a_{i_2})\cdots p(a_{i_N}) = \prod_{k=1}^{N} p(a_{i_k})$

信源的数学模型

信源是产生消息（符号）、**消息序列**（符号序列）以及时间连续的消息的来源

随机变量　　随机矢量　　随机过程

2. 用随机矢量（序列）描述信源输出的消息——输出为一系列符号

离散有记忆信源

信源输出的符号之间彼此依存、互不独立，可用联合概率分布或条件概率分布描述这种关联性。实际信源发出的消息符号往往只与前面若干个符号的依赖性较强，因此常研究**有限记忆信源**。

马尔可夫信源是一种特殊的离散非平稳有记忆信源，我们将在后面进行详细讲解。

信源的数学模型

信源是产生消息（符号）、消息序列（符号序列）以及**时间连续的消息**的来源

随机变量　　随机矢量　　随机过程

3. 用随机过程描述信源输出的消息——输出为时间连续的消息

设有一个时间过程 $X(t)$，如果对于每一个固定的时刻 $t_j(j=1,2,\ldots)$，$X(t_j)$是一个随机变量，则称 $X(t)$ 为随机过程。或者说，依赖于时间t的一族随机变量 $X(t)$为随机过程。

简单来说，随机过程就是随机变量与时间维度的结合。

例：某商店在 t_0 时刻到 t_n 这段时间内，接待顾客的人数。

每个时刻的顾客人数可以视为一个随机变量，那么各时刻的人数就可视为一个随机过程。

离散信源的信息熵

小节1 自信息与信息熵

小节2 信息熵的性质

离散信源的信息熵

小节1 自信息与信息熵

小节2 信息熵的性质

随机事件的自信息

随机事件的自信息是指信源发出一个消息本身所包含的信息量，由**该消息的不确定性**所决定。

自信息的定义式为 $I(x_i) = \log \dfrac{1}{p(x_i)} = -\log p(x_i)$

公式中的对数底数为2，即 $I(x_i) = \log_2 \dfrac{1}{p(x_i)} = -\log_2 p(x_i)$，实际书写往往省略数字2；

自信息是确切的物理量，单位为bit，在实际计算时一定要带单位；

r进制自信息的表达式为 $I_r(x_i) = \dfrac{I(x_i)}{\log r}$，单位为对应的$r$进制符号；

利用计算器可计算10为底的对数$\lg x$，则对数换算关系为 $\log_2 p(x_i) = \log_2 10 \cdot \lg p(x_i) = 3.322 \lg p(x_i)$。

随机事件的自信息

$$I(x_i) = \log \frac{1}{p(x_i)} = -\log p(x_i) \quad \begin{cases} p(x_i) = 0 & \Rightarrow & I(x_i) = \infty \\ p(x_i) = 1 & \Rightarrow & I(x_i) = 0 \\ p(x_1) > p(x_2) & \Rightarrow & I(x_1) < I(x_2) \\ I(x_i y_i) = I(x_i) + I(y_i) & & \text{（当两事件相互独立）} \end{cases}$$

自信息的意义 在事件发生前表征事件不确定性的大小

在事件发生后表征事件所提供的信息量

随机事件的自信息

联合自信息

$$I(x_i y_j) = -\log p(x_i y_j)$$

条件自信息

$$I(x_i | y_j) = -\log p(x_i | y_j) \qquad I(y_j | x_i) = -\log p(y_j | x_i)$$

联合自信息、条件自信息与自信息的关系

$$I(x_i y_j) = -\log p(y_j) p(x_i | y_j) = I(y_j) + I(x_i | y_j)$$

$$= -\log p(x_i) p(y_j | x_i) = I(x_i) + I(y_j | x_i)$$

例题2-1 箱子中有90个红球，10个白球，现在从箱子中随机地取出一个球，求：

(1) 事件“取出一个红球”的不确定性；

(2) 事件“取出一个白球”所供的信息量；

(3) 事件“取出一个红球”与“取出一个白球”的发生，哪个更难猜测？

解析2-1 取出一个球的颜色的概率空间为 $\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} X = ''red'' & X = ''white'' \\ 0.9 & 0.1 \end{bmatrix}$

(1) 事件“取出一个红球”的不确定性为 $I(''red'') = -\log 0.9 \approx 0.152bit$

(2) 事件“取出一个白球”所供的信息量为 $I(''white'') = -\log 0.1 \approx 3.322bit$

(3) 事件“取出一个红球”的信息量少于事件“取出一个白球”的信息量，因此事件“取出一个白球”更难猜测。

例题2-2 英文字母中，“a”出现的概率为0.064，“c”出现的概率为0.022；

(1) 试分别计算它们的自信息量；

(2) 假定前后字母出现是相互独立的，计算“ac”的自信息量；

(3) 若前后字母有关联性，当“a”出现后，“c”出现的概率为0.04，

试计算“a” 出现后，“c”出现的自信息量；计算此时“ac”的自信息量

解析2-2 (1) 事件“a出现”的自信息量 $I("a") = -\log 0.064 \approx 3.96bit$

事件“c出现”的自信息量 $I("c") = -\log 0.022 \approx 5.51bit$

(2) 若字母前后相互独立，则“ac”的自信息量 $I("ac") = I("a") + I("c") = 9.47bit$

(3) “a”出现后，“c”出现的自信息量 $I("c" \,|\, "a") = -\log 0.04 \approx 4.64bit$

此时“ac”的自信息量为 $I("ac") = -\log p("a")p("c" \,|\, "a") = I("a") + I("c" \,|\, "a") = 8.6bit$

离散信源的信息熵

信息熵为信源中各消息所包含自信息的数学期望，表征**信源整体的不确定度**，也称为平均自信息。

其定义式为 $H(X)=E[I(x_i)]=\sum_{i=1}^{n}p(x_i)I(x_i)=-\sum_{i=1}^{n}p(x_i)\log p(x_i)$ 。

说明：公式中的对数底数为2，实际书写往往省略；

信息熵的单位为 *bit* / 符号，在实际计算时一定要带单位；

r进制信源熵的表达式为 $H_r(X)=\dfrac{H(X)}{\log r}$ ，单位为r进制单位/符号

利用计算器可以计算10为底的对数lg$x$，则对数换算关系为

$$\log_2 p(x_i)=\log_2 10\cdot \lg p(x_i)=3.322\lg p(x_i)$$

离散信源的信息熵

信息熵的意义

信息熵表征信源输出前，信源的平均不确定性；

信息熵表征信源输出后，每个消息(符号)所包含的平均信息量；

信息熵表征信源，即随机变量 X 的**整体不确定度**。

信息熵本质是概率的函数

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

$$\Updownarrow$$

$$H(X) = H(p(x_1), p(x_2), \cdots, p(x_n)) = H(p_1, p_2, \cdots, p_n)$$

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} X = x_1 & X = x_2 \\ p & 1 - p \end{bmatrix}$$

$$\Updownarrow$$

$$H(X) = H(p, 1 - p)$$

例题2-3 一离散无记忆信源的概率空间为 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1=0 & a_2=1 & a_3=2 & a_4=3 \\ \frac{3}{8} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} \end{bmatrix}$

其发出的消息为（202120130213001203210110321010021032011223210）。

(1) 此消息的自信息是多少?

(2) 在此消息中平均每个符号携带的信息量是多少?

(3) 信源的信息熵 $H(X)$ 是多少?

解析2-3 消息（202120130213001203210110321010021032011223210）共有45个符号，其中有14个0，13个1，12个2，6个3

(1) $I(''a_1=0'') = -\log\frac{3}{8} \approx 1.415bit$ $\qquad I(''a_2=1'') = -\log\frac{1}{4} = 2bit$

$I(''a_3=2'') = -\log\frac{1}{4} = 2bit$ $\qquad I(''a_4=3'') = -\log\frac{1}{8} = 3bit$

信源无记忆，故符号之间相互独立，则该消息的自信息

$$I = 14I(''a_1=0'') + 13I(''a_2=1'') + 12I(''a_3=2'') + 6I(''a_4=3'')$$

$$= 14\times 1.415 + 13\times 2 + 12\times 2 + 6\times 3 = 87.81bit$$

例题2-3 一离散无记忆信源的概率空间为 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1=0 & a_2=1 & a_3=2 & a_4=3 \\ \frac{3}{8} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} \end{bmatrix}$

其发出的消息为（202120130213001203210110321010021032011223210）。

(1) 此消息的自信息是多少?

(2) 在此消息中平均每个符号携带的信息量是多少?

(3) 信源的信息熵 $H(X)$ 是多少?

解析2-3 消息（202120130213001203210110321010021032011223210）共有45个符号，其中有14个0，13个1，12个2，6个3

(2) 在此消息中平均每个符号携带的信息量

$$\bar{I} = \frac{I}{N} = \frac{87.81bit}{45} \approx 1.95bit/symbol$$

(3) 此信源的信息熵

$$H(X) = H(\frac{3}{8},\frac{1}{4},\frac{1}{4},\frac{1}{8}) = -\frac{3}{8}\log\frac{3}{8} - \frac{1}{4}\log\frac{1}{4}\times 2 - \frac{1}{8}\log\frac{1}{8}$$

$$= 1.91bit/symbol$$

离散信源的信息熵

小节1 自信息与信息熵

小节2 信息熵的性质

信息熵的性质「对称性」

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} X = x_1 & X = x_2 & \cdots & X = x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix} \Rightarrow \quad \text{概率矢量} \ \vec{P} = (p_1, p_2, \cdots, p_n)$$

当概率矢量中各概率分量的次序任意变更时，熵保持不变，即：

$$H(p_1, p_2, \cdots, p_n) = H(p_2, p_3, \cdots, p_n, p_1) = \cdots = H(p_n, p_1, \cdots, p_{n-1})$$

熵只与随机变量的总体结构有关，与信源的总体统计特性有关；

两个信源，只要发出的消息数相同，概率分布相同，那么信源熵必然相同；

信息熵不能述信源消息的具体含义和主观价值，只是客观的数学度量。

信息熵的性质「确定性」

$$H(1,0) = H(1,0,0) = \cdots = H(1,0,0,\cdots,0) = 0$$

信源的概率空间中，只要有一个事件是必然事件，那么其他事件必为不可能事件，熵为0。

信息熵的性质「非负性」

$H(p_1, p_2, \cdots, p_n) \geq 0$ 只有当随机变量为确知量时，等号成立。

注意：熵的非负性仅对离散信源成立，连续信源的熵可以为负。

信息熵的性质「扩展性」

$$\lim_{\varepsilon \to 0} H_{n+1}(p_1, p_2, \cdots, p_n - \varepsilon, \varepsilon) = H_n(p_1, p_2, \cdots, p_n)$$

在概率空间中增加一个发生概率极小的随机事件，信源熵保持不变。

信息熵的性质「强可加性」

$H(XY) = H(X) + H(Y|X) = H(Y) + H(X|Y)$ 条件熵

物理意义

(1) 在信源 X 和 Y 相关联时，信源 XY 每发一个符号组合所能供的平均信息量，等于信源 X (或 Y) 每发出一个符号所能供的平均信息量，再加上在 X (或 Y) 已知的条件下信源 Y (或 X) 再发出一个符号所供的平均信息量。

