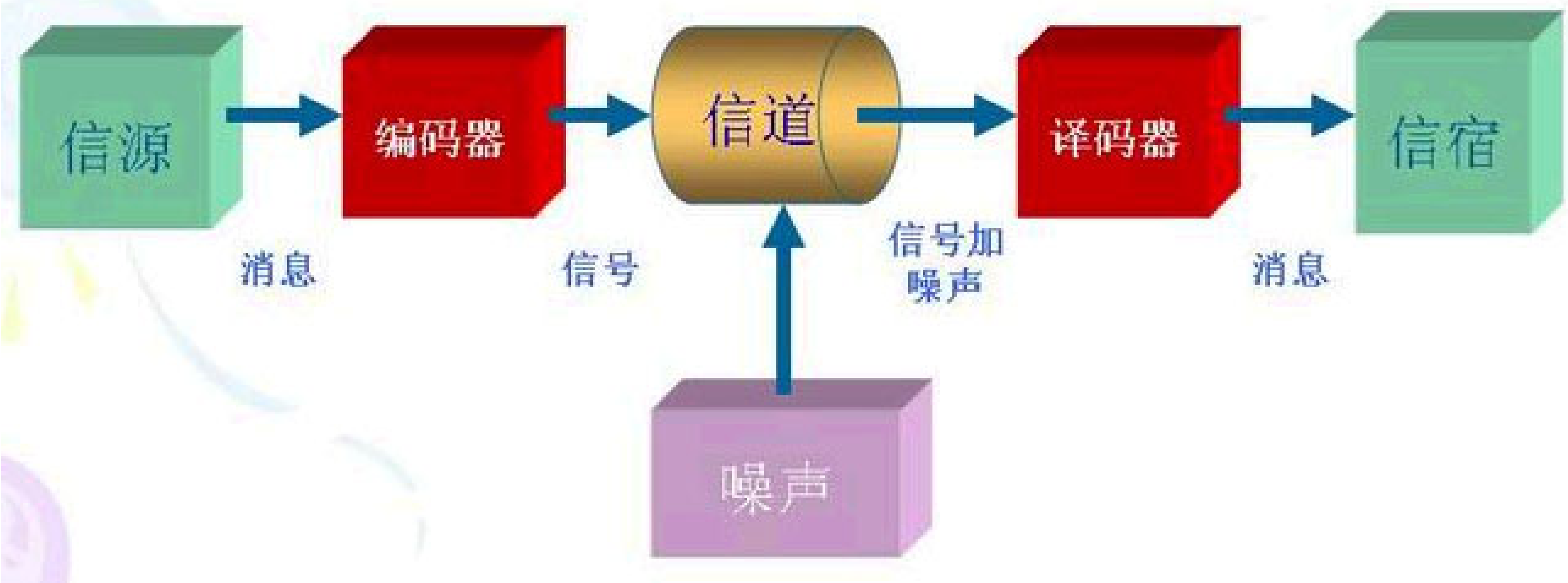
第一章 要点复习

- 什么是信息？
  - 是事物运动状态或存在方式下不确定性的描述
  - 如何理解“不确定性”？
- 通信系统的基本模型



概率论知识的回顾 －书写要求及格式

- 已知随机变量的联合概率分布 $p(x_i y_j)$ 如下所示，求条件概率 $p(y_j|x_i)$ 。

$p(x_i y_j)$		y_j		
		0	**1**	**2**
x_i	**0**	**1/8**	**0**	**1/8**
	1	**1/8**	**1/8**	**0**
	2	**0**	**1/4**	**1/4**

练习（1）

- 某年级有甲，乙，丙3个班级，各班人数分别占年级总人数的1/4，1/3，5/12。已知甲，乙，丙3个班级中集邮人数分别占该班总人数的1/2，1/4，1/5

（1）事件“某人为集邮者”的概率；

（2）事件“某人既为集邮者，又属于乙班”的概率。

（3）事件“某集邮者是乙班的”的概率.

练习（2）

- 某校入学考试中有25%考生被录取。被录取的考生中有50%来自本市，而落榜考生中有25%来自本市。所有本市的考生都学过英语，外地考生中录取和未录取的都只有50%学过英语。问：

（1）“某学生学过英语”这一事件发生的概率

（2）“学过英语的某学生被录取”这一事件发生的概率

（3）考生地域状态已知时英语学习状态的概率分布。

练习（3）

- 设有来自两个地区的考生报名表分别是10份和15份，其中女生的报名表分别是3份和7份。随机地取一个地区的报名表，从中先后抽出两份（不放回）,问：

（1）“先抽到的一份是女生表”的概率

（2）“后抽到的一份是男生表”的概率

（3）“先抽到的一份是女生表，后抽到的一份是男生表”的概率

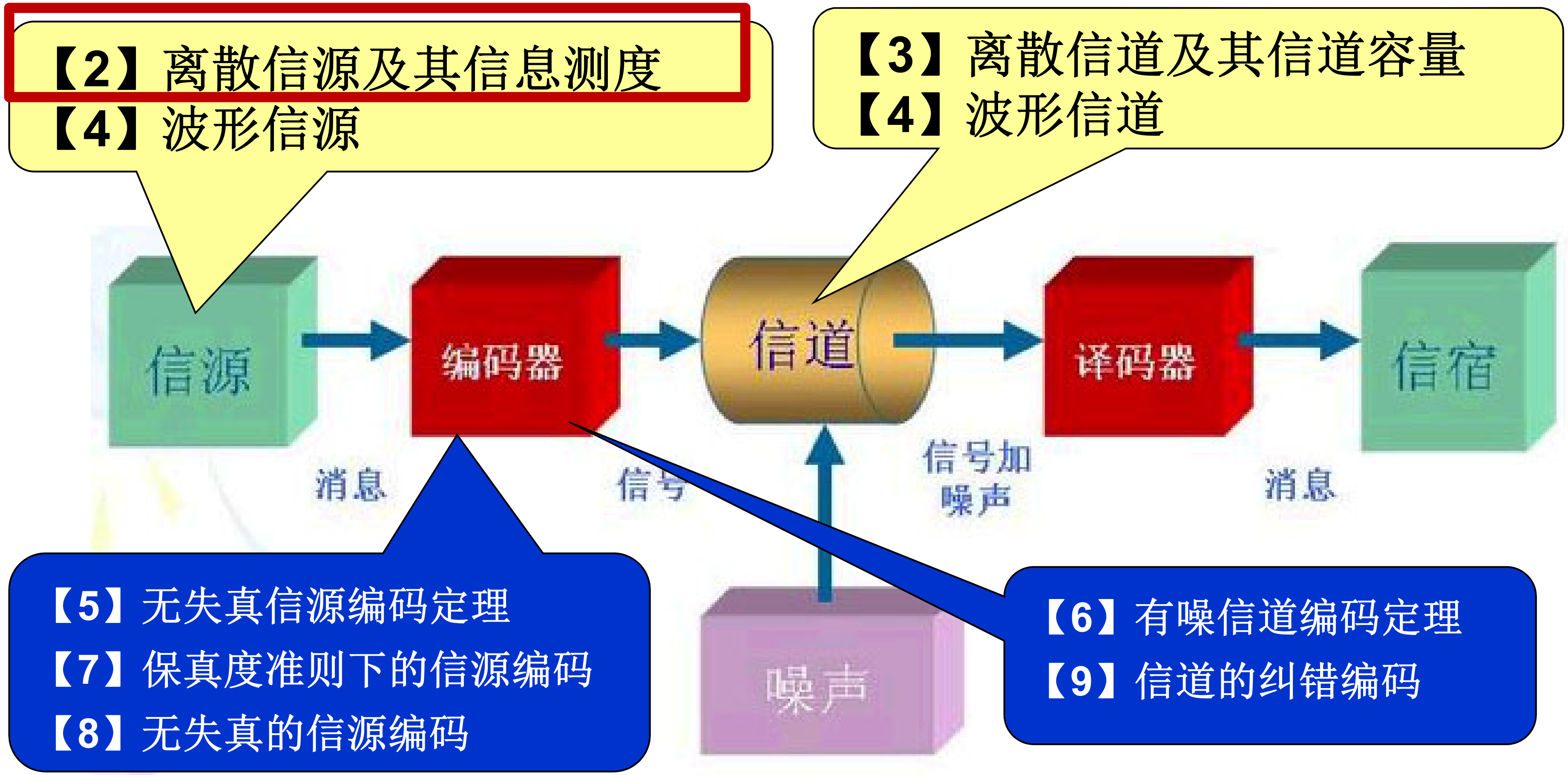
（4）“后抽到的一份是男生表的条件下，先抽到的一份是女生表”的概率

第二章　　离散信源及其信息测度

数智学院 孟放

信息论的旅程

- 从本章开始，我们将从有效而可靠地传输信息的观点出发，对组成信息传输系统的各个部分进行讨论。



本章的研究内容

- 什么是信源？
  - 信息的来源，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 。
  - 数学上，信源是产生随机变量，随机序列/矢量和随机过程的源。
- 信源的研究内容 ★
  - 只研究信源产生消息的不确定性！
  1. 信源的输出如何描述？
  2. 如何计算信源产生的信息量？
  3. 能否压缩？可压缩程度？

基本概念

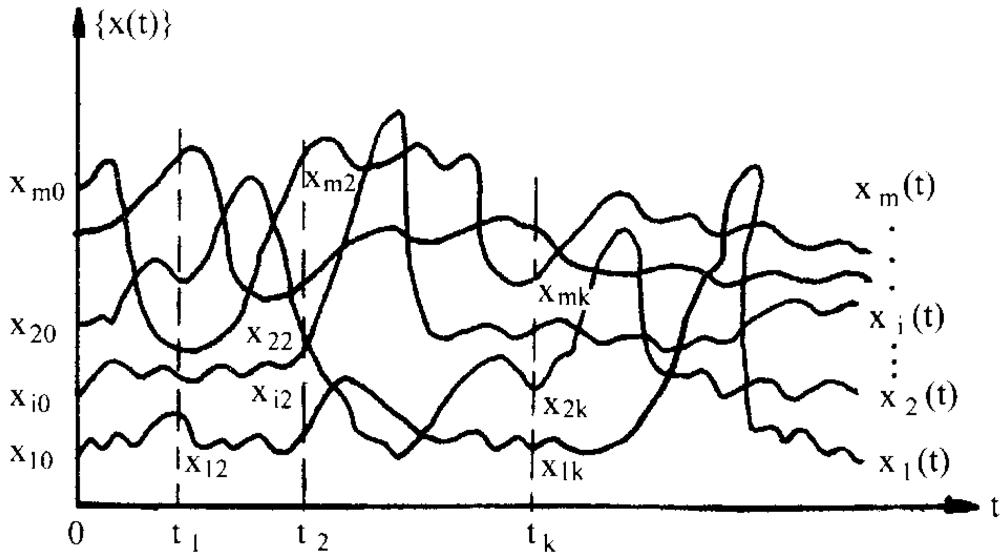
- 信源分类
  - 连续信源和离散信源 --- 本章只研究离散信源
  - 有记忆信源和无记忆信源
    - 记忆长度
  - 平稳信源和非平稳信源
- 信源的分析思路
  - 实际生活中信源：各种携带信息的消息表示形式
    - 在一个固定的时刻，信源发出的是一个随机变量。
    - 随着时间的延续，信源发出的是一个随机过程。
  - 如何分析信源的属性 --- 惯用思路
    - 简单 --〉具有一定实用性的复杂信源

实际常见信源

- 用随机过程描述信源输出的消息
  - 实际信源输出的消息常常是时间和取值都是连续的
    - 用随机过程$\{x(t)\}$来描述，即随机波形信源

如：语音信号，

模拟电视图像信号



一个随机过程$\{x(t)\}$

★信源的分类图

随机过程{$x(t)$} ⟷ 随机波形信源：信源输出消息在时间/空间上和取值上都是连续的函数

↓ 取样定理

- 随机序列X
  - 平稳信源
    - 连续平稳信源
    - ↓ 取值量化
    - 离散平稳信源
      - 无记忆扩展
      - 有限记忆
  - 非平稳信源 → 马尔可夫信源

- 随机变量X
  - 连续单符号信源
  - ↓ 取值量化
  - 离散单符号信源

本章主要内容

- 信源的数学模型及分类
- 离散信源的信息熵
- 熵的基本性质
- 离散无记忆扩展信源
- 离散平稳信源
- 马尔可夫信源
- 信源冗余度与自然语言的熵
- 加权熵