(2) 当 X 和 Y 相互独立时，$H(XY) = H(X) + H(Y)$ ，此公式称为信息熵的可加性。

信息熵的性质「强可加性」

条件熵的定义 条件自信息对联合概率的数学期望

定义式

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

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

计算常用公式

$$H(Y|X)=\sum_{i,j}p(x_iy_j)\log\frac{1}{p(y_j|x_i)}=\sum_{i,j}p(x_i)\boxed{p(y_j|x_i)\log\frac{1}{p(y_j|x_i)}}\Rightarrow H(Y|X)=\sum_{i=1}^{n}p(x_i)H(Y|x_i)$$

$$H(X|Y)=\sum_{i,j}p(x_iy_j)\log\frac{1}{p(x_i|y_j)}=\sum_{i,j}p(y_j)\boxed{p(x_i|y_j)\log\frac{1}{p(x_i|y_j)}}\Rightarrow H(X|Y)=\sum_{j=1}^{m}p(y_j)H(X|y_j)$$

例题2-4 已知随机变量 X, Y 的联合概率分布和条件概率分布如下所示，试求条件熵 $H(Y|X)$

$$P_{XY} = \begin{array}{c} \\ x_1 \\ x_2 \end{array}\!\!\begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{bmatrix} \frac{1}{8} & \frac{1}{8} & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \end{array} \qquad P_{Y|X} = \begin{array}{c} \\ x_1 \\ x_2 \end{array}\!\!\begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \end{array}$$

解析2-4 法一：利用熵的强可加性 $H(XY) = H(X) + H(Y|X)$

根据联合概率矩阵，易得 $P_X = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \end{bmatrix}$ $P_Y = \begin{bmatrix} \frac{1}{8} & \frac{3}{8} & \frac{1}{2} \end{bmatrix}$

$$H(Y|X) = H(XY) - H(X)$$

$$= H(\frac{1}{8}, \frac{1}{8}, \frac{1}{4}, \frac{1}{2}) - H(\frac{1}{4}, \frac{3}{4})$$

$$= (-\frac{2}{8}\log\frac{1}{8} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{2}\log\frac{1}{2}) - (-\frac{1}{4}\log\frac{1}{4} - \frac{3}{4}\log\frac{3}{4})$$

$$\approx 0.94 bit/symbol$$

例题2-4 已知随机变量 X, Y 的联合概率分布和条件概率分布如下所示，试求条件熵 $H(Y|X)$

$$P_{XY}=\begin{array}{c c} & \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \end{array} & \begin{bmatrix} \frac{1}{8} & \frac{1}{8} & 0 \\ 0 & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \end{array} \qquad P_{Y|X}=\begin{array}{c c} & \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \end{array} & \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \end{array}$$

解析2-4 法二：利用公式 $H(Y|X)=\sum_{i=1}^{n} p(x_i)H(Y|x_i)$

$$P_X=\begin{bmatrix} \frac{1}{4} & \frac{3}{4} \end{bmatrix} \qquad P_{Y|X}=\begin{array}{c c} & \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \end{array} & \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix} \end{array}$$

$$H(Y|X)=\frac{1}{4}H(\frac{1}{2},\frac{1}{2})+\frac{3}{4}H(\frac{1}{3},\frac{2}{3})$$

$$=\frac{1}{4}(-\frac{1}{2}\log\frac{1}{2}-\frac{1}{2}\log\frac{1}{2})+\frac{3}{4}(-\frac{1}{3}\log\frac{1}{3}-\frac{2}{3}\log\frac{2}{3})$$

$$\approx 0.94 bit/symbol$$

例题2-5 已知随机变量 X, Y 的联合概率分布如下所示

试求下列熵：$H(X)$、$H(Y)$、$H(XY)$、$H(X|Y)$、$H(Y|X)$

$$P_{XY} = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{array}\begin{array}{c} \begin{array}{cccc} y_1 & y_2 & y_3 & y_4 \end{array} \\ \begin{bmatrix} 0.25 & 0 & 0 & 0 \\ 0.10 & 0.30 & 0 & 0 \\ 0 & 0.05 & 0.10 & 0 \\ 0 & 0 & 0.05 & 0.10 \\ 0 & 0 & 0.05 & 0 \end{bmatrix} \end{array}$$

解析2-5 根据联合概率分布，可求得边缘概率 $P_X = [0.25 \quad 0.40 \quad 0.15 \quad 0.15 \quad 0.05]$ $P_Y = [0.35 \quad 0.35 \quad 0.20 \quad 0.10]$

$$H(X) = H(0.25,0.40,0.15,0.15,0.05) \approx 2.066 bit/symbol$$

$$H(Y) = H(0.35,0.35,0.20,0.10) \approx 1.856 bit/symbol$$

$$H(XY) = H(0.25,0.10,0.30,0.05,0.10,0.05,0.05,0.10) \approx 2.665 bit/2symbols$$

$$H(X|Y) = H(XY) - H(Y) = 0.809 bit/symbol$$

$$H(Y|X) = H(XY) - H(X) = 0.6 bit/symbol$$

信息熵的性质「强可加性」

强可加性的拓展结论

多随机变量联合熵链规则 $H(X_1X_2\cdots X_N)=H(X_1)+H(X_2|X_1)+H(X_3|X_2X_1)+\cdots+H(X_N|X_{N-1}X_{N-2}\cdots X_2X_1)$

N个变量相互独立时 $H(X_1X_2\cdots X_N)=\sum_{i=1}^{N}H(X_i)$

N个变量独立同分布时 $H(X_1X_2\cdots X_N)=NH(X)$

几个不等关系 $\begin{cases} H(X|Y)\le H(X) \\ H(XY)\le H(X)+H(Y) \\ H(X_1X_2\cdots X_N)\le H(X_1)+H(X_2)+\cdots+H(X_N) \end{cases}$ (N个变量独立时等号成立)

信息熵的性质「上凸性与极值性」

- **上凸性**

信息熵是概率分布的严格上凸函数。

- **极值性**

信息熵是概率分布的严格上凸函数，因此信息熵存在极大值。

离散无记忆信源输出n个不同的信息符号时，当且仅当各个符号出现的概率均等时，熵最大，即：

$$H(p_1, p_2, \cdots, p_n) \leq H(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}) = \log n$$

直观理解：概率空间内各随机事件发生机会均等时，不确定性最大。

信息熵的性质「上凸性与极值性」

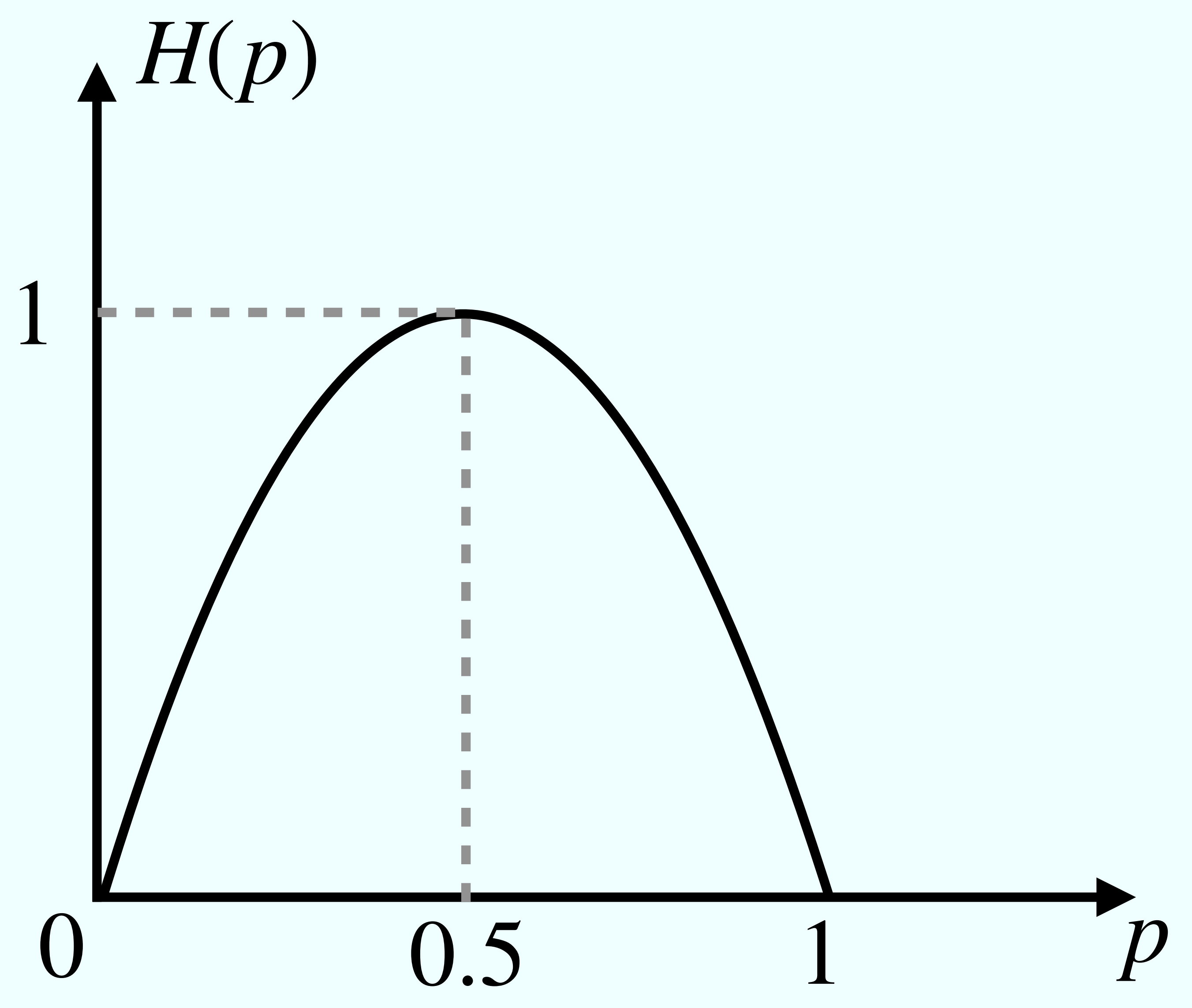
- **极值性**

离散无记忆信源输出n个不同的信息符号时，当且仅当各个符号出现的概率均等时，熵最大，即：

$$H(p_1, p_2, \cdots, p_n) \leq H(\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}) = \log n$$

例：二进制信源概率空间

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix} \quad H(X) = -p\log p - (1-p)\log(1-p)$$



当p=0.5时，熵最大，为1bit/symbol

离散无记忆信源的扩展信源

小节1 信源模型

小节2 信息熵

离散无记忆信源的扩展信源

离散无记忆信源的扩展信源的数学模型

离散无记忆信源

信源先后发出的符号之间**统计独立** $P(\vec{X}) = P(X_1X_2\cdots X_N) = P(X_1)P(X_2)\cdots P(X_N)$

离散无记忆信源的扩展信源

$\boxed{\begin{array}{l}\text{符号序列中各符号取自同一个信源空间 } [X,P] \\ \text{符号序列中前后符号的出现}\textbf{相互独立，彼此无关}\end{array}}$

独立同分布

扩展信源——输出为**多个符号为一组**的信源

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & \cdots & a_i & \cdots & a_q \\ p_1 & \cdots & p_i & \cdots & p_q \end{bmatrix} \Rightarrow \begin{bmatrix} X^N \\ P(\vec{X}) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \cdots & \alpha_i & \cdots & \alpha_{q^N} \\ p_1 & \cdots & p_i & \cdots & p_{q^N} \end{bmatrix}$$

每个 α_i 可视为由信源 X 中 N 个 a_i 组成的序列 $\alpha_i = (a_{i_1}a_{i_2}\cdots a_{i_N})$

α_i 的概率 $p(\alpha_i)$ 对应信源 X 中 N 个 a_i 组成的序列的概率

离散无记忆信源的扩展信源的数学模型

扩展信源中输出符号序列的数目 q^N

q: 符号集中的符号个数，N: 扩展序列长度

$$\alpha_i = (a_{i_1} a_{i_2} \cdots a_{i_N}) \quad \Rightarrow \quad p(\alpha_i) = p(a_{i_1}) p(a_{i_2}) \cdots p(a_{i_N}) = \prod_{k=1}^{N} p(a_{i_k})$$

离散无记忆信源的扩展信源

小节1 信源模型

小节2 信息熵

离散无记忆信源的扩展信源的熵

由于离散无记忆信源的扩展信源符号序列中前后符号的出现相互独立，彼此无关，因此有

$$H(X^N) = H(X_1X_2\cdots X_N) = H(X_1) + H(X_2) + \cdots + H(X_N)$$

符号序列中各符号取自同一个信源空间 $[X,P]$，故有

$$H(X_1) = H(X_2) = \cdots = H(X_N) = H(X)$$

故离散无记忆信源扩展信源的熵为 $H(X^N) = NH(X)$

离散无记忆信源扩展信源的熵和N长序列变量独立同分布的信源熵相同。

例题2-6 一离散无记忆信源的概率空间为 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$，试求其二次扩展信源的信息熵$H(X^2)$。

解析2-6 原始离散无记忆信源的信息熵为

$$H(X) = H(\frac{1}{2},\frac{1}{4},\frac{1}{4}) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = 1.5bit/symbol$$

则其二次扩展信源的信息熵为

$$H(X^2) = 2H(X) = 3bit/2symbols$$

例题2-7

每帧电视图像都可以认为是由500*600个像素点组成，所有像素点独立变化， 且每个像素点都取128个不同的亮度电平，且这128个亮度电平等概率出现，规定每秒传送30帧图像，问：

(1) 每帧图像含有多少信息量?

(2) 传递电视图像获得的信息率(bit/sec)是多少?

(3) 若广播员在约10000个汉字组成的字汇中选1000字来口述这一帧图像，那么口述此图像所供的信息量是多少?（假设汉字字汇为等概率分布，且彼此无依赖）

(4) 若要恰当地描述这帧图像，广播员在口述中至少需要多少汉字?

解析2-7

序列 独立 同分布 ⇒ 扩展信源

(1) 每个像素点的概率空间为 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_{128} \\ \frac{1}{128} & \frac{1}{128} & \cdots & \frac{1}{128} \end{bmatrix}$

则每个像素点的平均信息量为 $H(X) = \log 128 = 7bit/symbol$

所有像素点独立变化， 且每个像素点都取128个不同的亮度电平，且这128个亮度电平等概率出现，

因此每帧图像可以视为像素点（无记忆信源）的500*600次扩展；

则每帧图像的平均信息量为 $H(X^N) = NH(X) = 3 \times 10^5 \times 7 = 2.1 \times 10^6 bit/$帧

例题2-7 每帧电视图像都可以认为是由500*600个像素点组成，所有像素点独立变化， 且每个像素点都取128个不同的亮度电平，且这128个亮度电平等概率出现，规定每秒传送30帧图像，问：

(1) 每帧图像含有多少信息量?

(2) 传递电视图像获得的信息率(bit/sec)是多少?

(3) 若广播员在约10000个汉字组成的字汇中选1000字来口述这一帧图像，那么口述此图像所供的信息量是多少?（假设汉字字汇为等概率分布，且彼此无依赖）

(4) 若要恰当地描述这帧图像，广播员在口述中至少需要多少汉字?

解析2-7 (2) 每秒传送30帧图像，则传送电视图像获得的信息率为

$$30 \times H(X^N) = 30\text{帧}/sec \times 2.1 \times 10^6 bit/\text{帧} \ = 6.3 \times 10^7 bit/sec$$

(3) 每个汉字的概率空间为 $\begin{bmatrix} Y \\ P(Y) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_{10000} \\ \frac{1}{10000} & \frac{1}{10000} & \cdots & \frac{1}{10000} \end{bmatrix}$

则每个汉字的平均信息量为 $H(Y) = \log 10000 \approx 13.29 bit/symbol$

假设汉字字汇为等概率分布，且彼此无依赖，

因此1000字的口述稿可以视为汉字信源（无记忆）的1000次扩展；

则广播员提供的平均信息量为 $H(Y^M) = MH(Y) = 1000 \times 13.29 = 1.329 \times 10^4 bit/\text{千字}$

例题2-7

每帧电视图像都可以认为是由500*600个像素点组成，所有像素点独立变化， 且每个像素点都取128个不同的亮度电平，且这128个亮度电平等概率出现，规定每秒传送30帧图像，问：

(1) 每帧图像含有多少信息量?

(2) 传递电视图像获得的信息率(bit/sec)是多少?

(3) 若广播员在约10000个汉字组成的字汇中选1000字来口述这一帧图像，那么口述此图像所供的信息量是多少?（假设汉字字汇为等概率分布，且彼此无依赖）

(4) 若要恰当地描述这帧图像，广播员在口述中至少需要多少汉字?

解析2-7

(4) 若要准确述这幅画面，则所有汉字的平均信息量应与一帧画面所供的信息量相匹配

因此所需要的汉字数为 $M=\dfrac{H(Y^M)}{H(Y)}=\dfrac{H(X^N)}{H(Y)}=\dfrac{2.1\times10^6}{13.29}=1.58\times10^5$

离散平稳信源及其极限熵

小节1 离散平稳信源

小节2 极限熵

离散平稳信源及其极限熵

小节1 离散平稳信源

小节2 极限熵

离散平稳信源的定义

对于离散平稳信源，输出的随机序列 $\vec{X}=(X_1X_2\cdots X_N\cdots)$ 中每个随机变量都是取值离散的离散型随机变量，且随机矢量的各维概率分布都与时间起点无关。

对任意两个不同起始时刻 i, j, 信源发出消息的联合概率相同，
即对于任意的$N=0,1,2\ldots$, 消息$X_iX_{i+1}\cdots X_{i+N}$ 与$X_jX_{j+1}\cdots X_{j+N}$具有相同的概率分布

$P(X_i)=P(X_j)$ $\Rightarrow$ 一维平稳

$P(X_iX_{i+1})=P(X_jX_{j+1})$ $\Rightarrow$ 二维平稳

……

$P(X_iX_{i+1}\cdots X_{i+N-1})=P(X_jX_{j+1}\cdots X_{j+N-1})$ $\Rightarrow$ N 维平稳

N维平稳必然N-1维、N-2维、……、一维平稳

离散平稳信源的特点

$$P(X_i) = P(X_j)$$
$$P(X_iX_{i+1}) = P(X_jX_{j+1})$$
$$P(X_iX_{i+1}\cdots X_{i+N}) = P(X_jX_{j+1}\cdots X_{j+N})$$
$$\cdots\cdots$$
$$P(X_{i+1}|X_i) = P(X_{j+1}|X_j)$$
$$P(X_{i+N}|X_iX_{i+1}\cdots X_{i+N-1}) = P(X_{j+N}|X_jX_{j+1}\cdots X_{j+N-1})$$

$$H(X_1) = H(X_2) = \cdots = H(X_N)$$
$$H(X_1X_2) = H(X_2X_3) = \cdots = H(X_{N-1}X_N)$$
$$H(X_1X_2X_3) = H(X_2X_3X_4) = \cdots = H(X_{N-2}X_{N-1}X_N)$$
$$\cdots\cdots$$
$$H(X_2|X_1) = H(X_3|X_2) = \cdots = H(X_N|X_{N-1})$$
$$H(X_3|X_1X_2) = H(X_4|X_2X_3) = \cdots = H(X_N|X_{N-2}X_{N-1})$$

各维条件概率只与关联长度N有关，与时间起点无关。

因此各维条件熵只与关联长度N有关，与时间起点无关。

离散平稳信源及其极限熵

离散平稳信源的三种信息测度方式（考虑信源的记忆性）

序列 $\vec{X}=(X_1X_2\cdots X_N)$ 的联合熵

$$H(\vec{X})=H(X_1)+H(X_2|X_1)+H(X_3|X_2X_1)+\cdots+H(X_N|X_{N-1}X_{N-2}\cdots X_2X_1)$$

表示信源每发一个N长序列所提供的平均信息量

序列 $\vec{X}=(X_1X_2\cdots X_N)$ 的平均符号熵

$$H_N(\vec{X})=\frac{1}{N}H(\vec{X})=\frac{1}{N}H(X_1X_2\cdots X_N)$$

表示信源每发一个N长序列平均每个符号所提供的信息量

序列 $\vec{X}=(X_1X_2\cdots X_N)$ 的**极限熵**

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

表示信源每发一个N长序列，当N趋于无穷大时，信源所提供的平均信息量

离散平稳信源的极限熵

序列 $\vec{X} = (X_1X_2\cdots X_N)$ 的**极限熵**

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

对于离散平稳有限记忆长度为m+1的信源（m+1维平稳信源），其极限熵等于m阶条件熵，即：

$$H_\infty = H_{m+1} = H(X_{m+1} | X_1X_2\cdots X_m)$$

记忆长度：m+1
条件概率阶数：m

关于极限熵的说明

在实际情况中，信源在不断地发出符号，符号之间的统计关联关系伸向无穷远；

此时，极限熵才能确切表达多符号离散平稳有记忆信源平均每发一个符号所提供的信息量；

但在分析中，记忆长度无穷的信源熵求解难度较大，因此我们常用有限长度条件熵近似代替极限熵。

例题2-8 二维平稳信源的一维概率分布为 $\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ \frac{1}{4} & \frac{4}{9} & \frac{11}{36} \end{bmatrix}$

输出符号序列中，只有前后两个符号之间有记忆，条件概率见表格，

求此离散平稳信源的极限熵。

X_i \ X_{i+1}	x_1	x_2	x_3
x_1	$\frac{7}{9}$	$\frac{2}{9}$	0
x_2	$\frac{1}{8}$	$\frac{3}{4}$	$\frac{1}{8}$
x_3	0	$\frac{2}{11}$	$\frac{9}{11}$