数学模型

信源分类

1.1、信源的数学模型 － 离散信源 - 随机变量

- 离散信源的特点
  - 输出消息在时间和幅值上都是离散的
  - 消息数目有限或可数无穷
  - 用离散随机变量来描述
- 例如：文字、数字、事件等

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & \cdots & a_i & \cdots & a_q \\ p(a_1) & \cdots & p(a_i) & \cdots & p(a_q) \end{bmatrix}$$

$$0 \le p(a_i) \le 1, (i = 1, 2, \cdots, q)$$

$$\sum_{i=1}^{q} p(a_i) = 1$$

1.1、信源的数学模型 － 连续信源

- 特点
  - 连续信号
  - 消息数目不可数
    - 输出消息的取值是连续的
  - 用连续随机变量来描述
- 语音、模拟信号等

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} (a,b) \\ p(x) \end{bmatrix}$$

$$p(x) \geq 0$$

$$\int_a^b p(x)dx = 1$$

$p(x)$是随机变量x的概率密度函数

1.1、信源的数学模型 – 随机矢量

- 有限离散信源（N是有限的）
  - 输出为多个消息符号 – 单个符号无意义
    - 自然语言、图像、……
  - 用随机矢量来描述： $\mathbf{X}=(\boldsymbol{X}_1\boldsymbol{X}_2\ldots\boldsymbol{X}_\mathrm{N})$

$$\begin{bmatrix}\mathbf{X}\\ P\end{bmatrix}=\begin{bmatrix}\alpha_1 & \cdots & \alpha_i & \cdots & \alpha_{q^N}\\ p(\alpha_1) & \cdots & p(\alpha_i) & \cdots & p(\alpha_{q^N})\end{bmatrix}$$

注：α_i 为N维矢量，其中每个分量来自随机变量$\boldsymbol{X}_i$ 的取值空间

$$X=[a_1,a_2,\cdots,a_q]$$

该信源的输出消息个数 q^N

1.1、随机矢量所表示的信源 － 平稳信源

对于信源输出的随机序列 $\mathbf{X}=(X_1X_2\ldots X_N)$

- 离散平稳信源：随机变量X_i(i=1, 2, …, N)为取值离散的离散型随机变量，且随机矢量**X**的各维概率分布都与时间起点无关。
- 连续平稳信源
  - 每个随机变量X_i(i=1, 2, …, N)都是取值连续的连续型随机变量，且，随机矢量**X**的各维概率密度函数都与时间起点无关。
  - 每个连续取样值进行量化处理，可将连续平稳信源转换为离散信源来处理。

1.1、随机矢量所表示的信源 – 记忆信源

- 离散无记忆信源
  - 信源先后发出的符号彼此统计独立
- 离散有记忆信源
  - 符号间彼此依存、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 – 有限记忆信源
    - 特殊的离散非平稳信源 马尔可夫信源
    - 以有限状态的马尔可夫链来描述有限记忆信源

1.2、信源的分类

- 按信号取值集合和信号取值时刻的集合是“离散”还是“连续”来分类：

信号取值集合	信号取值时刻集合	信源种类
离散	离散	数字(digital) /离散(discrete)
连续	连续	模拟(analog) /波形(waveform)
连续	离散	连续(continuous)

1.2、信源的分类 － 总结

随机变量X描述信源输出的消息

- 离散信源：输出消息数为有限或可数的，每次只输出一个消息
- 分层（量化）
- 连续信源：输出消息数无限或不可数，每次只输出一个消息

随机序列X

- 非平稳信源：描述信源输出消息的随机序列是非平稳随机序列
  ——马尔可夫信源（记忆长度有限）
- 平稳信源：描述信源输出消息的随机序列X是平稳的随机序列
  - 离散平稳信源：输出随机序列X中每个随机变量X_i取值离散，且各维概率分布不随时间而改变。
    - 离散无记忆信源的N次扩展： X中各随机变量统计独立，且取自同一概率空间
    - 有限记忆信源：X中各随机变量之间有依赖关系，但记忆长度有限
  - 连续平稳信源：输出随机序列X中每个随机变量X_i取值连续，且各维概率密度函数不随时间而改变。

取样定理

随机过程$\{x(t)\}$：随机波形信源，信源输出消息是时间和取值都连续的函数

本章主要内容

- 信源的数学模型及分类
- 离散信源的信息熵
  - 自信息
  - 信息熵
- 熵的基本性质
- 离散无记忆扩展信源
- 离散平稳信源
- 马尔可夫信源
- 信源冗余度与自然语言的熵
- 加权熵

单符号无记忆离散信源

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a_1 & \cdots & a_i & \cdots & a_q \\ p(a_1) & \cdots & p(a_i) & \cdots & p(a_q) \end{bmatrix}$$

$$0 \le p(a_i) \le 1, (i = 1, 2, \cdots, q)$$

$$\sum_{i=1}^{q} p(a_i) = 1$$

自信息：信源发出的某个符号（事件），所包含/提供的信息量。

信息熵：信源平均每个符号所包含/提供的信息量，也称平均自信息。

2.1、 自信息 － 直观感觉

- 一个事件 x_i 本身所包含的信息量，由该事件的不确定性决定的。
- 自信息量 $I(x_i)$ 与该事件发生概率 $p(x_i)$ 之间所构成的函数 $f[p(x_i)]$ 应满足以下四个条件：
  - 若 $p(x_i) > p(x_j)$，则应该有 $I(x_i) < I(x_j)$；
  - 若 $p(x_i) = 0$，则 $I(x_i) = \infty$ ；
  - 若 $p(x_i) = 1$，则 $I(x_i) = 0$ ；
  - 两个独立事件的联合信息量，应等于它们各自信息量之和：$I(x_i y_j) = I(x_i) + I(y_j)$

2.1、 自信息 – 定义

- 定义：任意**随机事件**的自信息量为该事件发生概率的对数的负值

$$\boxed{I(x_i) = -\log p(x_i) = \log \frac{1}{p(x_i)}}$$

- 自信息量的单位与所用对数的底有关
  - 以2为底，单位为比特 (**bit**) – **binary unit**
  - 以*e*为底，单位为奈特 (**nat**) – **natural unit**
  - 以10为底，单位为哈特 (**hat**) – **Hartley unit**

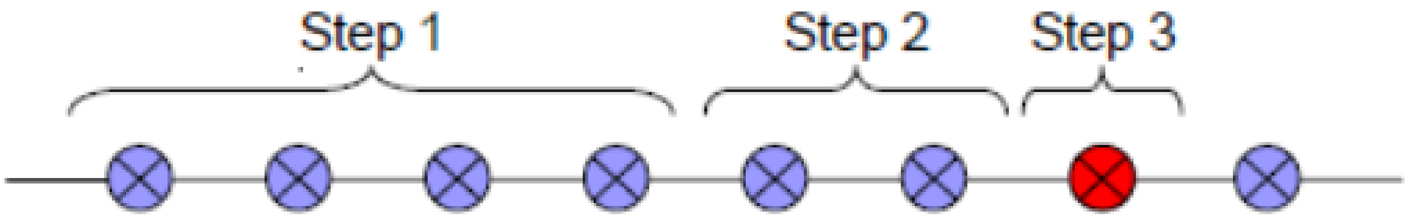
- 各单位之间的换算：多用以2为底，简写log
  - 1奈特 = $\log_2 e$ 比特 = 1.443 比特
  - 1哈特 = $\log_2 10$ 比特 = 3.322 比特
  - 若以*r*为底，则1 *r* 进制单位 = $\log_2 r$ 比特

2.1、 自信息 － 计算机中的bit单位?

- 信息论中“比特”是指抽象的信息量单位；
  - 如果 $p(x_i)=1/2$，则 $I(x_i)=1bit$ 。所以**1*bit***信息量就是两个互不相容的等可能事件之一发生时所提供的信息量。
- 计算机术语中“比特”是代表二元数字；
  - 计算机中数据存储的最小单位“位”；
  - 每个二元数字所能提供的最大平均信息量为**1**比特。

2.1、 自信息 – 分析

信息论中“比特”的意义



- 8个灯泡串联，其中一个灯丝坏了。如何用最少的步骤定位出哪个坏了？
- 最少需要用三次二元判定来定位故障。因此，这个事件所含有的信息量是3比特。
  - 足球比赛猜冠军？

2.1、 自信息 – 含义

- 从三个不同角度理解
  - 事件发生前
    - 描述该事件发生的不确定性的大小
  - 事件发生后
    - 表示该事件所含有/提供的信息量
  - 在理想信道中
    - 等于收信者接受到该消息后所获取的信息量

2.1、 自信息 － 联合自信息和条件自信息

- 定义：二维联合空间*XY*中的事件 $(x_i y_j)$的联合自信息

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

- 定义：事件 y_j给定的条件下事件 x_i 的条件自信息为

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

- 自信息、条件自信息和联合自信息：

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

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

2.1、 自信息 － 例题2.1

- 箱中有90个红球，10个白球。现从箱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”所提供的信息量；
  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？

提示：确定概率空间！

2.1、 自信息 – 例题2.1 – 求解

1. 根据应用条件**构建概率空间**

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 = \text{红} & x_2 = \text{白} \\ 0.9 & 0.1 \end{bmatrix}$$

2. 代入求解公式

$$\boxed{I(x_i) = -\log p(x_i) = \log \frac{1}{p(x_i)}}$$

(1) $I(x_1) = -\log 0.9 = 0.152 \quad bit$

(2) $I(x_2) = -\log 0.1 = 3.323 \quad bit$

3. 白球更难猜测，所含信息量更大！

2.1、 自信息 – 例题2.1（续1）

- 箱中有90个红球，10个白球。现从箱中随机地取出两个球。求：
  1. 事件“两个球中有红、白球各一个”的不确定性；
  2. 事件“两个球都是白球”所提供的信息量；
  3. 事件“两个球都是白球”和“两个球都是红球”的发生，哪个事件更难猜测？

定义概率空间：

- 假定：X_1表示第一个球的颜色，X_2表示第二个球的颜色。

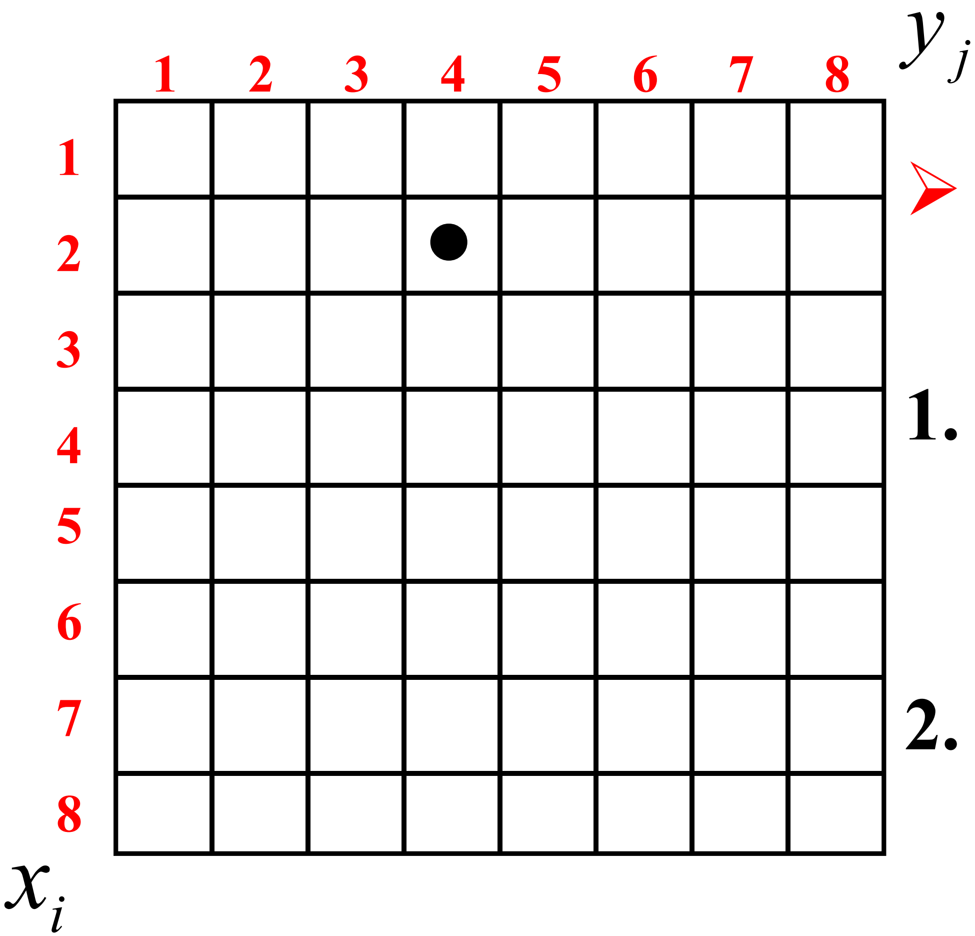
2.1、 自信息 – 例题2.1（续2）

- 箱中有90个红球，10个白球。现从箱中先拿出一球，再拿出一球，求：
  1. 事件“在第一个球是红球条件下，第二个球是白球”的不确定性；
  2. 事件“在第一个球是红球条件下，第二个球是红球”所提供的信息量。