解析2-8 该平稳信源记忆长度为2，因此其极限熵为一阶条件熵；

$$H_\infty = H_2 = H(X_2|X_1) = \sum_{i=1}^{3} p(x_i)H(X|x_i)$$

$$= \frac{1}{4}H(\frac{7}{9},\frac{2}{9}) + \frac{4}{9}H(\frac{1}{8},\frac{3}{4},\frac{1}{8}) + \frac{11}{36}H(\frac{2}{11},\frac{9}{11})$$

$$= 0.87bit/symbol$$

马尔可夫信源

小节1 马尔可夫链

小节2 马尔可夫信源及其信源熵

马尔可夫信源

小节1 马尔可夫链

小节2 马尔可夫信源及其信源熵

马尔可夫链的定义

马尔可夫链是一种特殊的随机过程

设 $\{X_n,\ n \geq 0,\ n \in Z\}$ 为一个随机状态序列；

时间参数集为 $\{0,1,2,\cdots\}$，状态空间为可数集 $S = \{S_0, S_1, S_2, \cdots, S_j, \cdots\}$；

对所有 $n \geq 1$，有如下统计特性：

$$P(X_n = S_{i_n} | X_{n-1} = S_{i_{n-1}}, X_{n-2} = S_{i_{n-2}}, \cdots, X_2 = S_{i_2}, X_1 = S_{i_1}) = P(X_n = S_{i_n} | X_{n-1} = S_{i_{n-1}})$$

则称 $\{X_n,\ n \geq 0,\ n \in Z\}$ 为马尔可夫链。

其特点是：随机变量的状态仅与上一时刻所处的状态有关联，与再之前的状态则无关联。

马尔可夫链的状态转移概率

已知状态 S_i 时，n-m 步后为状态 S_j 的条件概率为 $P(X_n = S_j | X_m = S_i) = p_{ij}(m,n) = p_{ij}^{(n-m)}$

我们记状态 S_i 时，k步后为状态 S_j 的条件概率为k步转移概率 $p_{ij}^{(k)} = P(X_{m+k} = S_j | X_m = S_i)$；

那么，状态 S_i 时，1步后为状态 S_j 的条件概率为1步转移概率 $p_{ij} = P(X_{m+1} = S_j | X_m = S_i)$ ；

一般来说，状态转移概率与起始时刻、转移步长和所处状态都有关。

时齐马尔可夫链的定义

- **时齐（齐次）马尔可夫链**

若马尔可夫链的状态转移概率与初始时刻无关，则称此马氏链为时齐（齐次）马尔可夫链，此时的状态转移概率不随时间的改变而改变，也称为平稳转移概率的马尔可夫链。

我们主要研究的就是**时齐马尔可夫链**，它的统计特性具有十分重要的地位。

时齐马尔可夫链的状态转移矩阵

因为时齐马尔可夫链的状态转移概率与初始时刻无关，其转移概率是平稳的，因此我们可以利用**转移概率矩阵**或**状态转移图**对其进行表征。

其中，一步状态转移概率 $p_{ij} = P(X_{m+1} = S_j | X_m = S_i)$ 对应的矩阵如下：

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{m1} & p_{m2} & \cdots & p_{mn} \end{bmatrix}$$

其中，矩阵中每个元素非负，且每行元素之和为1。

我们可以推得时齐马尔可夫链一步转移概率与二步转移概率的关系是：$P^{(2)} = P \cdot P = P^2$

对于任意两个正整数m和n，且k=m+n，有k步转移概率 $P^{(k)} = P^{(m)} \cdot P^{(n)}$，该关系称为切普曼-柯尔莫哥洛夫方程（简称C-K方程）

根据切普曼-柯尔莫哥洛夫方程，我们可以知道，k步转移，可以等价于先进行m步转移，再进行n步转移，然后两次转移的矩阵相乘，其中k=m+n。

时齐马尔可夫链的状态转移矩阵

对于任意两个正整数m和n，且k=m+n，有k步转移概率 $P^{(k)}=P^{(m)}\cdot P^{(n)}$，该关系称为切普曼-柯尔莫哥洛夫方程（简称C-K方程）

根据切普曼-柯尔莫哥洛夫方程，我们可以知道，k步转移，可以等价于先进行m步转移，再进行n步转移，然后两次转移的矩阵相乘，其中k=m+n。

因此我们有 $P^{(k)}=P^{(m)}\cdot P^{(n)}=P\cdot P^{(k-1)}=P\cdot P\cdot P^{(k-2)}=\cdots=P^{k}$ ；

对于一般马尔可夫链来说，只有已知**各时刻的一步状态概率**，以及**初始分布**，才可以完整描述马尔可夫链的整个统计特性。

但对于时齐马尔可夫链来说，其**一步状态概率** P 决定了多步状态概率，因此**由初始分布以及一步转移概率**，即可描述时齐马尔可夫链的整个统计特性。

马尔可夫链的各态历经性

一般来说，马尔可夫链是非平稳的随机过程，时齐马尔可夫链也是如此。

但**时齐有限状态马尔可夫链**在一定条件下，最终绝对概率分布会达到稳态，便于近似为平稳信源。

- **各态历经性定理**

当**时齐**、**有限状态**的马尔可夫链满足各态历经性，即存在正整数r，使得状态转移矩阵P^r中所有元素都大于零时，在步长足够大的情况下，最终各状态的绝对概率分布会达到稳态

$$\lim_{n\to\infty} p_{ij}^{(n)} = W_j$$

定理表明：只要转移步长足够大，那么经过n步转移后处在第j个状态的概率分布已和初始处于什么状态无关，且趋于某一个固定分布。

无论之前处于哪个状态，经过足够长的步数后跳转到第j个状态的概率相同。

马尔可夫链的各态历经性

- **稳态分布的求解**

在判别各态历经性之后，利用如下的矩阵求解稳态分布。

$$\begin{cases} \sum_i W_i p_{ij} = W_j \\ \sum_j W_j = 1 \end{cases} \Leftrightarrow \begin{cases} WP = W \\ \sum_j W_j = 1 \end{cases}$$

P 为状态转移矩阵

稳态分布为 $W = [W_1 \quad W_2 \quad \cdots \quad W_N]$

马尔可夫链的分析思路

1. 判断是否时齐——转移概率是否平稳——状态转移矩阵P
2. **判断是否遍历——极限分布是否存在——利用状态转移矩阵元素判别**（很关键！）
3. 利用方程组进行稳态分布的求解

$$\begin{cases} \sum_i W_i p_{ij} = W_j \\ \sum_j W_j = 1 \end{cases} \Leftrightarrow \begin{cases} WP = W \\ \sum_j W_j = 1 \end{cases}$$

P 为状态转移矩阵

稳态分布为 $W = [W_1 \quad W_2 \quad \cdots \quad W_N]$

例题2-9 设有一个时齐的马尔可夫链，其状态转移矩阵如右所示。

试判断其稳态分布是否存在，若存在，请求出该稳态分布。

$$P=\begin{bmatrix}0 & 0 & 1\\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6}\\ \frac{1}{2} & \frac{1}{2} & 0\end{bmatrix}$$

解析2-9 首先判断其稳态分布是否存在：

$$P=\begin{bmatrix}0 & 0 & *\\ * & * & *\\ * & * & 0\end{bmatrix}, P^2=\begin{bmatrix}0 & 0 & *\\ * & * & *\\ * & * & 0\end{bmatrix}\begin{bmatrix}0 & 0 & *\\ * & * & *\\ * & * & 0\end{bmatrix}=\begin{bmatrix}* & * & 0\\ * & * & *\\ * & * & *\end{bmatrix},$$

$$P^3=P^2P=\begin{bmatrix}* & * & 0\\ * & * & *\\ * & * & *\end{bmatrix}\begin{bmatrix}0 & 0 & *\\ * & * & *\\ * & * & 0\end{bmatrix}=\begin{bmatrix}* & * & *\\ * & * & *\\ * & * & *\end{bmatrix},$$

P^3中各元素都大于零，因此稳态分布存在。

例题2-9 设有一个时齐的马尔可夫链，其状态转移矩阵如右所示。

试判断其稳态分布是否存在，若存在，请求出该稳态分布。

$$P=\begin{bmatrix}0 & 0 & 1\\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6}\\ \frac{1}{2} & \frac{1}{2} & 0\end{bmatrix}$$

解析2-9 求解稳态分布：

根据 $\begin{cases}WP=W\\ \sum_j W_j=1\end{cases} \Leftrightarrow \begin{cases}[W_1 \quad W_2 \quad W_3]\begin{bmatrix}0 & 0 & 1\\ \frac{1}{2} & \frac{1}{3} & \frac{1}{6}\\ \frac{1}{2} & \frac{1}{2} & 0\end{bmatrix}=[W_1 \quad W_2 \quad W_3]\\ W_1+W_2+W_3=1\end{cases}$

$$\begin{cases}0\cdot W_1+1/2\cdot W_2+1/2\cdot W_3=W_1\\ 0\cdot W_1+1/3\cdot W_2+1/2\cdot W_3=W_2\\ 1\cdot W_1+1/6\cdot W_2+0\cdot W_3=W_3\\ W_1+W_2+W_3=1\end{cases} \Rightarrow \begin{cases}W_1=\frac{1}{3}\\ W_2=\frac{2}{7}\\ W_3=\frac{8}{21}\end{cases}$$

因此稳态分布为 $[W_1 \quad W_2 \quad W_3]=\begin{bmatrix}\frac{1}{3} & \frac{2}{7} & \frac{8}{21}\end{bmatrix}$

稳态分布不存在的时候，方程组也会有对应解，因此必须先验证稳态分布存在，再求稳态分布！

马尔可夫信源

小节1 马尔可夫链

小节2 马尔可夫信源及其信源熵

马尔可夫信源的定义

- ***M*** **阶马尔可夫信源**

当信源的记忆长度为 M+1 时，该时刻的符号只依赖于前面M 个符号（条件长度为M），而不取决于更前面的符号，这样的信源称为M阶马尔可夫信源。

$$\cdots X_1 X_2 \cdots X_m X_{m+1} X_{m+2} \cdots$$

$$p(x_{i_{m+1}} | x_{i_1} x_{i_2} \cdots x_{i_m}) = \boxed{p(x_{i_{m+1}} | S_i)}$$

符号条件概率

$$\cdots\cdots x_{i_1} x_{i_2} \cdots x_{i_m} x_{i_{m+1}} x_{i_{m+2}} \cdots\cdots$$

S_i S_j

状态示意图

$$p(x_{i_{m+1}} x_{i_m} x_{i_{m-1}} \ldots x_{i_2} | x_{i_m} x_{i_{m-1}} x_{i_{m-2}} \ldots x_{i_1}) = \boxed{p(S_j | S_i)}$$

状态转移概率

当信源为一阶马尔可夫信源时，每个状态只有一个符号

此时二者完全等价

把时刻 i_1 到 i_m 的m个变量组成的符号序列看做一个状态，则后一个状态 S_j，即时刻 i_2 到 i_{m+1}的m个变量组成的符号序列仅与状态 S_i 有关，与再往前的状态则无关，服从马尔可夫链的定义。

马尔可夫信源的定义

- **时齐马尔可夫信源**

若条件概率 $p(x_{i_{m+1}}|x_{i_1}x_{i_2}\cdots x_{i_m})=p(S_j|S_i)$与时间起点无关，则信源的输出状态序列可看作时齐的马尔可夫链，则称信源为时齐马尔可夫信源。

举例 设有一个二元二阶时齐马尔可夫信源

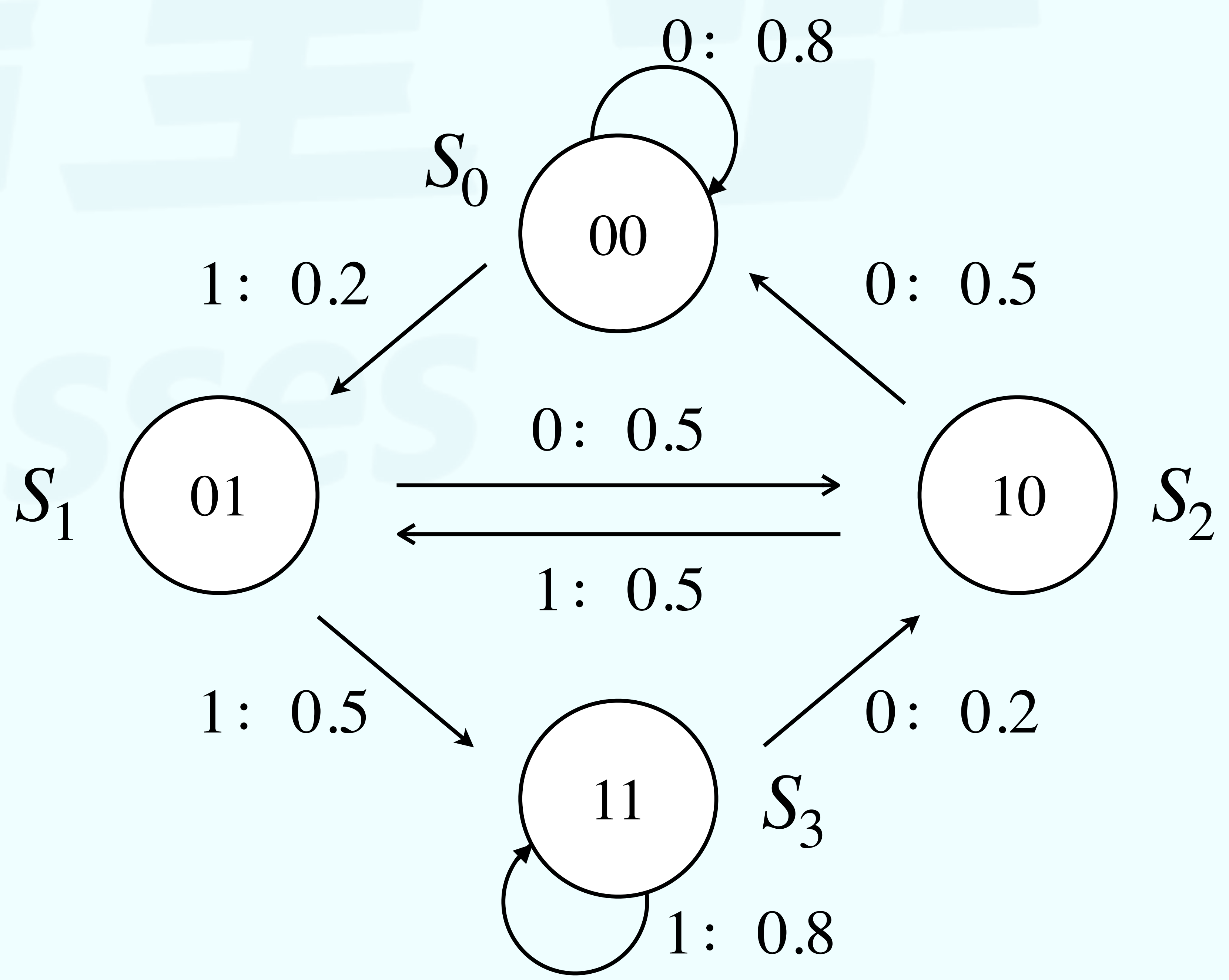
符号集 $X=\{0,1\}$

状态集 $S=\{00,01,10,11\}$

$p(0|00)=p(1|11)=0.8$

$p(1|00)=p(0|11)=0.2$

$p(0|01)=p(1|01)=p(0|10)=p(1|10)=0.5$



马尔可夫信源的定义

举例

$$p(0|00)=p(1|11)=0.8$$

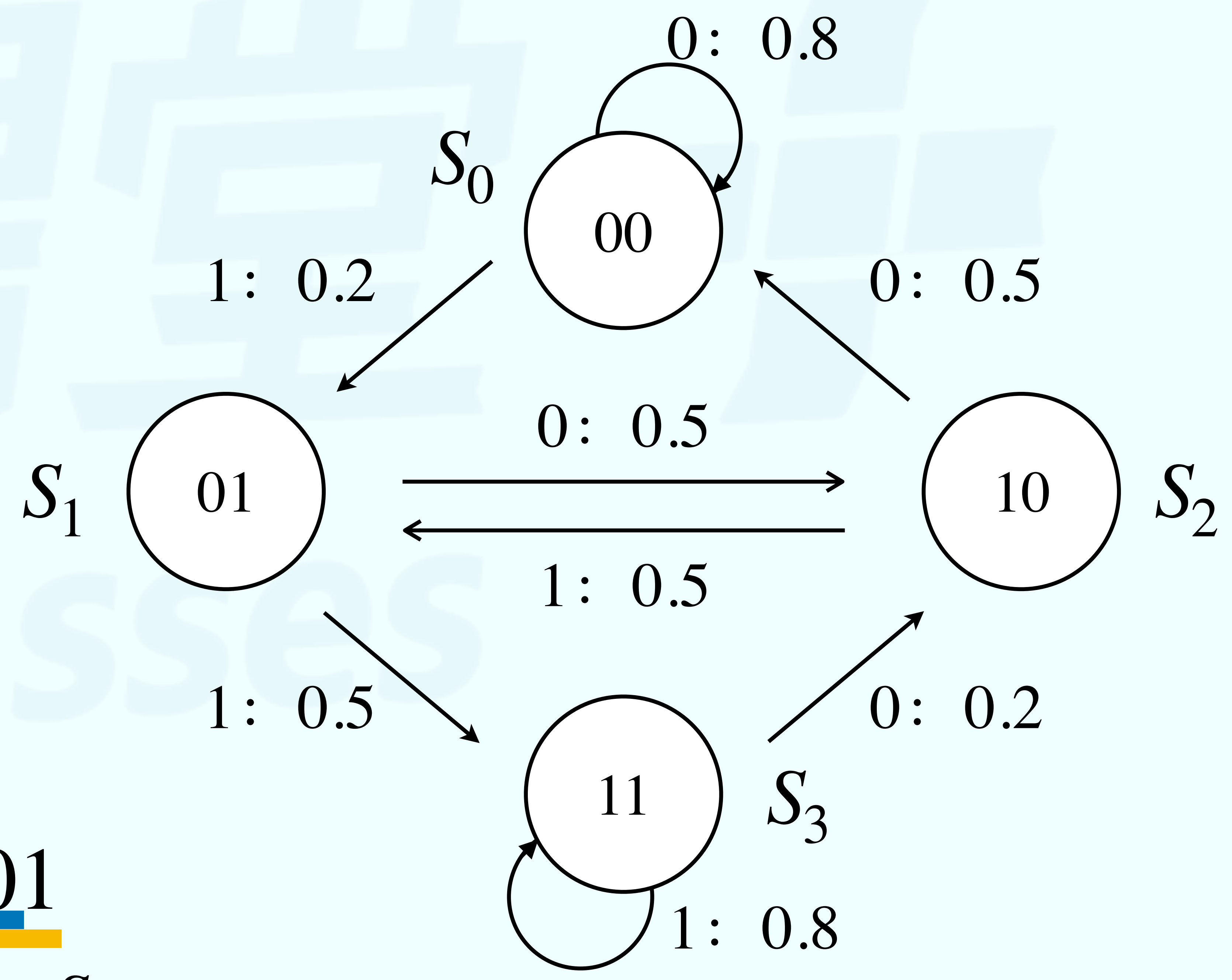
000 $S_0 \to S_0$　　111 $S_3 \to S_3$

$$p(1|00)=p(0|11)=0.2$$

001 $S_0 \to S_1$　　110 $S_3 \to S_2$

$$p(0|01)=p(1|01)=p(0|10)=p(1|10)=0.5$$

010 $S_1 \to S_2$　　011 $S_1 \to S_3$　　100 $S_2 \to S_0$　　101 $S_2 \to S_1$



马尔可夫信源的定义

举例

$$p(0|00)=p(1|11)=0.8$$

$$\underset{S_0\to S_0}{000}\qquad \underset{S_3\to S_3}{111}$$

$$p(1|00)=p(0|11)=0.2$$

$$\underset{S_0\to S_1}{001}\qquad \underset{S_3\to S_2}{110}$$

$$p(0|01)=p(1|01)=p(0|10)=p(1|10)=0.5$$

$$\underset{S_1\to S_2}{010}\qquad \underset{S_1\to S_3}{011}\qquad \underset{S_2\to S_0}{100}\qquad \underset{S_2\to S_1}{101}$$

状态转移概率矩阵

$$P(S_j|S_i)=\begin{array}{c c} & \begin{array}{cccc} 00 & 01 & 10 & 11 \end{array} \\ \begin{array}{c} S_0=00 \\ S_1=01 \\ S_2=10 \\ S_3=11 \end{array} & \begin{bmatrix} 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.8 & 0.2 \end{bmatrix} \end{array}$$

符号条件概率矩阵

$$P(x|S_i)=\begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} S_0=00 \\ S_1=01 \\ S_2=10 \\ S_3=11 \end{array} & \begin{bmatrix} 0.8 & 0.2 \\ 0.5 & 0.5 \\ 0.5 & 0.5 \\ 0.8 & 0.2 \end{bmatrix} \end{array}$$

请注意状态转移概率矩阵与符号条件概率矩阵的不同！

当信源为一阶马尔可夫信源时，每个状态只有一个符号，此时符号条件概率矩阵与状态转移概率矩阵相等。

马尔可夫信源的信源熵

一般的马尔可夫信源是非平稳的，但当**时齐、遍历的马尔可夫信源**达到稳态分布以后，可以用有限记忆长度的平稳信源的有关结论对其进行分析。

设有一个 m 阶时齐、遍历的马尔可夫信源最终能达到平稳，则可利用有限记忆长度的离散平稳信源的极限熵表征此信源。

$$H_\infty = H_{m+1} = H(X_{m+1} | X_1X_2\cdots X_m) = H(X_{m+1} | S) = \sum_i p(S_i)H(X | S_i)$$