  注意：

  - 定义概率空间：X_1和X_2分别表示第一个球和第二个球的颜色。
  - 确定求解事件对应的概率/概率空间！

2.1、 自信息 – 练习2.1



➢ 下面两种情况下的自信息：

1. 随机放棋子，棋子出现在i行j列时所提供的信息量。

2. 已知棋子的行号i，棋子出现在第j列所提供的信息量。

2.2、平均自信息 – 引入

- 自信息的取值是一个随机“变量”
  - 自信息是指信源发出的某一消息所含有的信息量；不同的消息，它们所含有的信息量也就不同。

- 平均自信息 – 统计平均量（数学期望）
  - 信息熵/信源熵/香农熵/无条件熵/熵函数/熵
  - 定义整个信源的不确定度

2.2、平均自信息 － 定义

- 随机变量X有q个可能取值，概率空间为

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & \cdots & x_i & \cdots & x_q \\ p(x_1) & \cdots & p(x_i) & \cdots & p(x_q) \end{bmatrix} \quad \boxed{\begin{array}{l} 0 \le p(x_i) \le 1, \\ \sum_{i=1}^{q} p(x_i) = 1 \end{array}}$$

- 随机变量X的平均自信息 – 统计平均值

$$H(X) = E[I(x_i)] = -\sum_{i=1}^{q} p(x_i) \log p(x_i)$$ ★

- 单位：取决于对数选取的底
  - 比特/符号、奈特/符号、哈特/符号

2.2、平均自信息 － 对于信源熵的进一步理解

- 意义：信源的信息熵是从整个信源的统计特性来考虑的。它是从平均意义上来表征信源的总体特性的。对于某特定的信源，其信息熵只有一个。不同的信源因统计特性不同，其信息熵也不同。

- *H*(*X*)三种物理含义
  - 表示信源输出前，信源的平均不确定性；
  - 是表示信源输出后，每个消息(符号)所提供的平均信息量；
  - 表示随机变量*X*的随机性；

2.2、平均自信息 - 对于熵的进一步理解（续）

- 注意：
  - 信源熵是信源的平均不确定性的描述。
  - 一般情况下，它并不等于平均获得的信息量。
  - 只有在理想情况下，接收者才能正确无误地接收到信源所发出的消息，消除了$H(X)$大小的平均不确定性。

2.2、平均自信息 – 例题2.2

- 已知：一信源X有6种输出符号，概率分别为

$$p(a)=0.5 \quad p(b)=0.25 \quad p(c)=0.125$$

$$p(d)=p(e)=0.05 \quad p(f)=0.025$$

求解 $H(X)=E[I(x_i)]=-\sum_{i=1}^{q}p(x_i)\log p(x_i)$

– 计算 $H(X)$。

– 求符号序列***ababba***和***fddfdf***的信息量，并将之与6位符号的信息量期望值相比较

2.2、平均自信息 – 例题2.2（续）

1. 由信息熵定义，该信源输出的信息熵为

$$H(X)=\sum_{i=1}^{6}p(x_i)\log\frac{1}{p(x_i)}$$

$$=0.5\log 2+0.25\log 4+0.125\log 8+2\times 0.05\log 20+0.025\log 40$$

$$=1.94 \quad \text{bit/symbol}$$

2. 符号序列***ababba***所含的信息量为

$$I_1=3I_a+3I_b=3\times\left[-\log p(a)-\log p(b)\right]$$

$$=3\times\left(\log 2+\log 4\right)=9\ \text{bit}$$

符号序列***fddfdf***所含的信息量为

$$I_2=3I_d+3I_f=3\left[-\log p(d)-\log p(f)\right]$$

$$=3\left[\log 20+\log 40\right]=28.932\ \text{bit}$$

2.2、平均自信息 – 例题2.2（续）

- 长度为6的符号序列所含信息量的平均值为

$$\bar{I} = 6H(X) = 11.64 \ bit/sym$$

$$\boxed{I_1 < \bar{I} < I_2}$$

- 分析：

$$I_1 = 3I_a + 3I_b = 3 \times [-\log p(a) - \log p(b)]$$

$$= 3 \times (\log 2 + \log 4) = 9 \ \text{bit}$$

$$I_2 = 3I_d + 3I_f = 3[-\log p(d) - \log p(f)]$$

$$= 3[\log 20 + \log 40] = 28.932 \ \text{bit}$$

2.2、平均自信息 – 练习2.2，练习2.3

- 掷一个均匀的硬币，直到出现“正面”为止，令X表示所需掷的次数，求熵$H(X)$。解答

- 电视屏上约有500*600个格点，按每点有10个不同的灰度等级考虑，设每个格点独立变化，且各灰度等级等概率出现。
  - 问：每个画面含有多少信息量？
  - 提示：
    1. 共有多少个画面？（样本个数）
    2. 每个画面的出现概率是多少？（样本对应的概率值）

本章主要内容

- 信源的数学模型及分类
- 离散信源的信息熵
- 信息熵的基本性质
- 离散无记忆扩展信源
- 离散平稳信源
- 马尔可夫信源
- 信源冗余度与自然语言的熵
- 加权熵

熵函数的定义
对称性
确定性
非负性
扩展性
强可加性 – 可加性
递增性
上凸性
极值性

3.1、信息熵的基本性质－ 熵函数的定义

- 熵函数：信息熵$H(X)$是随机变量X的概率分布函数

$$H(X) = -\sum_{i=1}^{q} p_i \log p_i = H(p_1, p_2, \cdots, p_q) = H(\mathbf{P})$$

- 当n=2时，熵函数可以表示为$H(p)$ ★

3.2、熵函数的性质 － 对称性

- 当概率矢量 $P=(p_1,p_2,\cdots,p_q)$ 中各分量的次序任意变更时，熵函数的值不变，即

$$H(p_1,p_2,\cdots,p_q)=H(p_2,p_1,\cdots,p_q)=H(p_q,p_1,\cdots,p_{q-1})$$

$$\boxed{H(X)=-\sum_{i=1}^{q}p(x_i)\log p(x_i)}$$

- 该性质说明：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。如果某些信源的统计特性相同（含有的符号数和概率分布相同），那么这些信源的熵就相同。

3.2、熵函数的性质 – 对称性 – 例题2.3

- 三个信源：

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} \text{红} & \text{黄} & \text{蓝} \\ 1/3 & 1/6 & 1/2 \end{bmatrix}$$

$$\begin{bmatrix} Y \\ P(Y) \end{bmatrix} = \begin{bmatrix} \text{红} & \text{黄} & \text{蓝} \\ 1/6 & 1/2 & 1/3 \end{bmatrix}$$

$$\begin{bmatrix} Z \\ P(Z) \end{bmatrix} = \begin{bmatrix} \text{晴} & \text{雾} & \text{雨} \\ 1/3 & 1/6 & 1/2 \end{bmatrix}$$

① ***X*与*Z*信源**：具体消息其含义不同；
② ***X*与*Y*信源**：同一消息的概率不同；
③ 但它们的信息熵是相同的。

3.2、熵函数的性质 － 对称性 －例题2.4

- *A*、*B*两个城市的天气情况分别如下：

	晴	多云	雨	冰雹
城市*A*	½	¼	1/8	1/8
城市*B*	½	1/8	1/8	1/4

? 计算所得的信息熵相同，但实际天气却相差甚大

! 信息熵未能描述事件本身的具体含义和主观价值

3.3、熵函数的性质 – 确定性

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

- 在概率空间中，只要有一个事件是必然发生事件，那么其它事件必然是不可能事件，因此该信源没有不确定性，其熵必为**0**。

3.4、熵函数的性质 – 非负性

$$H(P)=H(p_1,p_2,\cdots,p_q)\geq 0$$

$$0\leq p_i\leq 1 \qquad -\log p_i\geq 0$$

$$\boxed{H(P)=-\sum_{i=1}^{q}p_i\log p_i\geq 0}$$

1、当随机变量是一个确知量时，熵$H(X)$=0。

2、离散信源的熵满足非负性，而连续信源的熵可能为负。

3.5、熵函数的性质 – 扩展性

$$\lim_{\varepsilon->0} H_{q+1}\left(p_1, p_2, \cdots, p_q-\varepsilon, \varepsilon\right)=H_q\left(p_1, p_2, \cdots, p_q\right)$$

$$\lim_{\varepsilon->0} \varepsilon \log \varepsilon=0$$

- 扩展性说明，增加一个概率接近于零的事件，信源熵保持不变。

- 单个小概率事件具有较大信息量；
- 但从总体来考虑时，小概率事件几乎不会出现，所以它对于离散集的熵的贡献可以忽略不计。这也是熵的总体平均性的一种体现。

3.6、熵函数的性质 – 强可加性

◆ 设有两个相互关联的信源*X*和*Y*

$$\boxed{P_{XY} \qquad P_{Y|X}}$$

- 如何理解“相关关联”？
  - 袋子里装了两种颜色的球，采用不同方式取两个球。
    - 情况一：**X** — 从中随机取出一个球，看颜色，放回；**Y** — 再从中随机取出一球，看颜色。
    - 情况二：**X**—从中随机取出一个球，看颜色，不放回；**Y**—再从中随机取出一球，看颜色。

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

$$\begin{bmatrix} Y \\ P(Y) \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \cdots & y_m \\ p(y_1) & p(y_2) & \cdots & p(y_m) \end{bmatrix} \quad 0 \le p(y_j) \le 1, \quad \sum_{j=1}^{m} p(y_j) = 1$$

3.6、熵函数的性质 － 强可加性（续一）

- 互相关联的含义：
  - 信源***X***发某一符号 x_i 的前提下，信源***Y***按一定的概率发某一符号y_j，用条件概率$p\left(y_j|x_i\right)$表示。
  - ***m*n***个条件概率$p\left(y_j|x_i\right)$，且 $\sum_{j=1}^{m} p\left(y_j|x_i\right)=1$
  - 联合信源***XY***发出某一个符号 $x_i y_j$ 的概率为：

$$p\left(x_i y_j\right)=p\left(x_i\right)p\left(y_j|x_i\right)$$

$$p\left(x_i y_j\right)=p_i p_{ij}$$

3.6、熵函数的性质 – 强可加性（续二）

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

联合熵

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

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

$$= -\sum_i p(x_i)\log p(x_i) + \left\{\sum_i p(x_i)\left[\underbrace{-\sum_j p(y_j|x_i)\log p(y_j|x_i)}_{H(Y|x_i)}\right]\right\}$$

$$= H(X) + \sum_i p_i H(Y|x_i)$$

$$= H(X) + \underbrace{H(Y|X)}_{\text{条件熵}}$$

3.6、熵函数的性质 – 强可加性（续三）

- 物理意义：
  - X和Y相关联时

$$H(XY)=H(X)+H(Y|X)=H(X)+\sum_i p(x_i)H(Y|x_i)$$

信源X每发一个符号所能提供的平均信息量

信源(XY)每发一个符号所能提供的平均信息量

符号x已知时，信源Y再发一个符号所提供的平均信息量

  - 当X和Y相互独立时　（可加性）

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

3.6、熵函数的性质 – 强可加性 – 注意!

- 条件熵的正确公式和错误公式之对比

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

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

- 多变量的数学期望

$$H(XY)=E[I(x_iy_j)]=-\sum_{XY}p(x_iy_j)\log p(x_iy_j)$$

★

$$H(Y|X)=E[I(y_j|x_i)]=-\sum_{XY}p(x_iy_j)\log p(y_j|x_i)$$

总结：条件熵的三种求解方法

$$H(Y|X) = E\left[I\left(y_j|x_i\right)\right]$$
$$= \sum_{XY} p\left(x_i y_j\right) I\left(y_j|x_i\right) = -\sum_{XY} p\left(x_i y_j\right) \log p\left(y_j|x_i\right)$$

$$H(Y|X)$$
$$= -\sum_{i} p\left(x_i\right) \sum_{j} p\left(y_j|x_i\right) \log p\left(y_j|x_i\right) = \sum_{i} p(x_i) H(Y|x_i)$$

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

3.6、熵的强可加性 – 例题2.5

- 已知联合概率分布如下，求：$H(XY)$，$H(X), H(Y), H(Y|X), H(X|Y)$？

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

- 理解解题思路，掌握解题步骤

3.6、熵的强可加性 – 例题2.5（续一）

1. $H(XY)=2.665\ bit/sym$

2. $$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0.25 & 0.40 & 0.15 & 0.15 & 0.05 \end{bmatrix}$$

   $H(X)=2.066\ bit/sym$

3. $$\begin{bmatrix} Y \\ P \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 \\ 0.35 & 0.35 & 0.20 & 0.10 \end{bmatrix}$$

   $H(Y)=1.856\ bit/sym$

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

3.6、熵的强可加性－

$$\begin{bmatrix} Y \\ P \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 \\ 0.35 & 0.35 & 0.20 & 0.10 \end{bmatrix}$$

4）

$$\begin{array}{cc} p(x_i \mid y_j) & \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \end{array} \\ \begin{array}{c} y_1 \\ y_2 \\ y_3 \\ y_4 \end{array} & \begin{bmatrix} 5/7 & 2/7 & 0 & 0 & 0 \\ 0 & 6/7 & 1/7 & 0 & 0 \\ 0 & 0 & 1/2 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

H*(*X*|*Y*)=0.809 *bit/sym

5）

$$\begin{array}{cc} p(y_j \mid x_i) & \begin{array}{cccc} y_1 & y_2 & y_3 & y_4 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{array} & \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ 0.25 & 0.40 & 0.15 & 0.15 & 0.05 \end{bmatrix}$$