$p(S_i)$ 极限稳态概率分布 $H(X | S_i)$ 符号条件熵

马尔可夫信源的分析思路

1 用条件概率描述，一步状态转移矩阵，符号条件矩阵、状态转移图等表征马尔可夫信源。

2 **根据一步状态转移矩阵判断极限分布是否存在**（很关键！）

3 根据一步状态转移矩阵求解极限分布 $\begin{cases} WP = W \\ \sum_j W_j = 1 \end{cases}$

4 根据符号条件矩阵求解极限熵 $H_\infty = H_{m+1} = \sum_i p(S_i)H(X|S_i)$（使用符号条件熵和稳态分布）

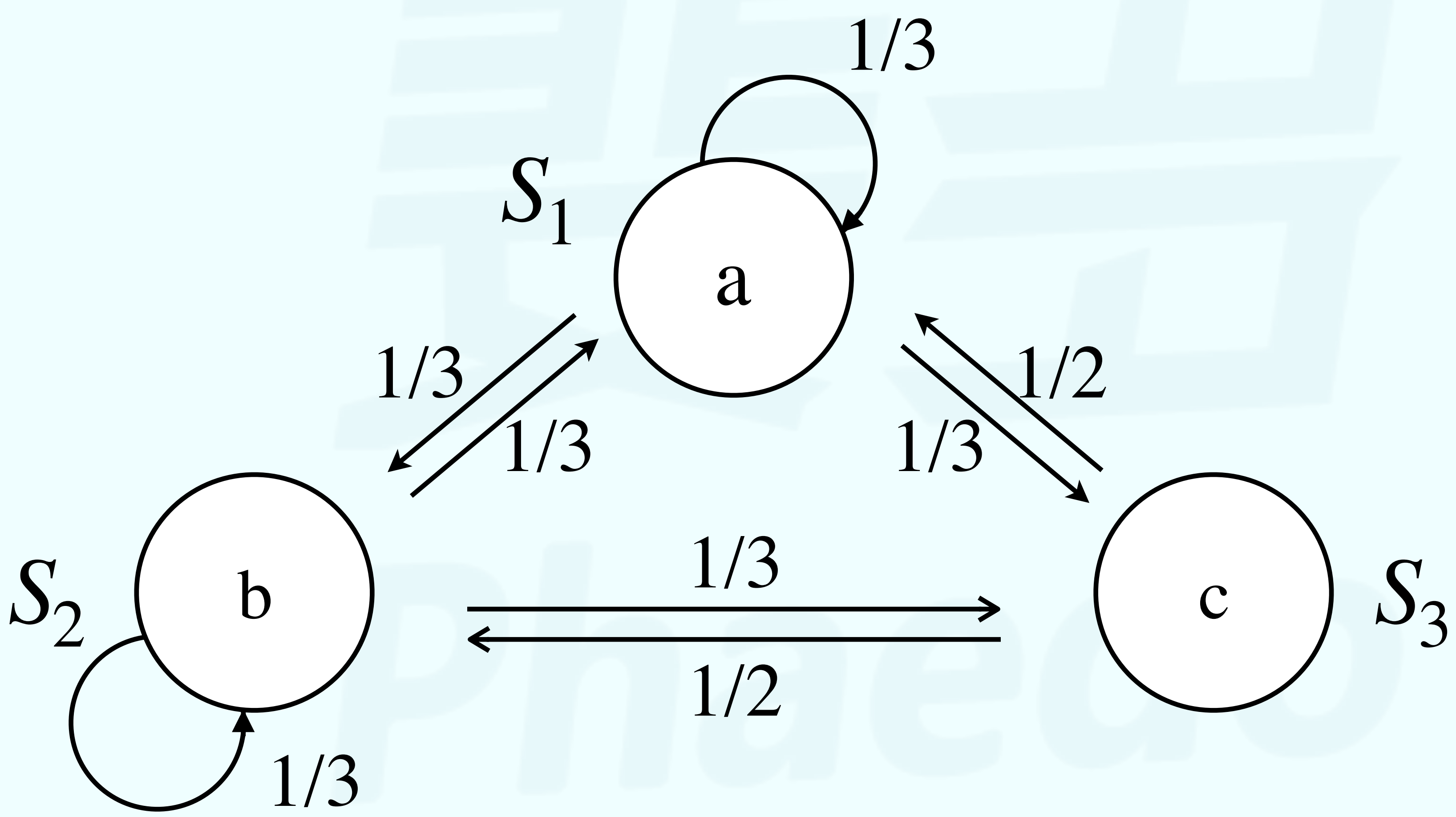
5 根据符号条件矩阵求解符号平稳分布 $P_X = WP_{X|S}$（使用符号转移矩阵）

6 求解信源的冗余度 $R = 1 - \dfrac{H_\infty}{H_0} = 1 - \dfrac{H_{m+1}}{H_0}$

研究一阶时齐马尔可夫信源时，符号条件概率矩阵与状态转移概率矩阵相等。

例题2-10 设有一个信源，它在开始时以$P(a)=0.6$，$P(b)=0.3$，$P(c)=0.1$的概率输出 X_1，如果X_1为a，则X_2为a , b , c 的概率为1/3；如果X_1为b，则X_2为a , b , c的概率为 1/3；如果X_1为c，则X_2为a, b的概率为1/2，为c的概率为0，而且后面输出 X_i 的概率只与 $X_{i\text{-}1}$ 有关，且 $P(X_i \mid X_{i\text{-}1})=P(X_2 \mid X_1)$，$i \ge 3$，试作出此信源的状态转移图，并求解其极限熵。

解析2-10 依题意得，该信源为一阶时齐马尔可夫信源，状态转移图如下图所示，其状态转移矩阵（暨符号条件矩阵）如下：



$$P=\begin{array}{c}\\ S_1=a\\ S_2=b\\ S_3=c\end{array}\begin{array}{c}\begin{array}{ccc}a & b & c\end{array}\\ \begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ \frac{1}{2} & \frac{1}{2} & 0\end{bmatrix}\end{array}$$

P^2中各元素都大于零，因此稳态分布存在

例题2-10 设有一个信源，它在开始时以$P(a)=0.6$，$P(b)=0.3$，$P(c)=0.1$的概率输出 X_1，如果X_1为a，则X_2为a，b，c 的概率为1/3；如果X_1为b，则X_2为a，b，c的概率为 1/3；如果X_1为c，则X_2为a, b的概率为1/2，为c的概率为0，而且后面输出 X_i 的概率只与 X_{i-1} 有关，且 $P(X_i \mid X_{i-1})=P(X_2 \mid X_1)$，$i\ge 3$，试作出此信源的状态转移图，并求解其极限熵。

解析2-10

$$P=\begin{array}{c} \\ S_1=a \\ S_2=b \\ S_3=c \end{array}\begin{array}{c} \begin{array}{ccc} a & b & c \end{array} \\ \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \end{array}$$

求解稳态分布：

$$\begin{cases} [W_1 \quad W_2 \quad W_3]\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}=[W_1 \quad W_2 \quad W_3] \\ W_1+W_2+W_3=1 \end{cases} \Rightarrow [W_1 \quad W_2 \quad W_3]=\begin{bmatrix} \frac{3}{8} & \frac{3}{8} & \frac{1}{4} \end{bmatrix}$$

则该信源的极限熵

$$H_\infty=H_2=H(X_{m+1}\mid X_m)=H(X_{m+1}\mid S)=\sum_i p(S_i)H(X\mid S_i)$$

$$=\frac{3}{8}H(\frac{1}{3},\frac{1}{3},\frac{1}{3})+\frac{3}{8}H(\frac{1}{3},\frac{1}{3},\frac{1}{3})+\frac{1}{4}H(\frac{1}{2},\frac{1}{2})$$

$$=1.44bit/symbol$$

例题2-11 设有一个二阶时齐的马尔可夫信源，其状态集为$S=\{00,01,10,11\}$，其符号转移条件概率如下表所示：

(1) 判断稳态分布是否存在；

(2) 若稳态分布存在，求状态的稳态分布；

(3) 求符号的稳态分布；

(4) 求此信源的极限熵。

起始状态	符号	
	0	1
00	1/2	1/2
01	1/3	2/3
10	1/4	3/4
11	1/5	4/5

解析2-11 (1) 二阶马尔可夫信源状态转移矩阵与符号条件矩阵不同

状态转移概率如下表，由此可以写出状态转移矩阵

起始状态	终止状态			
	00	01	10	11
00	1/2	1/2	0	0
01	0	0	1/3	2/3
10	1/4	3/4	0	0
11	0	0	1/5	4/5

$$\Rightarrow \quad P=\begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 1/4 & 3/4 & 0 & 0 \\ 0 & 0 & 1/5 & 4/5 \end{bmatrix}$$

P^2中各元素都大于零，因此稳态分布存在

例题2-11 设有一个二阶时齐的马尔可夫信源，其状态集为$S=\{00,01,10,11\}$，其符号转移条件概率如下表所示：

(1) 判断稳态分布是否存在；

(2) 若稳态分布存在，求状态的稳态分布；

(3) 求符号的稳态分布；

(4) 求此信源的极限熵。

起始状态	符号	
	0	1
00	1/2	1/2
01	1/3	2/3
10	1/4	3/4
11	1/5	4/5

解析2-11 (2)

$$\begin{cases} [W_1 \quad W_2 \quad W_3 \quad W_4]\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{5} & \frac{4}{5} \end{bmatrix} = [W_1 \quad W_2 \quad W_3 \quad W_4] \\ W_1+W_2+W_3+W_4=1 \end{cases}$$

解得稳态分布为
$$\begin{cases} W_1=p(00)=\frac{3}{35} \\ W_2=p(01)=\frac{6}{35} \\ W_3=p(10)=\frac{6}{35} \\ W_4=p(11)=\frac{4}{7} \end{cases}$$

例题2-11 设有一个二阶时齐的马尔可夫信源，其状态集为 $S=\{00,01,10,11\}$，其符号转移条件概率如下表所示：

(1) 判断稳态分布是否存在；

(2) 若稳态分布存在，求状态的稳态分布；

(3) 求符号的稳态分布；

(4) 求此信源的极限熵。

起始状态	符号	
	0	1
00	1/2	1/2
01	1/3	2/3
10	1/4	3/4
11	1/5	4/5

解析2-11 (3) 根据符号转移矩阵概率可以写出符号转移矩阵

起始状态	符号	
	0	1
00	1/2	1/2
01	1/3	2/3
10	1/4	3/4
11	1/5	4/5

$$\Rightarrow \quad P_{a|S}=\begin{array}{c} \\ S_0=00 \\ S_1=01 \\ S_2=10 \\ S_3=11 \end{array}\begin{array}{c} a=0 \quad a=1 \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \\ \frac{1}{4} & \frac{3}{4} \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix} \end{array}$$

解得符号的稳态分布为 $P_a=WP_{a|S}=[3/35 \quad 6/35 \quad 6/35 \quad 4/7]\begin{bmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \\ 1/4 & 3/4 \\ 1/2 & 4/5 \end{bmatrix}=\begin{bmatrix} \frac{9}{35} & \frac{26}{35} \end{bmatrix}$

例题2-11 设有一个二阶时齐的马尔可夫信源，其状态集为$S=\{00,01,10,11\}$，其符号转移条件概率如下表所示：

(1) 判断稳态分布是否存在；

(2) 若稳态分布存在，求状态的稳态分布；

(3) 求符号的稳态分布；

(4) 求此信源的极限熵。

起始状态	符号 0	符号 1
00	1/2	1/2
01	1/3	2/3
10	1/4	3/4
11	1/5	4/5

解析2-11 (4)