H*(*Y*|*X*)=0.600 *bit/sym

3.6、熵的强可加性 – 练习2.4

- 随机变量*X*,*Y*的联合概率分布如下图，求联合熵$H(XY)$和条件熵$H(Y|X)$？

$P(XY)$

X \ Y	0	1
0	1/4	1/4
1	1/2	0

联合熵：　　*bit/sym*

条件熵：　　*bit/sym*

- ***条件熵的求解可采用多种方法***

3.7、熵函数的性质 － 递增性

$$H(p_1,p_2,\cdots,p_{n-1},q_1,q_2,\cdots,q_m)$$

$$= H(p_1,p_2,\cdots,p_{n-1},p_n) + p_n H\left(\frac{q_1}{p_n},\frac{q_2}{p_n},\cdots\frac{q_m}{p_n}\right)$$

$$\boxed{\sum_{i=1}^{n} p_i = 1 \quad \sum_{j=1}^{m} q_j = p_n}$$

说明：若信源原来的**n**个符号中，有一个符号$\boldsymbol{p_n}$被划分/分割成**m**个$\boldsymbol{q}$符号（这**m**个符号的概率之和等于$\boldsymbol{p_n}$的概率）。则信源的熵将增加，其增加项为新划分而产生的不确定性量。

3.7、熵的递增性 – 例题2.6

- 运用熵函数的递增性，计算*H(1/3,1/3,1/6,1/6)*的数值。

$$H\left(\frac{1}{3},\frac{1}{3},\frac{1}{6},\frac{1}{6}\right)$$

$$=H\left(\frac{1}{3},\frac{2}{3}\right)+\frac{2}{3}H\left(\frac{1}{2},\frac{1}{2}\right)+\frac{1}{3}H\left(\frac{1}{2},\frac{1}{2}\right)$$

$$=H\left(\frac{1}{3},\frac{2}{3}\right)+H\left(\frac{1}{2},\frac{1}{2}\right)=1.918\quad bit/sym$$

熵函数的两个重要性质

- 引言 – 函数的极值问题
  1. 对于一个熵函数，是否存在最大值？
  2. 在熵函数取最大值时，对应信源的概率分布（自变量）取值是怎样的？
- 两个重要性质
  - 上凸性
    - 凸函数中的极值点即为最值点！
  - 极值性
    - 这个极值是多少？

3.7、熵函数上凸性 – 预备知识

- 重要概念 – 凸函数
- 定义：设 $f(X)=f(x_1,x_2,\cdots,x_n)$ 为一个多元函数，若对任一个小于 **1** 的正数 $\alpha(0<\alpha<1)$ 以及函数 $f(X)$ 定义域中的任意两个矢量 X_1,X_2，有

$$f[\alpha X_1+(1-\alpha)X_2]\geq \alpha f(X_1)+(1-\alpha)f(X_2)$$

则称 $f(X)$ 为定义域上的上凸函数。

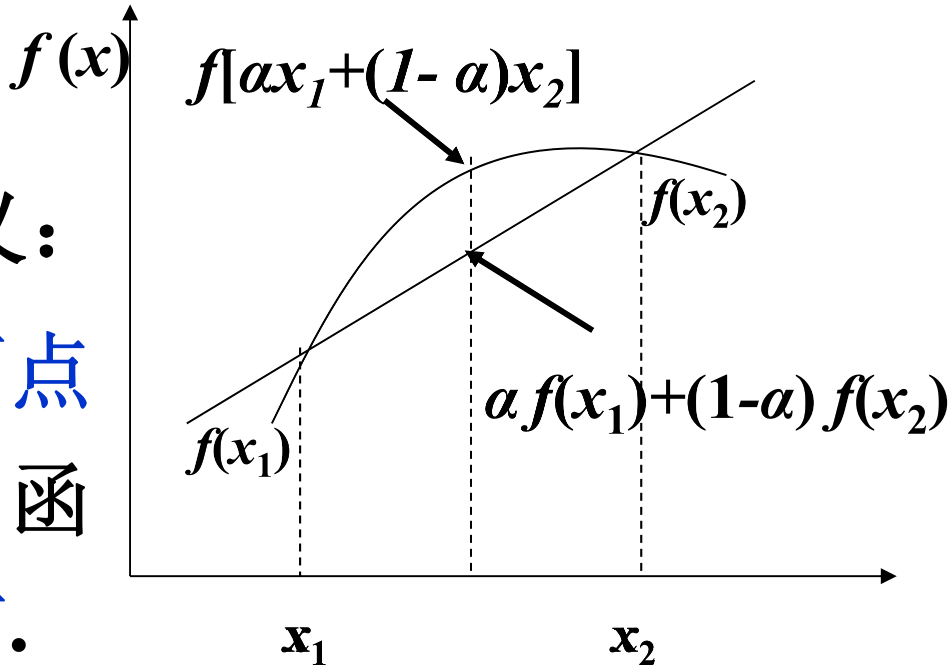
- （严格）上凸函数、（严格）下凸函数

《高等数学》中的凸函数说明；判断：x^2 $\ln x$ x^3

3.7、熵函数上凸性－ 预备知识（续一）

◆ 上凸性的几何意义：

在上凸函数的任两点之间画一条割线，函数总在割线的上方.



一元上凸函数示意图

- 上凸函数在定义域内的极值必为最大值！
- 熵函数是上凸函数吗？（上凸性）

3.7、熵函数上凸性－预备知识（续二）

- 引理：詹森不等式

数学归纳法

若 $f(X)$是定义在区间[***a*, *b***]上的实值连续上凸函数，则对于任一组 $x_1, x_2, \cdots, x_q \in [a,b]$和任意一组非负实数 $\lambda_1, \lambda_2, \cdots, \lambda_q$，$\sum_{k=1}^{q} \lambda_k = 1$，下列不等式成立：

$$\sum_{k=1}^{q} \lambda_k f(x_k) \le f\left[\sum_{k=1}^{q} \lambda_k x_k\right]$$

当***f*(*x*)**为***log***形式时：$E[\log x_i] \le \log(E[x_i])$

3.7、熵函数上凸性－预备知识（续三）

- 引理1：对于任意实数x>0，有 $\boxed{\ln x \le x-1}$

证明：

令 $f(x)=\ln x-(x-1)$

则 $f'(x)=\dfrac{1}{x}-1$ => $x=1$为极值点

因为 $f''(x)=-\dfrac{1}{x^2}<0$，所以$f(x)$是上凸函数，

$x=1$ 是 $f(x)$ 的极大值点。

因此有 $f(x)=\ln x-(x-1)\le 0$，即 $\ln x \le x-1$

3.7、熵函数上凸性－预备知识（续四）

◆ 引理2（香农辅助定理）：

$$H(p_1,p_2,\ldots p_n) = \boxed{-\sum_{i=1}^{n} p_i \log p_i \le -\sum_{i=1}^{n} p_i \log q_i} \bigstar$$

$$\sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} q_i = 1$$

$$H(p_1,p_2,\ldots p_n) + \sum_{i=1}^{n} p_i \log q_i = -\sum_{i=1}^{n} p_i \log p_i + \sum_{i=1}^{n} p_i \log q_i$$

$$= \sum_{i=1}^{n} p_i \log \frac{q_i}{p_i}$$

引入：信息散度，相对熵
(Kullback-Leibler距离)

$$\le \underline{\log} e \sum_{i=1}^{n} p_i [\frac{q_i}{p_i} - 1] = \log e \sum_{i=1}^{n} [p_i - q_i] = 0$$

任一集合的任一概率分布所定义的熵<u>必小于等于</u>该分布对其他概率分布的自信息取数学期望。

$$\boxed{\ln x \le x - 1}$$

3.7、熵函数上凸性

$$(\because -\sum_{i=1}^{n} p_i \log p_i \le -\sum_{i=1}^{n} p_i \log q_i)$$

★ $H(p_1, p_2, \ldots, p_n)$是概率分布$(p_1, p_2, \ldots, p_n)$的严格上凸函数，即

$$H\left[\alpha \mathbf{p} + (1-\alpha)\mathbf{p}'\right] > \alpha H(\mathbf{p}) + (1-\alpha) H(\mathbf{p}')$$

证明： $\alpha \mathbf{p} + (1-\alpha)\mathbf{p}'$ 可以被看做是一种新的概率分布

$$H\left[\alpha \mathbf{p} + (1-\alpha)\mathbf{p}'\right] = -\sum_{i=1}^{n}\left[\alpha p_i + (1-\alpha)\ p_i'\right]\log\left[\alpha p_i + (1-\alpha)\ p_i'\right]$$

$$= -\alpha\sum_{i=1}^{n} p_i \log\left[\alpha p_i + (1-\alpha)\ p_i'\right] - (1-\alpha)\sum_{i=1}^{n} p_i' \log\left[\alpha p_i + (1-\alpha)\ p_i'\right]$$

$$\ge -\alpha\sum_{i=1}^{n} p_i \log p_i - (1-\alpha)\sum_{i=1}^{n} p_i' \log p_i'$$

$$= \alpha H(\mathbf{p}) + (1-\alpha) H(\mathbf{p}')$$

3.8、熵函数的性质 － 极值性

- 最大离散熵定理
- 定理： 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概率相等时(即 $p_i = \frac{1}{n}$)，熵最大，即

★

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

3.8、熵函数的性质 – 极值性（续一）

- 熵的极值性的证明

$$H(\mathbf{P})=H(p_1,p_2,\cdots,p_n)=\boxed{\sum_{i=1}^{n}p_i\log\frac{1}{p_i}\le\log\sum_{i=1}^{n}p_i\frac{1}{p_i}}=\log n$$

- 等概率分布时，等号成立；
- 直观理解：概率空间内各随机事件发生机会均等，则不确定性最大。

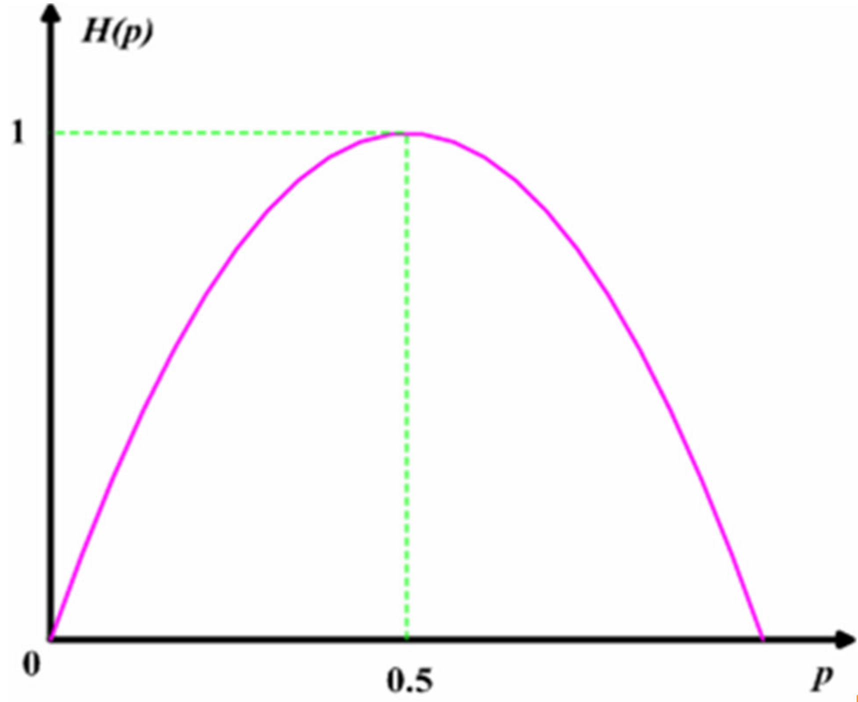
$$\boxed{E[\log x_i]\le\log(E[x_i])}$$

3.8、熵函数的性质 – 极值性 – 例题

- **例题2.7：**

  二进制信源概率空间为

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



- 二进制信源的信息熵为

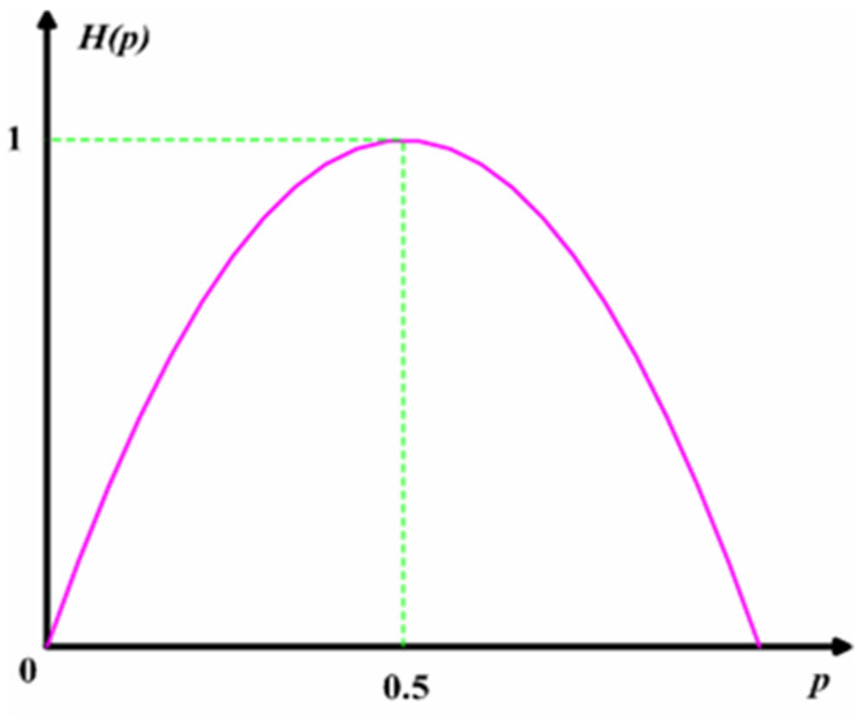
$$H(X) = -[p\log p + (1-p)\log(1-p)]$$

- 信息熵$H(X)$是p的函数，熵函数$H(p)$的曲线如上

在MATLAB中观察含有三个变量的熵函数的绘制曲面

3.8、熵函数的性质 – 极值性 – 例题分析

- 从图中可以得出熵函数的一些性质：
  - 若二进制信源的输出是确定的，则该信源不提供任何信息；
  - 当信源符号0和1等概率发生时，信源的熵达到最大值，等于1比特/符号；



本章主要内容

- 信源的数学模型及分类
- 离散信源的信息熵
- 熵的基本性质
- 离散无记忆扩展信源
  - 定义
  - 熵的求解
- 离散平稳信源
- 马尔可夫信源
- 信源冗余度与自然语言的熵
- 加权熵

4、离散无记忆信源的扩展信源

- 定义
  - 实际的信源输出为符号序列
  - 假定：符号序列中前后符号彼此无关
  - 扩展信源 – 多符号信源（同一概率空间）
- 用随机矢量/随机变量序列来描述

$$\mathbf{X} = X_1 X_2 X_3 \cdots$$

$$\begin{bmatrix} X_i \\ P(X_i) \end{bmatrix} = \begin{bmatrix} X_{i_1} = x_1 & \cdots & X_{i_i} = x_i & \cdots & X_{i_q} = x_q \\ p(x_1) & \cdots & p(x_i) & \cdots & p(x_q) \end{bmatrix}$$

4.1、离散无记忆信源的扩展信源－二进制信源

- 把信源输出序列看成是一组组发出的
- 二进制信源：信源输出符号‘**0**’，‘**1**’
  - 未扩展二进制信源 $a_1 = 0, a_2 = 1$
  - 二次扩展信源 $\alpha_1 = 00, \alpha_2 = 01, \alpha_3 = 10, \alpha_4 = 11$
  - 三次扩展信源

$$\alpha_1 = 000, \alpha_2 = 001, \alpha_3 = 010, \alpha_4 = 011,$$
$$\alpha_5 = 100, \alpha_6 = 101, \alpha_7 = 110, \alpha_8 = 111$$

  - N次扩展信源 2^N

4.1、离散无记忆信源的扩展信源 – 数学模型

- 单符号离散信源的数学模型

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} a_1 \cdots a_i \cdots a_q \\ p_1 \cdots p_i \cdots p_q \end{bmatrix} \qquad \sum_{i=1}^{q} p(x_i) = 1$$

- N次扩展信源对应的数学模型 ★

$$\begin{bmatrix} X^N \\ P(\mathbf{X}) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \cdots & \alpha_i & \cdots & \alpha_{q^N} \\ p(\alpha_1) & \cdots & p(\alpha_i) & \cdots & p(\alpha_{q^N}) \end{bmatrix}$$

  - 每个符号α_i 对应于某一个由N个a_i 组成的序列
  - α_i 的概率 $p(\alpha_i)$ 是对应的N个 a_i组成序列的概率

4.1、无记忆信源扩展信源 – 数学模型（续）

- 扩展信源的输出符号数目 q^N
  - 序列长度 $\boldsymbol{N}$
  - 符号集中的符号个数 $\boldsymbol{q}$

输出符号的概率

– $\boldsymbol{N}$重概率空间 $X^N=\left(X_1X_2\cdots X_N\right)$

– 离散无记忆信源

$\alpha_i=(x_{i_1}x_{i_2}\cdots x_{i_N})$的概率为

$$p(\alpha_i)=p(x_{i_1})p(x_{i_2})\cdots p(x_{i_N})=\prod_{k=1}^{N}p_{i_k}$$

4.2、N次扩展信源的熵

$$H(\mathbf{X})=H(X^N)=-\sum_{X^N}P(\mathbf{X})\log P(\mathbf{X})=-\sum_{X^N}p(\alpha_i)\log p(\alpha_i)$$

- 信源X的熵与扩展信源 X^N的熵之间的关系

$$H\left(X^N\right)=NH\left(X\right)$$

证明过程

完备性说明：

$$\sum_{X^N}p\left(\alpha_i\right)=\sum_{X^N}p_{i_1}p_{i_2}\cdots p_{i_N}=\sum_{i_1=1}^{q}\sum_{i_2=1}^{q}\cdots\sum_{i_N=1}^{q}p_{i_1}p_{i_2}\cdots p_{i_N}$$

$$=\sum_{i_1=1}^{q}p_{i_1}\sum_{i_2=1}^{q}p_{i_2}\cdots\sum_{i_N=1}^{q}p_{i_N}=1$$

4.2、N次扩展信源的熵 – 例题2.8

- 设有一离散无记忆信源**X**，其概率空间为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix} \qquad \sum_{j=1}^{3} p_j = 1$$

- 求该信源**X**的二次扩展信源的熵

求解思路

4.2、N次扩展信源的熵 － 例题（续）

- 信源X的原始信源熵

$$H(X)=\sum_{j=1}^{3}p(a_j)\log\frac{1}{p(a_j)}=1.5\quad \text{bit/sym}$$

- 二次扩展信源对应的概率空间

$$H(X^2)=\sum_{X^2}p(\alpha_i)\log\frac{1}{p(\alpha_i)}=\cdots=3\quad \underline{\text{bit/sym}}$$

- 二次扩展信源熵与原始信源熵

$$H(X^2)=2H(X)$$

本章主要内容

- 信源的数学模型及分类
- 离散信源的信息熵
- 熵的基本性质
- 离散无记忆扩展信源
- 离散平稳信源
- 马尔可夫信源
- 信源冗余度与自然语言的熵
- 加权熵

基本概念

熵的性质

信源熵的求解

5.1、基本概念 – 前言

- 实际信源输出往往是符号序列
- 离散多符号信源可以用随机矢量/随机变量序列来描述 $\mathbf{X} = X_1 X_2 \cdots X_N \cdots$
- 一般来说，信源的统计特性随着时间的推移而有所变化。为了便于研究，我们常常假定在一个较短的时间段内，信源是平稳信源。

5.1、基本概念 － 离散平稳信源 ★

- 对于离散随机变量序列 $X_1X_2\cdots X_n\cdots$，若任意两个不同时刻***i***和***j***(大于1的任意整数)信源发出消息的概率分布完全相同，即对于任意的 $N=0,\ 1,\ 2,\ \cdots$，$X_iX_{i+1}\cdots X_{i+N}$ 和 $X_jX_{j+1}\cdots X_{j+N}$ 具有相同的概率分布。也就是

$$P(X_i)=P(X_j)$$

一维平稳信源

$$P(X_iX_{i+1})=P(X_jX_{j+1})$$

二维平稳信源

$$\vdots$$

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

即各维联合概率分布均与时间起点无关的信源称为离散平稳信源。

5.1、基本概念 － 离散平稳信源（续）

- 对离散平稳信源，由联合概率与条件概率的关系可以推出

$$P(X_{i+1} \mid X_i) = P(X_{j+1} \mid X_j)$$

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

因此

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

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

$$H(X_3 \mid X_1 X_2) = H(X_4 \mid X_2 X_3) = \cdots = H(X_N \mid X_{N-2} X_{N-1})$$

5.1、基本概念 – 平均符号熵、极限熵

- **平均符号熵**：随机变量序列中，对前N个随机变量的联合熵求平均，即

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

如果当 $N \to \infty$ 时上式极限存在，则可定义 $\lim_{N\to\infty} H_N(\mathbf{X})$为**熵率**，或**极限熵**，记为

$$H_\infty = \lim_{N\to\infty} H_N(\mathbf{X})$$

➡ **思考：离散平稳无记忆扩展信源的熵率是多少？**

5.1、基本概念 – 离散平稳有记忆信源

- 前提：
  - 实际信源常常是有记忆信源
  - 设信源输出N长的符号序列，则可以用N维随机矢量 $\mathbf{X}=X_1X_2\cdots X_N$来表示信源。

- N维随机矢量的联合熵为

$$H(\mathbf{X})=H(X_1X_2\cdots X_N)$$

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

5.2、离散平稳信源的性质

- 对任意离散平稳信源，若$H_1(\mathbf{X})<\infty$，则具有以下性质：

1. $H\left(X_N \mid X_1 X_2 \cdots X_{N-1}\right)$ 随N的增加是非递增的；
2. N给定时，平均符号熵大于等于条件熵，

$$H_N(\mathbf{X}) \geq H\left(X_N \mid X_1 X_2 \cdots X_{N-1}\right)$$

3. 平均符号熵$H_N(\mathbf{X})$随N的增加是非递增，

$$H_N(\mathbf{X}) \leq H_{N-1}(\mathbf{X})$$

4. 极限熵 $\lim\limits_{N\to\infty} H_N(\mathbf{X}) = \lim\limits_{N\to\infty} H\left(X_N \mid X_1 X_2 \cdots X_{N-1}\right)$

5.3、极限熵/熵率的求解

- 对任意离散平稳信源，若 $H_1(\mathbf{X})<\infty$，则

$$\lim_{N\to\infty} H_N(\mathbf{X}) = \lim_{N\to\infty} H(X_N | X_1 X_2 \cdots X_{N-1})$$

- 证明思路
  - 首先证明存在性
  - 极限值的证明

5.3、极限熵的求解公式－ 证明

- 第一步：极限条件熵是否存在？
  - 只要**X** 的样本空间有限，则必有 $H_1(\mathbf{X}) = H(X_1) < \infty$
  - 根据条件熵的性质，以及信源的平稳性有

$$0 \le H(X_N \mid X_1 X_2 \ldots X_{N-1}) \le H(X_{N-1} \mid X_1 X_2 \ldots X_{N-2}) \le \ldots \le H(X_1) < \infty$$

$\therefore H(X_N \mid X_1 X_2 \ldots X_{N-1}), \quad N = 1,2,\ldots$ 是单调有界数列，

极限 $\lim_{N\to\infty} H(X_N \mid X_1 X_2 \ldots X_{N-1})$必然存在，且极限为**0**和 $H(X_1)$ 之间的某一个值。

5.3、极限熵的求解公式－ 证明（续）

- 第二步：极限值的证明？
  - 对于收敛的实数列，有以下结论成立：

    如果 $a_1, a_2, a_3, \cdots$ 是一个收敛的实数列，那么

$$\lim_{N\to\infty}\frac{1}{N}(a_1+a_2+\cdots+a_N)=\lim_{N\to\infty}a_N$$

由上述结论可知

$$\lim_{N\to\infty}H_N(\mathbf{X})=\lim_{N\to\infty}\frac{1}{N}\left[H(X_1)+H(X_2\mid X_1)+H(X_3\mid X_1X_2)+\cdots+H(X_N\mid X_1X_2\cdots X_{N-1})\right]$$

$$=\lim_{N\to\infty}H(X_N\mid X_1X_2\cdots X_{N-1})$$

当记忆长度 ***m*+1** 有限时

$$\boxed{\begin{aligned}\lim_{N\to\infty}H_N(X)&=\lim_{N\to\infty}H(X_N|X_1X_2\cdots X_{N-1})\\&=H(X_{m+1}|X_1X_2\cdots X_m)\end{aligned}}$$

5.3、极限熵的计算 – 例题2.9

- 平稳信源X的一维概率分布

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} a & b & c \\ \frac{1}{4} & \frac{4}{9} & \frac{11}{36} \end{bmatrix}$$

若该信源在输出符号序列时，只有前后两个符号之间有关联，条件概率如右表所示：

- 求信源极限熵/熵率（先判断存在否？）
- 求解$H(X)$ 和$H(X_2|X_1)$；假定以两个符号为一组，求解$1/2H(X_1X_2)$。

条件概率 $P(X_2 \mid X_1)$

X_1 \ X_2	a	b	c
a	7/9	2/9	0
b	1/8	3/4	1/8
c	0	2/11	9/11

5.3、极限熵的计算 － 例题（续）

1） $$H_\infty = \lim_{N\to\infty} H(X_N \mid X_1 X_2 \ldots X_{N-1}) = \lim_{N\to\infty} H(X_N \mid X_{N-1})$$

$$= H(X_2 \mid X_1) = 0.870$$ 比特/符号

2）如果不考虑符号间的相关性，则信源熵为

$$H(X) = H(\frac{1}{4}, \frac{4}{9}, \frac{11}{36}) = 1.542$$ 比特/符号

3）如果把信源发出的符号看成是分组发出的，每两个符号为一组，这个新信源的熵为

$$H(X_1 X_2) = H(X_1) + H(X_2 \mid X_1) = 2.412$$ 比特/两个符号

结论： $$H_\infty < \frac{1}{2} H(X_1 X_2) < H(X)$$

如何从理论上解释这个结果？

5.3、极限熵的含义

- 代表了一般离散平稳有记忆信源平均每发一个符号提供的信息量。
- 实际多符号离散平稳信源
  - 符号之间的统计关联关系伸向无穷远；
  - 研究记忆长度有限的信源更具有应用价值；
    - 信源无穷阶联合/条件概率分布极难测定
    - 多用条件熵或平均符号熵作为极限熵的近似值
- 非平稳离散信源
  - 马尔可夫信源