已知 $\begin{cases} W_1=p(00)=\frac{3}{35} \\ W_2=p(01)=\frac{6}{35} \\ W_3=p(10)=\frac{6}{35} \\ W_4=p(11)=\frac{4}{7} \end{cases}$

$$P_{a|S}=\begin{array}{c} \\ S_0=00 \\ S_1=01 \\ S_2=10 \\ S_3=11 \end{array}\begin{array}{c} \begin{array}{cc} a=0 & a=1 \end{array} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \\ \frac{1}{4} & \frac{3}{4} \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix} \end{array}$$

则信源的极限熵

$$H_\infty=H_3=H(X_{m+1}|X_{m-1}X_m)$$

$$=H(X_{m+1}|S)=\sum_i p(S_i)H(a|S_i)$$

$$=\frac{3}{35}H(\frac{1}{2},\frac{1}{2})+\frac{6}{35}H(\frac{1}{3},\frac{2}{3})+\frac{6}{35}H(\frac{1}{4},\frac{3}{4})+\frac{4}{7}H(\frac{1}{5},\frac{4}{5})$$

$$=0.79bit/symbol$$

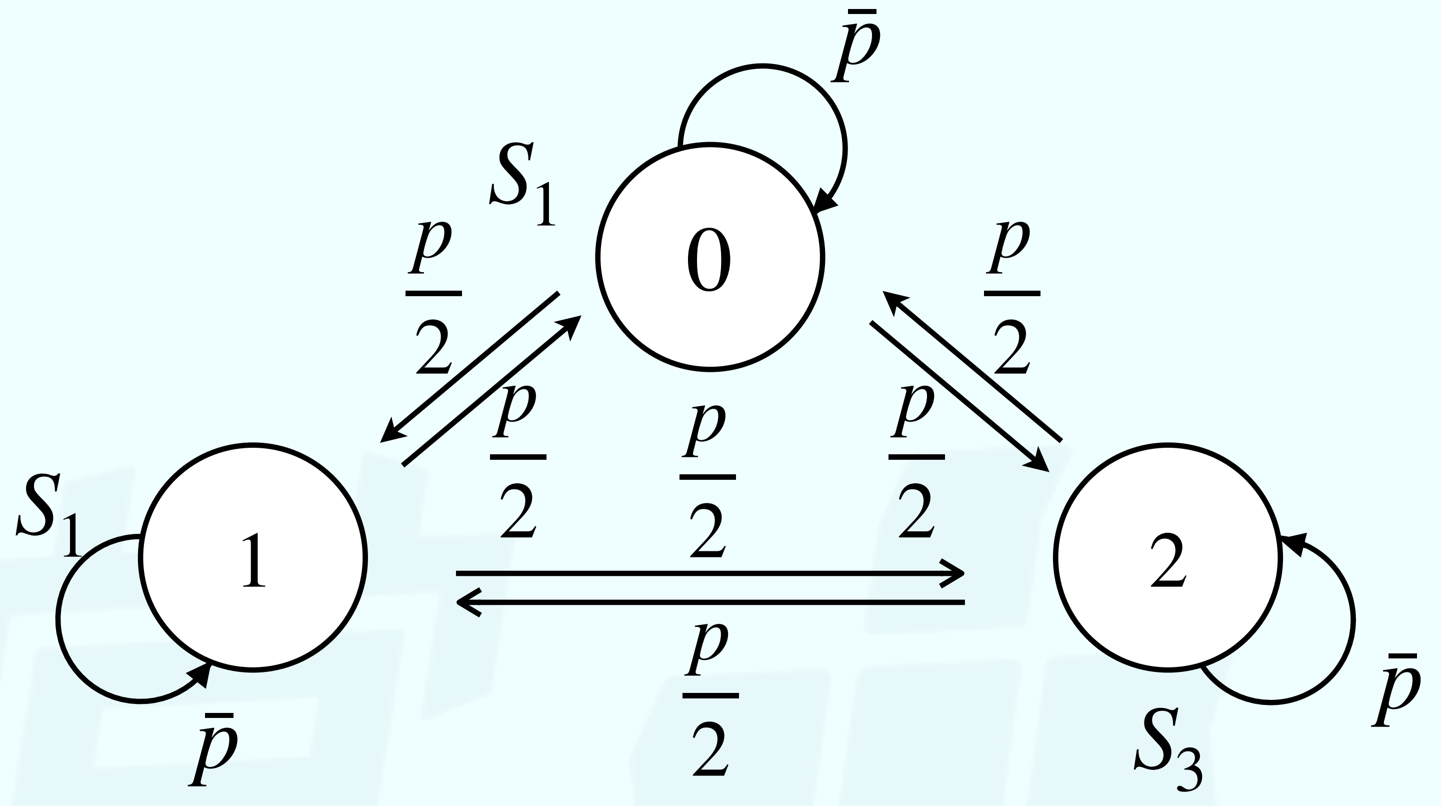
例题2-12 某一阶马尔可夫信源的状态转移图如图，已知信源X的符号集为$\{0,1,2\}$，并定义$\bar{p}=1-p$ 。

(1) 试求信源的极限概率分布$P(0)$，$P(1)$，$P(2)$；

(2) 求此信源的极限熵；

(3) 对一阶马尔可夫信源，p取何值时极限熵最大?

并计算当p取0和1时，极限熵的值。



解析2-12 (1)

状态转移矩阵为 $\begin{bmatrix} \bar{p} & \frac{p}{2} & \frac{p}{2} \\ \frac{p}{2} & \bar{p} & \frac{p}{2} \\ \frac{p}{2} & \frac{p}{2} & \bar{p} \end{bmatrix}$，为了判断各态历经性，需要对参数 p 进行分类讨论：

当$0<p<1$时，P中各元素均大于零，故极限分布存在；

当$p=0$时，P为单位矩阵，其正整数次幂也为单位矩阵无法满足所有元素大于零，故此时极限分布不存在；

当$p=1$时，P^2 中各元素大于零，极限分布存在；

因此当$0<p\le 1$时，极限分布存在。

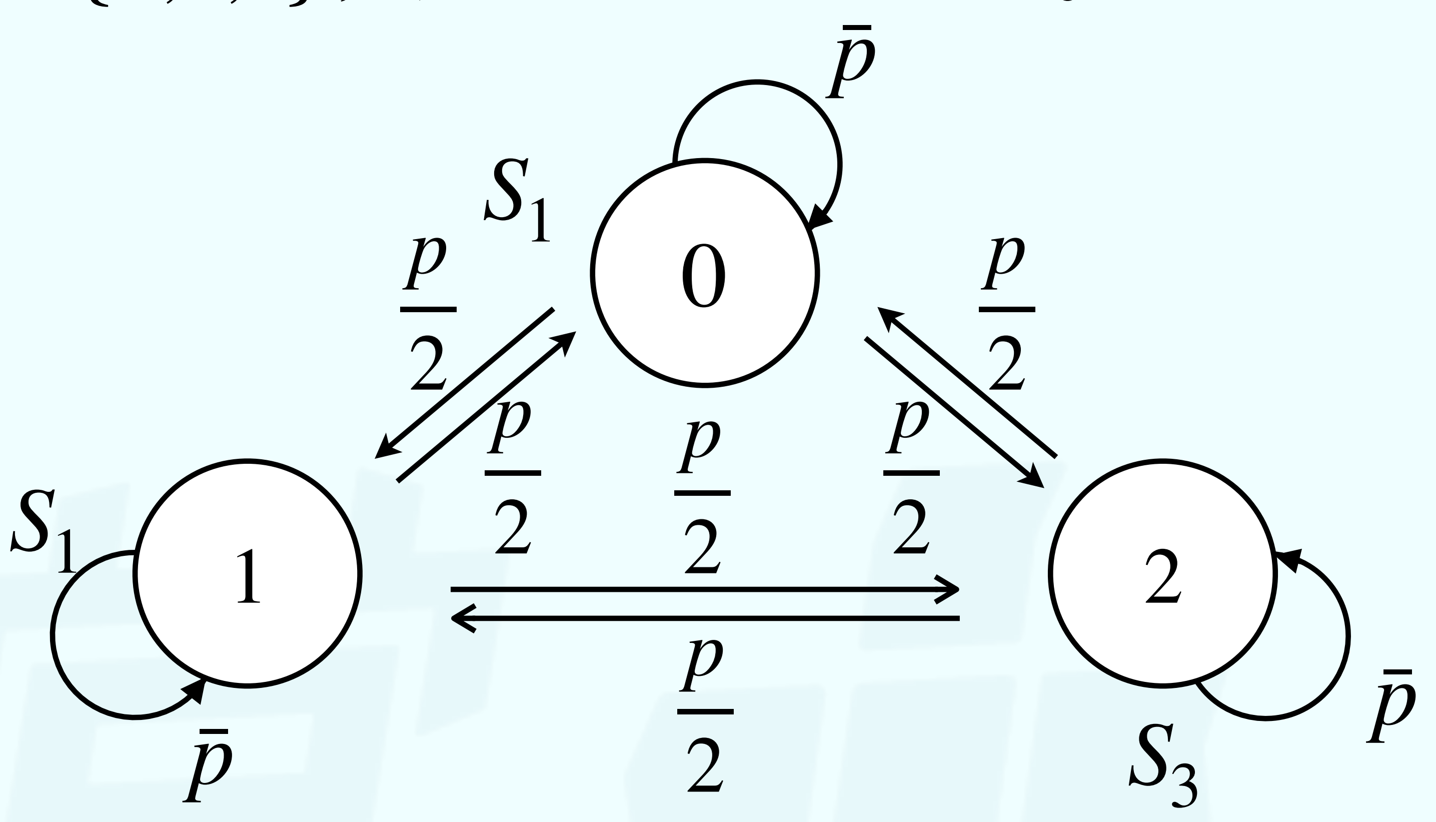
例题2-12 某一阶马尔可夫信源的状态转移图如图，已知信源X的符号集为$\{0,1,2\}$，并定义$\bar{p}=1-p$ 。

(1) 试求信源的极限概率分布$P(0)$，$P(1)$，$P(2)$；

(2) 求此信源的极限熵；

(3) 对一阶马尔可夫信源，p取何值时极限熵最大?