本章主要内容

- 信源的数学模型及分类
- 离散信源的信息熵
- 熵的基本性质
- 离散无记忆扩展信源
- 离散平稳信源
- 马尔可夫信源
- 信源冗余度与自然语言的熵
- 加权熵

基本概念

马尔可夫信源 – 马尔可夫链

马尔可夫信源熵求解

6.1、基本概念 －马尔可夫信源的定义

- 一种特殊的非平稳离散信源

1. 记忆长度有限

$$p\left(x_i \mid x_{i-1}x_{i-2}\cdots x_{i-m},\cdots\right)=p\left(x_i \mid x_{i-1}x_{i-2}\cdots x_{i-m}\right)$$

2. 马尔可夫链
   - 已知现在，将来与过去无关
   - 在给定当前知识或信息的情况下，过去（即当前点之前的历史状态）对于预测将来（即当前点以后的未来状态）是无关的。
   - 例：气象－预报－带伞；汉字输入法；……

6.1、基本概念 – 信源状态

- 重要概念 – 状态★

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

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

$$\cdots \quad s_i \qquad s_j \quad \cdots$$

$$\cdots S_1 S_2 \cdots S_m S_{m+1} S_{m+2} \cdots$$

思考：

状态的数目？

$$\boxed{\begin{aligned} & p(x_{i_{m+1}} \mid x_{i_1} x_{i_2} \cdots x_{i_m}) \\ & = p(x_{i_{m+1}} \mid s_i) \\ & = p(s_j \mid s_i) \end{aligned}}$$

马尔可夫链的引入？

6.1、基本概念 – 状态转移图 – 例题2.10

- 设一个二元一阶马尔可夫信源，信源符号集 $X=\{0,1\}$，输出符号的条件概率为

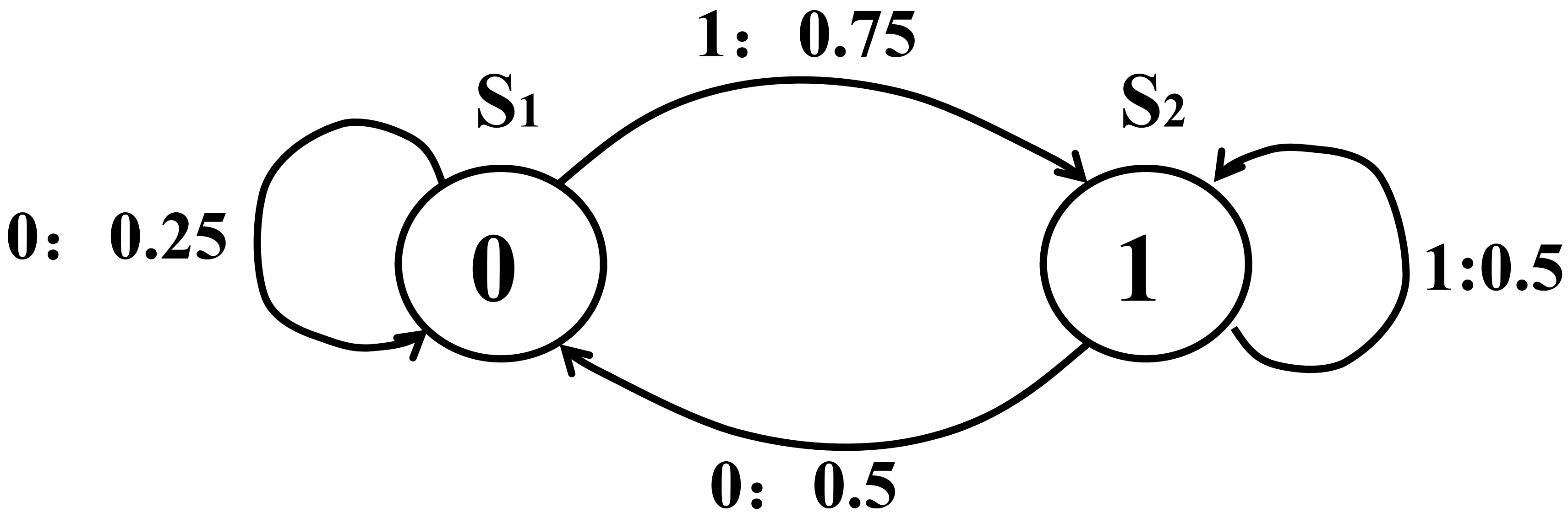
$$p(0|0)=0.25 \quad p(1|0)=0.75 \quad p(0|1)=0.5 \quad p(1|1)=0.5$$

- 要求：用状态转移图表示该信源
- 解题步骤：
  - 状态确定：状态数目、状态表示；
  - 状态间的转移：转移条件、转移概率；

6.1、基本概念－状态转移图－例题（续）

$$X=\{0,1\}$$

$$p(0|0)=0.25 \quad p(1|0)=0.75 \quad p(0|1)=0.5 \quad p(1|1)=0.5$$



二元一阶马尔可夫信源状态转移图

6.1、基本概念 － 状态转移图 － 练习

- 设一个二元二阶马尔可夫信源，信源符号集为 $X=\{0,\ 1\}$，输出符号的条件概率如下，请画出该信源的状态转移图。

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

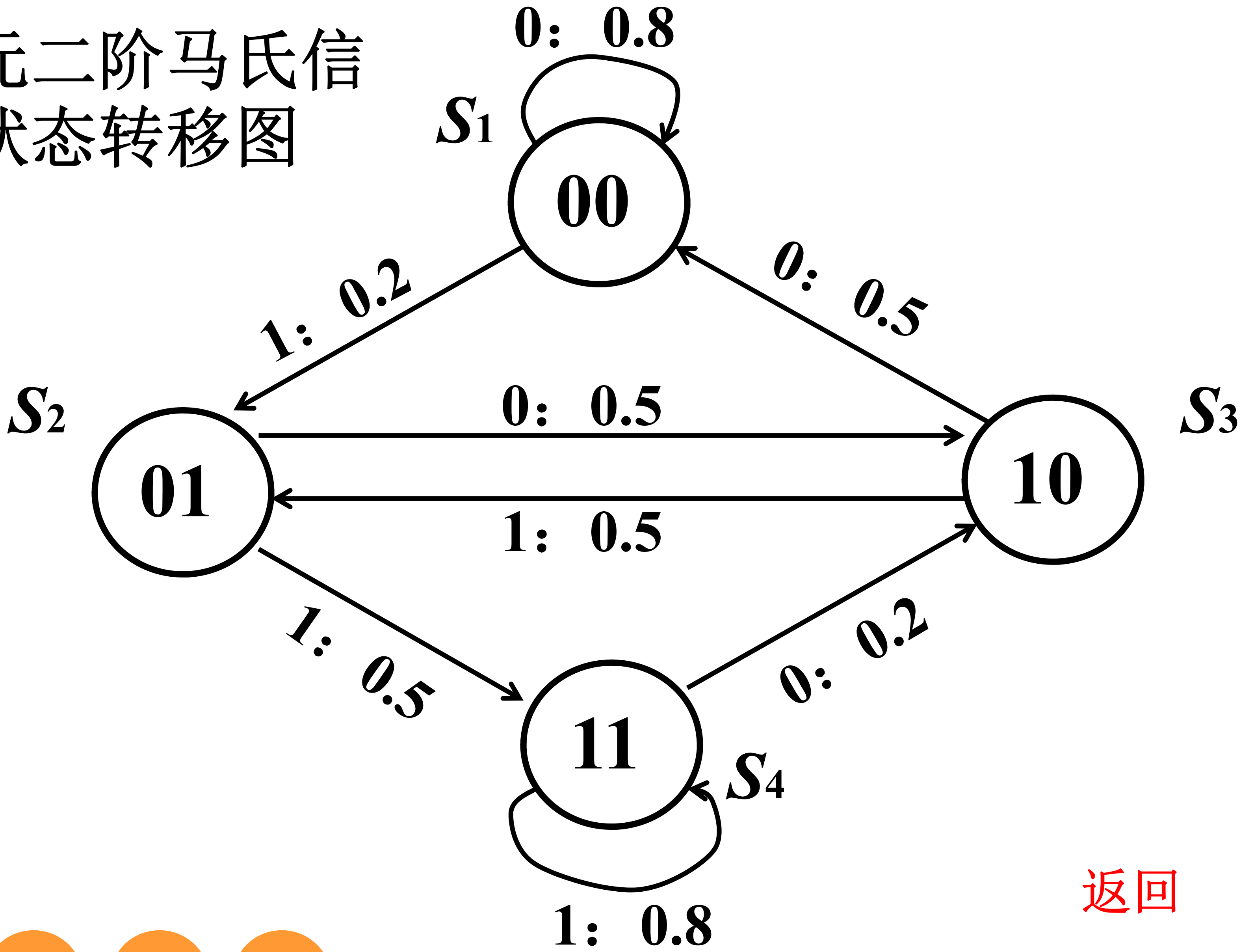
$$p(0|01)=p(0|10)=p(1|01)=p(1|10)=0.5,$$

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

- 考虑：状态数目、状态表示、转移概率

4.1、基本概念－状态转移图－练习（续）

二元二阶马氏信源状态转移图



返回

讨论

- 状态转移图 – 描述马尔可夫信源
  - 与时间有关吗？
  - 各不同状态条件下，每发一个符号所提供的信息量？
  - 状态的出现概率与什么有关？
  - 系统对外每发一个符号所提供的信息量？
  - 系统对外所发符号的概率分布？

6.2 有限状态马尔可夫链

- 设 $\{X_n, n\in N^+\}$为一个随机序列，时间参数集 $N^+=\{0,1,2,\cdots\}$，状态空间 $S=\{S_1,S_2,\cdots,S_J\}$，若对所有$n\in N^+$，有

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

则称 $\{X_n, n\in N^+\}$ 为马尔可夫链★

- 物理意义：随机序列{***X***(***n***)}的“将来”只通过“现在”与“过去”发生联系。

6.2、时齐马尔可夫链

- 如果马尔可夫链状态转移概率与起始时刻无关，即对任意*m*，有

$$p_{ij}(m)=p(X_{m+1}=S_j\,|\,X_m=S_i)=p_{ij}\qquad i,j\in 1,2,\cdots q^m$$

则称这类马尔可夫链为时齐(齐次)马尔可夫链。也称为平稳转移概率的马尔可夫链。

- 注意：与平稳信源概念不同

6.2、与平稳信源的关系

- 平稳信源是各维联合概率分布特性具有时间推移不变性；
- 齐次马尔可夫链要求转移概率具有时间推移不变性；
- 注意：平稳包含齐次，而齐次不包含平稳。即，一般来说齐次马尔可夫信源不一定是平稳信源。