并计算当p取0和1时，极限熵的值。



解析2-12 (1)

$$\begin{cases} \begin{bmatrix} W_1 & W_2 & W_3 \end{bmatrix} \begin{bmatrix} \bar{p} & \frac{p}{2} & \frac{p}{2} \\ \frac{p}{2} & \bar{p} & \frac{p}{2} \\ \frac{p}{2} & \frac{p}{2} & \bar{p} \end{bmatrix} = \begin{bmatrix} W_1 & W_2 & W_3 \end{bmatrix} \\ W_1 + W_2 + W_3 = 1,\ \bar{p} = 1 - p \end{cases} \Rightarrow \begin{cases} W_1 = \frac{1}{3} \\ W_2 = \frac{1}{3} \\ W_3 = \frac{1}{3} \end{cases}$$

因此极限分布为 $P(0)=P(1)=P(2)=\dfrac{1}{3}$ 。

(2) 信源的极限熵为

$$H_\infty = H_2 = H(X_{m+1}|X_m) = H(X_{m+1}|S) = \sum_i p(S_i)H(a|S_i)$$

$$= \frac{1}{3}H(\bar{p},\frac{p}{2},\frac{p}{2}) \times 3 \ = (-\bar{p}\log\bar{p} - p\log\frac{p}{2}) \ = (-\bar{p}\log\bar{p} - p\log p + p) bit/symbol$$

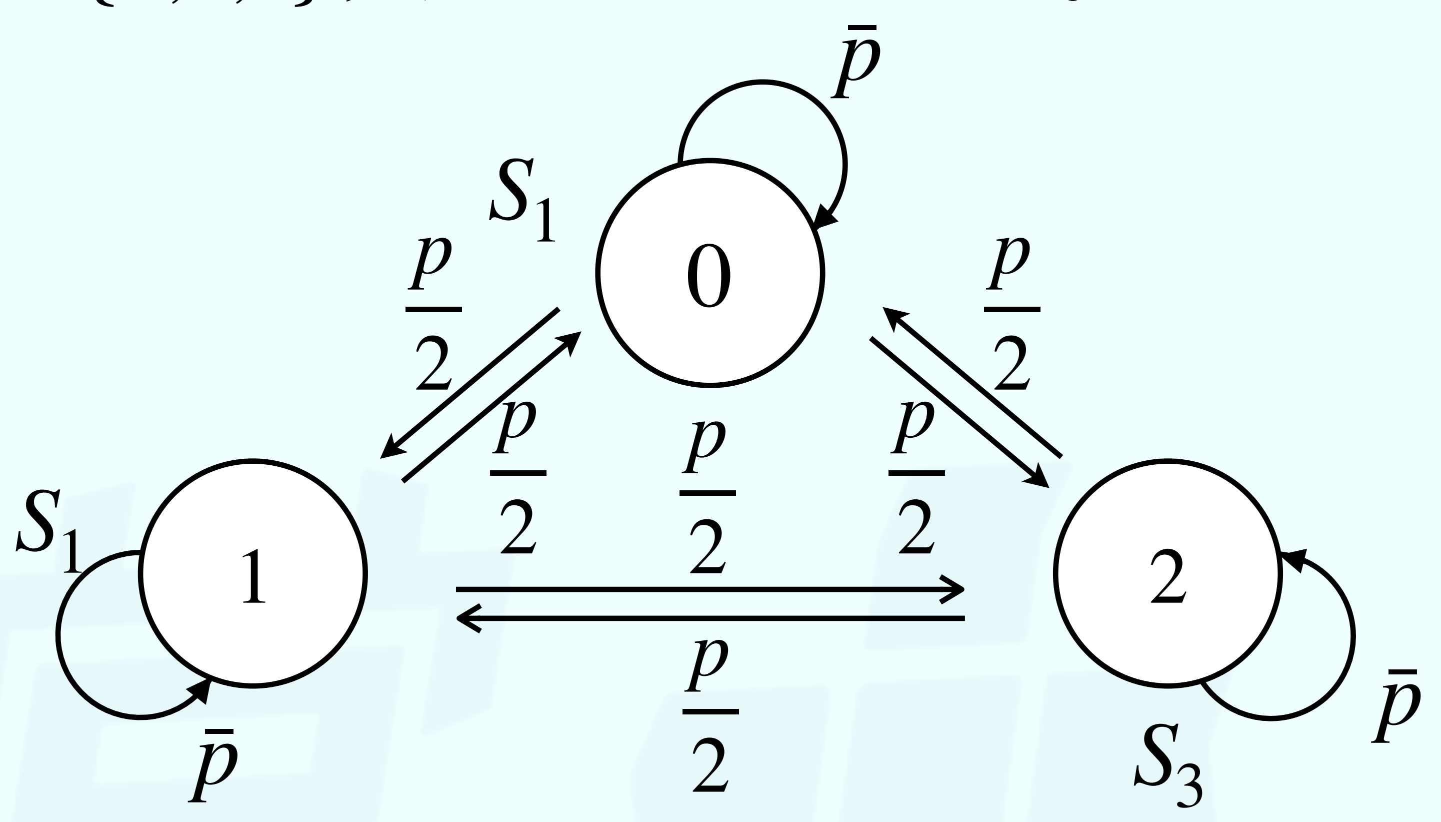
例题2-12 某一阶马尔可夫信源的状态转移图如图，已知信源X的符号集为$\{0,1,2\}$，并定义$\bar{p}=1-p$ 。

(1) 试求信源的极限概率分布$P(0)$，$P(1)$，$P(2)$；

(2) 求此信源的极限熵；

(3) 对一阶马尔可夫信源，p取何值时极限熵最大?

并计算当p取0和1时，极限熵的值。



解析2-12

(3) $$\frac{dH_\infty}{dp}=\frac{d}{dp}[-(1-p)\log(1-p)-p\log p+p]$$

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

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

当 $\log\frac{2(1-p)}{p}=0,\ p=\frac{2}{3}$ 时，极限熵最大 $(H_\infty)_{max}=\log 3\approx 1.59bit/symbol$

当 p=0 时，极限熵为各符号初始概率分布的熵；

当 p=1 时，极限熵为 1 $bit/symbol$。

信息冗余度

信息冗余度

小节1 信源的相关性与冗余度

小节2 自然语言的熵

信源的相关性与冗余度

- **信源的相关性** 信源输出符号之间的依赖程度

熟记几个离散平稳信源的符号及其熵（设信源有q 个符号）

$H_0 = \log q$	独立等概信源
$H_1 = H(X)$	一维平稳信源/独立同分布信源
$H_2 = H(X_2 \mid X_1)$	二维平稳信源
$H_{m+1} = H(X_{m+1} \mid X_1X_2 \cdots X_m)$	m+1 维平稳信源
$H_\infty = \lim\limits_{n\to\infty} H(X_{m+1} \mid X_1X_2 \cdots X_m)$	记忆长度无限的平稳信源

$H_0 \geq H_1 \geq H_2 \geq \cdots \geq H_{m+1} \geq \cdots \geq H_\infty$ 符号间相关性越大，信息熵越小

信源的相关性与冗余度

- **信源的冗余度** $R = 1 - \dfrac{H_\infty}{H_0}$

熵的相对率 $\eta = \dfrac{H_\infty}{H_0}, \quad R = 1 - \eta$

对一般离散平稳信源，H_∞ 就是实际信源熵，它本身就可以传输信源包含的所有信息，但由于记忆长度为无限的信源熵难以确定，因此需要用有限记忆长度的信源熵代替，必然会使熵增大，从而在传输中产生浪费

信息冗余度

小节1 信源的相关性与冗余度

小节2 自然语言的熵

自然语言的熵——以英语信源为例

- **英语信源的熵**

英语信源由26个字母和1个空格符号组成。

假设27个符号等概分布，则信源熵 $H_0 = \log 27 = 4.76 bit/symbol$ 。

实际情况下，绝非等概出现，各符号出现的概率统计如下

符号	概率	符号	概率	符号	概率
空格	0.2	S	0.052	Y,W	0.012
E	0.105	H	0.047	G	0.011
T	0.072	D	0.035	B	0.0105
O	0.0654	L	0.029	V	0.008
A	0.063	C	0.023	K	0.003
N	0.059	F,U	0.0225	X	0.002
I	0.055	M	0.021	J,Q	0.001
R	0.054	P	0.0175	Z	0.001

若将英语信源视为离散无记忆信源，

即不考虑符号之间的关系，则有：

$$H_1 = -\sum_{i=1}^{27} p_i \log p_i = 4.03 bit/symbol$$

虽然输出序列是英语字母和空格，但由于没有考虑符号之间的关联性，序列很难构成英语。

自然语言的熵——以英语信源为例

- **英语信源的熵**

如果按照离散无记忆信源输出英语符号，即相当于按照上表中的概率分布随机选择英语字母并将其排列，可得到类似下面的输出序列：

AI_NGAE_ITE_NNR_ASAEV_OTE_BAINTHA_HYROO_PORE_SETRYGAIETR WCO_EHDUARU_EUEU_C_FT_NSREM_DIY_EESE_F_O_SRIS_R_UNNASHOR ...

这个序列看起来是由英语信源的符号组成的，但并不是英语。

在实际中，英语的某个字母出现后，其后面的字母并为完全按照概率分布随机出现，而是会按照满足一定关系的条件概率分布出现，例如T后面出现h的概率较高，出现n的概率较小。

也就是说，英语字母之间具有强烈的依赖关系，上述序列只考虑了字母出现的概率，而没有考虑字母之间的依赖关系。

自然语言的熵——以英语信源为例

- **英语信源的熵**

若将英语信源视为一阶、二阶、…、无穷阶马尔可夫信源， 即考虑符号之间的关联性，且关联长度一直伸向无穷远，则有：

$$H_2 = 3.32bit/symbol$$

$$H_3 = 3.1bit/symbol$$

$$\cdots\cdots$$

$$H_\infty = 1.4bit/symbol$$

随着关联性这一因素的加入，输出序列会有像样的英语单词出现，有依赖关系的字母数越高，即马尔可夫信源的阶数越高，输出序列就会越趋近于实际情况。若依赖关系延伸到无穷远，则信源就会输出真正的英语，此时极限熵为1.4 *bit*/符号。

自然语言的熵——以英语信源为例

- **英语信源的熵**

例如，把英语信源近似为二阶马尔可夫信源，则可以得到类似下面的输出序列

IANKS_CAN_OU_ANG_RLER_THTTED_OF_TO_SHOR_OF_TO_HAVE
MEM_A_I_MAND_AND_BUT_WHISS_ITABLY_THERVEREER...

二阶马尔可夫信源的记忆长度为3，因此，这个序列中被空格分开的两字母组或三字母组，组成的大多都是有意义的英文单词，而4个以及以上的字母组，就很难查出对应的单词了。

因为该序列仅仅考虑了3个及以下字母之间的依赖关系，而实际英语字母之间的关系可以延伸到更多符号，且单词与单词之间也存在依赖关系

我们不难发现，有依赖的字母数越多，即马尔可夫信源阶次越高，那么输出序列就越接近实际的英语，当依赖关系延伸至无穷远时，信源输出的就是真正的英语，此时极限熵为$1.4 bit/symbol$ 。

自然语言的冗余度——以英语信源为例

- **英语信源的冗余度**

$$R = 1 - \frac{H_\infty}{H_0} = 1 - \frac{1.4}{4.76} = 0.71$$

写英语文章时，71%是由语言结构定好的，只有29%是写文字的人可以自由选择的；

100 页的书，大约只传输29 页就可以了，其余71 页可以压缩掉。

信息的冗余度表示信源可压缩的程度。相关性越大，信息熵越小，冗余度越大；

从提高传输效率的观点出发，总是希望减少或去掉冗余度；

但冗余度大的消息抗干扰能力强，能通过前后符号间的关联关系纠正错误。