6.2、时齐马尔可夫链 － 状态转移矩阵

- 齐次马氏链可以用转移概率矩阵或状态转移图来描述。一步转移概率P_{ij}对应的转移矩阵如下（假定共有J个状态）：

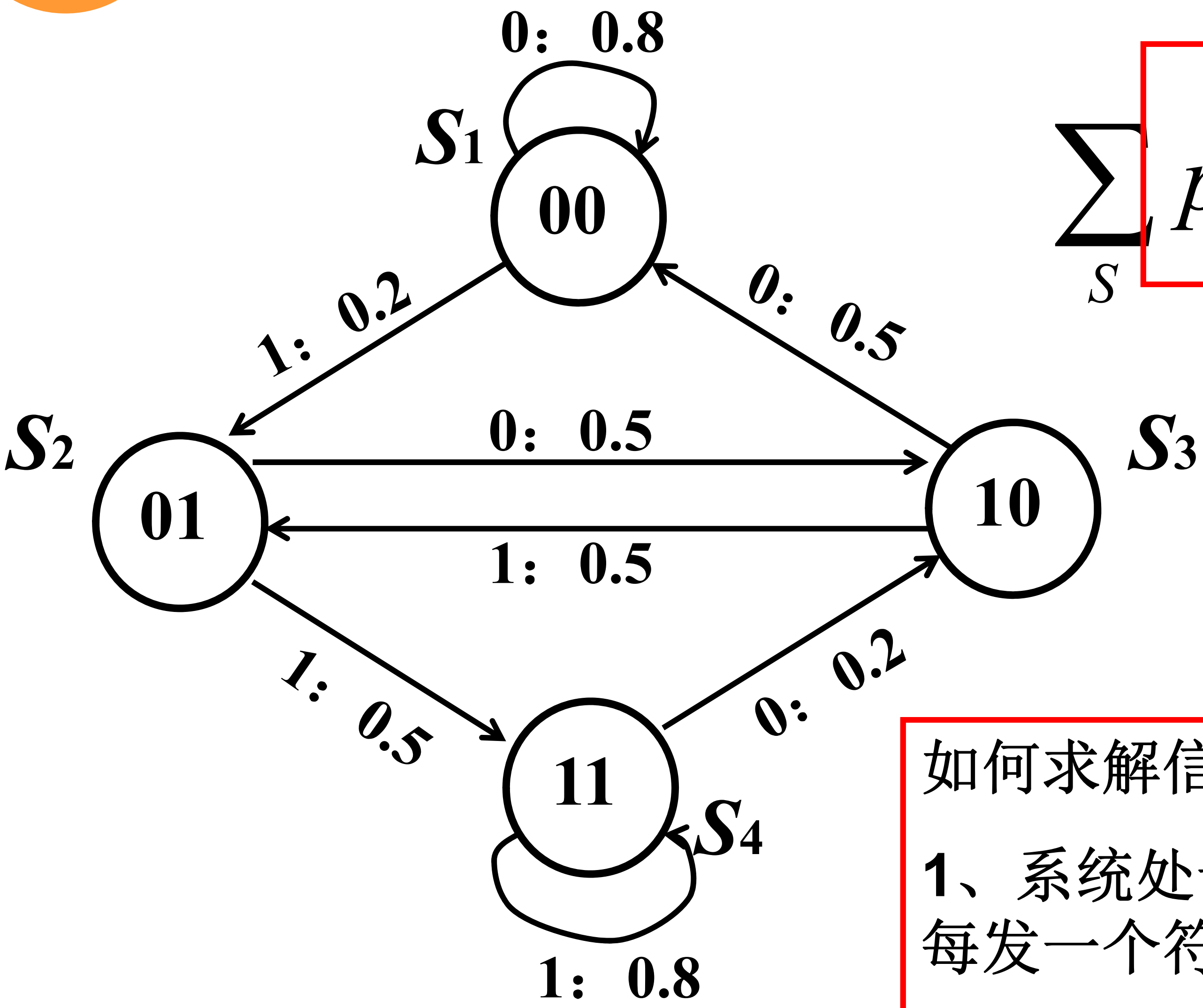
★

$$\mathbf{P}=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1J} \\ p_{21} & p_{22} & \cdots & p_{2J} \\ \cdots & \cdots & \cdots & \cdots \\ p_{J1} & p_{J2} & \cdots & p_{JJ} \end{bmatrix}$$

$$(i) \qquad 0 \le p_{ij}(m,n) \le 1$$

$$(ii) \qquad \sum_{j\in S} p_{ij}(m,n)=1$$

6.2、 马尔可夫信源熵求解分析 – 二元二阶为例



$$\sum_{S} \boxed{p(S_i)} H(X|S_i)$$

$$H(X|S_i)$$

如何求解信源熵？

1、系统处于某个状态时，对外每发一个符号所提供的信息量。

2、状态的出现概率。

6.2、 马尔可夫信源熵 – 稳态?

- 对于二元二阶马氏信源，四个状态[00，01，10，00]的概率即为$P(X_iX_{i+1})$。
- 马尔可夫信源不是平稳信源，即，前后两个时刻的联合概率分布与时间起点有关。
- 状态的出现概率能否与时间无关?
  - 如何求解马尔可夫信源的状态出现概率?

提问：状态转移图与状态转移矩阵

- 二元二阶马尔可夫信源，$X \in \{0, 1\}$，状态$S=(00, 01, 10, 11)$。下表为各状态下的符号发生概率，请画出该信源的状态转移图，并写出对应的状态转移矩阵。

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

6.2、 状态概率求解 – 状态转移概率★

- 描述马氏链最重要的参数（*m*, *n*表示时刻）

$$p_{ij}(m,n) = p\left\{X_n = S_j \mid X_m = S_i\right\} = p\{X_n = j \mid X_m = i\}$$

- 状态转移概率的性质

$$(i) \qquad 0 \le p_{ij}(m,n) \le 1$$

$$(ii) \qquad \sum_{j \in S} p_{ij}(m,n) = 1$$

- 一步转移概率

$$p_{ij}(m,m+1) \Rightarrow p_{ij}(m)$$

- *k* 步转移概率

$$p_{ij}^{(k)}(m) = P\left\{X_{m+k} = j \mid X_m = i\right\}$$

6.2、状态概率求解 – 多步转移概率求解

- 切普曼-柯尔莫哥洛夫方程（***Chapman-Kolmogorov***）

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

证明过程

$$p_{ij}^{(2)} = \sum_{k\in S} p_{ik} p_{kj} \qquad (i,j\in S)$$ 矩阵的乘法运算

$$p_{ij}^{(m+1)} = \sum_{k\in S} p_{ik}^{(m)} p_{kj} = \sum_{k\in S} p_{ik} p_{kj}^{(m)} \qquad (i,j\in S)$$

6.2、状态概率求解－多步转移概率求解（续）

$$p^{(k)} = p \cdot p^{(k-1)} = p \cdot p \cdot p^{(k-2)} = \cdots = p^{k}$$

表明：对于时齐马尔可夫链来说，一步转移概率完全决定了k 步转移概率。为了确定各状态的概率，还需知道状态的初始概率。

初始概率分布

$$P(X_n = S_j) = \sum_i P(X_n = S_j, X_0 = S_i) = \sum_i p_{0i} p_{ij}^{(n)}$$

$$j = 1,2,\cdots,J$$

即：由初始分布及各时刻的一步转移概率就可以完整描述马尔可夫链的统计特性。

6.2、状态概率求解 – 平稳分布的定义

- 若齐次马尔可夫链对于一切 i，j 存在不依赖于 i 的极限 $\lim_{n\to\infty} p_{ij}^{(n)} = p_j$
  且满足

$$p_j \geq 0, \qquad p_j = \sum_{i=0}^{J-1} p_{0i} p_{ij}, \qquad \sum_j p_j = 1$$

则称其具有遍历性（各态历经性），称为平稳分布，其中 p_{0i} 为该马尔可夫链的初始分布。

6.2、状态概率求解 – 平稳分布的分析

- 时齐性 – 平稳转移概率 – 转移矩阵
- 遍历性 – 平稳分布 – 对于状态而言
- 遍历性的直观意义
  - 转移步数n充分大时，可用一常数作为n步转移概率的近似值；
  - 转移步数足够大时，信源中所有状态出现的概率达到一种稳定分布，与信源的初始状态无关。
- 如何判断：遍历性/平稳分布的存在？

6.2、稳态分布存在的充要条件

- 设 $\mathbf{P}$ 为马氏链的状态转移矩阵，则该马氏链平稳分布存在的充要条件是，存在一个正整数 N，使矩阵 $\mathbf{P}^N$ 中的所有元素均大于零。

- 上述条件等价于，对于任意一个状态 $S_i(i=1\cdots J)$，存在同一个正整数 N，使得从任意初始状态出发，经过 N 步转移之后，一定可以到达该状态 S_i。

即，具有可遍历性。

6.2、平稳分布概率的符号引入

若齐次马尔可夫链对一切i, j存在不依赖与i的极限，

$$\lim_{n\to\infty} p_{ij}^{(n)} = W_j \qquad (i,j = 1,2,\cdots,J)$$

且满足

$$W_j > 0$$

$$\mathbf{WP} = \mathbf{W}$$

$$\sum_j W_j = 1$$

则称其具有遍历性，W_j称为稳态分布。

6.2、平稳分布的求解

$\lim\limits_{n\to\infty} p_{ij}^{(n)} = W_j$ 从定义上很难求。但是如果知道其极限存在，即可由如下方程组求得。

$$\bigstar \begin{cases} \mathbf{WP} = \mathbf{W} \quad \text{即} \quad \sum\limits_i W_i p_{ij} = W_j \\ \sum\limits_j W_j = 1 \end{cases}$$

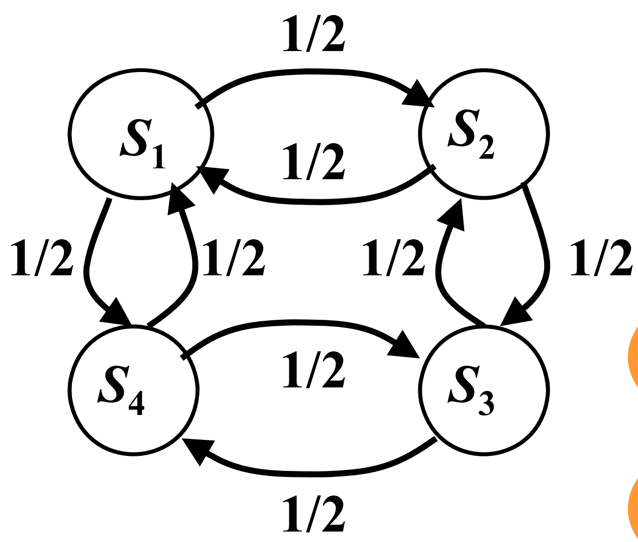
$\mathbf{W} = (W_1, W_2, \cdots W_J)$是该马氏链的稳态分布概率矢量且是唯一稳态分布。

该方程有解只是稳态分布存在的必要条件，而非充分条件。

6.2、平稳分布的求解 – 例题2.11

- 设有一马尔可夫链，其状态转移矩阵如右所示，试判断该马尔可夫链是否存在稳态分布？

$$\mathbf{P}=(p_{ij})=\begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}$$



$$\begin{cases} \sum_{i=1}^{4} W_i p_{ij} = W_j \\ \sum_{j=1}^{4} W_j = 1 \end{cases}$$ 有解，

$$W_j = 1/4,\ j = 1,2,3,4$$

6.2、平稳分布的求解 – 例题（续）

$$P^{(k)} = P^k = \boxed{\begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}} \quad k\text{为奇数}$$

$$P^{(k+1)} = P^{k+1} = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}$$

$$= \boxed{\begin{bmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{bmatrix}} \quad k\text{为偶数}$$

故，该马尔可夫链的稳态分布并不存在！

6.2、稳态分布存在的充要条件 – 例题2.12

- 设有一马尔可夫链，其状态转移矩阵如下所示，请判断是否存在稳态分布，若存在，求解对应的稳态分布？

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

- 解题步骤
  - 验证稳态分布是否存在？
  - 稳态分布求解！

6.2、稳态分布求解 － 例题（续）

- 求解第一步 - 验证是否存在稳态分布

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

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

存在正整数N=3，使得P^N中的所有元素均大于零

6.2、稳态分布求解 － 例题（续）

- 求解第二步 - 稳态分布的求解

$$\begin{array}{cc} & \begin{array}{ccc} p_{11} & p_{12} & p_{13} \end{array} \\ \mathbf{P} = \begin{array}{c} p_{11} \\ p_{21} \\ p_{31} \end{array} & \begin{bmatrix} 0 & 0 & 1 \\ 1/2 & 1/3 & 1/6 \\ 1/2 & 1/2 & 0 \end{bmatrix} \end{array}$$

$$\begin{cases} \sum_{i=1}^{3} W_i p_{ij} = W_j \\ \sum_{j=1}^{3} W_j = 1 \end{cases} \Rightarrow \begin{cases} 0 \cdot W_1 + 1/2 \cdot W_2 + 1/2 \cdot W_3 = W_1 \\ 0 \cdot W_1 + 1/3 \cdot W_2 + 1/2 \cdot W_3 = W_2 \\ 1 \cdot W_1 + 1/6 \cdot W_2 + 0 \cdot W_3 = W_1 \\ W_1 + W_2 + W_3 = 1 \end{cases}$$

$$W_1 = 1/3 \qquad W_2 = 2/7 \qquad W_3 = 8/21$$

与n步状态转移矩阵的关系

6.2、稳态分布求解 – 例题2.13

- 有一个二阶马尔可夫链 $X\in\{0, 1\}$，其条件概率如表所示，状态变量$S=(00, 01, 10, 11)$。则状态转移概率如表所示，求系统稳定后对外发送0、1符号的概率。

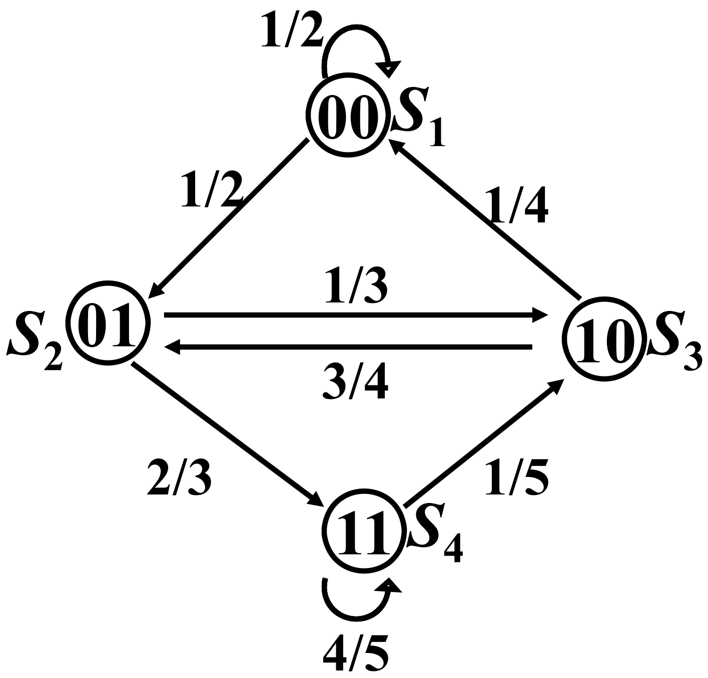
起始状态	符号	
	0	1
00	1/2	1/2
01	1/3	2/3
10	1/4	3/4
11	1/5	4/5

起始状态	终止状态			
	$S_1(00)$	$S_2(01)$	$S_3(10)$	$S_4(11)$
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

$$\mathbf{P1}=[p(a_j\mid S_i)]=\begin{bmatrix}1/2 & 1/2\\ 1/3 & 2/3\\ 1/4 & 3/4\\ 1/5 & 4/5\end{bmatrix}\quad \mathbf{P2}=[p(S_j\mid S_i)]=\begin{bmatrix}1/2 & 1/2 & 0 & 0\\ 0 & 0 & 1/3 & 2/3\\ 1/4 & 3/4 & 0 & 0\\ 0 & 0 & 1/5 & 4/5\end{bmatrix}$$

6.2、稳态分布求解 － 例题（续）

$$\mathbf{P1}=[p(a_j\,|\,S_i)]=\begin{bmatrix}1/2 & 1/2\\ 1/3 & 2/3\\ 1/4 & 3/4\\ 1/5 & 4/5\end{bmatrix}\qquad \mathbf{P2}=[p(S_j\,|\,S_i)]=\begin{bmatrix}1/2 & 1/2 & 0 & 0\\ 0 & 0 & 1/3 & 2/3\\ 1/4 & 3/4 & 0 & 0\\ 0 & 0 & 1/5 & 4/5\end{bmatrix}$$



显然，状态转移矩阵与符号转移矩阵是不同的，不能混淆。

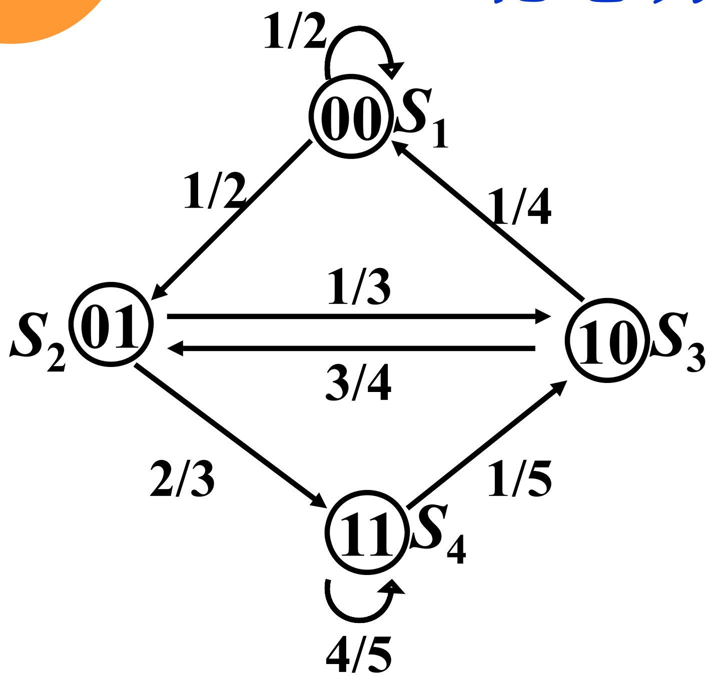
状态转移矩阵： $Q{*}Q$

符号条件转移概率：$Q{*}q$

其中，$Q=q^m$，

$q\in A=(a_1.a_2,\ldots a_q)$

6.2、稳态分布求解 － 例题（续）



$$\begin{cases} 1/2\cdot W_1 + 1/4\cdot W_3 = W_1 \\ 1/2\cdot W_1 + 3/4\cdot W_3 = W_2 \\ 1/3\cdot W_2 + 1/5\cdot W_4 = W_3 \\ 2/3\cdot W_2 + 4/5\cdot W_4 = W_4 \\ W_1 + W_2 + W_3 + W_4 = 1 \end{cases}$$

$W_1 = 3/35 \quad W_2 = 6/35 \quad W_3 = 6/35 \quad W_4 = 4/7$

解得，稳定后的状态概率分布$p(Si)$,而稳定后的符号概率是：

$$p(a_1=0)=\sum_i p(S_i)p(a_1\,|\,S_i)=\frac{1}{2}\times\frac{3}{35}+\frac{1}{3}\times\frac{6}{35}+\frac{1}{4}\times\frac{6}{35}+\frac{1}{5}\times\frac{4}{7}=\frac{9}{35}$$

$$p(a_2=1)=\sum_i p(S_i)p(a_2\,|\,S_i)=\frac{1}{2}\times\frac{3}{35}+\frac{2}{3}\times\frac{6}{35}+\frac{3}{4}\times\frac{6}{35}+\frac{4}{5}\times\frac{4}{7}=\frac{26}{35}$$

6.2、稳态分布求解 － 总结

- 先确定其稳态概率存在！
- 两种求解方法：

1. 看状态图：每一个状态的稳态概率＝转移到该状态的各个状态的概率统计平均值。
2. 公式法

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

6.2、说明 – 时齐遍历的马尔可夫信源

- *m* 阶马尔可夫信源在起始的有限时间内，信源不是平稳和遍历/各态历经性的，状态的概率分布有一段起始渐变过程。经过足够长时间之后，信源处于什么状态已与初始状态无关，这时每种状态出现的概率已达到一种稳定分布。
- 一般马尔可夫信源并非是平稳信源。但当时齐遍历的马尔可夫信源达到稳定后，这时就可以看成是平稳信源。

6.3、马尔可夫信源的信源熵 － 引言

- 实际信源近似为平稳信源
  - 将记忆长度*N*取足够大

$$H_\infty = \lim_{N\to\infty} H_N(\mathbf{X}) = \lim_{N\to\infty} \tfrac{1}{N} H(X_1 X_2 \ldots X_{N-1} X_N)$$

$$= \lim_{N\to\infty} H(X_N \mid X_1 X_2 \ldots X_{N-1})$$

$$= \lim_{N\to\infty} \left\{ -\sum_{k_1=1}^{n} \cdots \sum_{k_{m+1}=1}^{n} p(x_{k_1} \cdots x_{k_N}) \log_2 p(x_{k_N} \mid x_{k_1} \cdots x_{k_{N-1}}) \right\}$$

- 时齐遍历的马尔可夫信源在达到稳态分布时：
  - 计算***m*** 阶马尔可夫信源熵以近似信源熵
  - 信源的记忆长度越长，熵值越小

6.3、马尔可夫信源的极限熵 －定义

- *m* 阶马尔可夫信源

$$X: \begin{bmatrix} a_1 & a_2 & \cdots & a_q \\ & p\left(a_{i_{m+1}} \mid a_{i_1} a_{i_2} \cdots a_{i_m}\right) & & \end{bmatrix}$$

- 对应的马尔可夫信源状态空间

$$\begin{bmatrix} S_1 & S_2 & \cdots & S_{q^m} \\ & p\left(S_j \mid S_i\right) = p\left(a_{i_{m+1}} \mid a_{i_1} a_{i_2} \cdots a_{i_m}\right) & & \end{bmatrix} \qquad \begin{aligned} & S_i = \left(a_{i_1}, a_{i_2}, \cdots, a_{i_m}\right) \\ & i_1, \cdots, i_m \in (1,2,\cdots,q) \end{aligned}$$

- 信源的极限熵

$$H_\infty = \lim_{N \to \infty} H\left(X_N \mid X_1 X_2 \cdots X_{N-1}\right)$$
$$= H\left(X_{m+1} \mid X_1 X_2 \cdots X_m\right) = H_{m+1}$$

- 求解前提 - 存在平稳分布

6.3、马尔可夫信源的极限熵 – 公式推导

$$
\begin{aligned}
H_{m+1} &= H\left(X_{m+1} \middle| X_1 X_2 \cdots X_m\right) \\
&= E\left[-\log p\left(x_{i_{m+1}} \middle| x_{i_1} x_{i_2} \cdots x_{i_m}\right)\right] \\
&= E\left[-\log p\left(x_{i_{m+1}} \middle| S_i\right)\right] \\
&= -\sum_{i=1}^{q^m} \sum_{i_{m+1}}^{q} p(S_i)\, p\left(x_{i_{m+1}} \middle| S_i\right) \log p\left(x_{i_{m+1}} \middle| S_i\right) \\
&= \sum_i p(S_i) H(X|S_i)
\end{aligned}
$$

6.3、马尔可夫信源的极限熵 －定义（续）

$$H_{m+1}=\sum_{S_j} p(S_j)H(X \mid S_j)$$

说明：

1、马尔可夫特性 - 极限熵求解中的无限大(*N*)变为有限值*m*；

2、*p* (S_j)是马尔可夫信源稳定后各状态的极限概率；

3、该极限熵表示*m*阶马尔可夫信源每发一个符号提供的平均信息量，单位：比特/符号。

6.3、马尔可夫信源的极限熵 – 例题2.13（续）

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

稳定后的状态概率分布$p(S_i)$

$$W_1=\frac{3}{35}\quad W_2=\frac{6}{35}\quad W_3=\frac{6}{35}\quad W_4=\frac{4}{7}$$

$$H_\infty=H_{m+1}=H_3$$

$$=-\sum_j\sum_i p(a_i,S_j)\log p(a_i\mid S_j)=-\sum_j\sum_i p(S_j)p(a_i\mid S_j)\log p(a_i\mid S_j)$$

$$=\sum_j p(S_j)H(X\mid S_j)$$

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

6.3、马尔可夫信源的极限熵 － 例题2.14

- 某二进制二阶马尔可夫信源X的符号集为{0、1}，四个状态 S_1,S_2,S_3,S_4 分别为00、01、10、11，并测定一步转移概率为：

$$p(S_1|S_1)=0.8 \quad p(S_2|S_1)=0.2 \quad p(S_3|S_2)=0.5 \quad p(S_4|S_2)=0.5$$

$$p(S_1|S_3)=0.5 \quad p(S_2|S_3)=0.5 \quad p(S_4|S_4)=0.8 \quad p(S_3|S_4)=0.2$$

求信源极限熵、信源平稳时的符号概率。

6.3、马尔可夫信源的极限熵 – 例题（续）

首先判断是否存在平稳分布！

状态转移图

$$P=\begin{bmatrix} 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \end{bmatrix}$$

约定：

$W_1=p(S_1)\quad W_2=p(S_2)\quad W_3=p(S_3)\quad W_4=p(S_4)$

$$\begin{cases} W_1+W_2+W_3+W_4=1 \\ W_i>0 \qquad (i=1,2,3,4) \end{cases}$$

$$\because \begin{cases} WP=W \\ \sum_i W_i=1 \end{cases}$$

$$[W_1 \quad W_2 \quad W_3 \quad W_4]\times\begin{bmatrix} 0.8 & 0.2 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \end{bmatrix}=[W_1 \quad W_2 \quad W_3 \quad W_4]$$

6.3、马尔可夫信源的极限熵 – 例题（续）

解答：

$$W_1 = W_4 = \frac{5}{14}, \qquad W_2 = W_3 = \frac{1}{7}$$

$$H_\infty = H_{m+1}$$

$$= H_3 = \sum_{Si} p(S_i) H(X \mid S_i)$$

$$= \frac{5}{14} H(0.8,0.2) + \frac{1}{7} H(0.5,0.5) + \frac{1}{7} H(0.5,0.5) + \frac{5}{14} H(0.8,0.2)$$

$$= 0.8(bit/\text{信符})$$

符号的平稳概率分布为：

$$p(0) = 0.8 * W_1 + 0.5 * W_2 + 0.5 * W_3 + 0.2 * W_4 = 0.5$$

$$p(1) = 0.2 * W_1 + 0.5 * W_2 + 0.5 * W_3 + 0.8 * W_4 = 0.5$$

6.3、马尔可夫信源的极限熵 – 例题2.15

- 某二元二阶马氏信源X的符号集{0、1}:

1. 起始概率（一维分布）：

$$p(0)=p(1)=0.5$$

2. 下一单位时间的随机变量X_2与X_1的依赖关系由条件概率$p(X_2|X_1)$表示：

$$p(0|0)=0.3, p(1|0)=0.7,$$
$$p(0|1)=0.4, p(1|1)=0.6$$

6.3、马尔可夫信源的极限熵 – 例题（续）

3. 再下一单位时间的随机变量X_3与X_1X_2有依赖关系，且由二阶条件概率$p(X_3 \mid X_1X_2)$表示：

$$p(0|00)=p(0|S_1)=0.4, p(1|00)=p(1|S_1)=0.6,$$

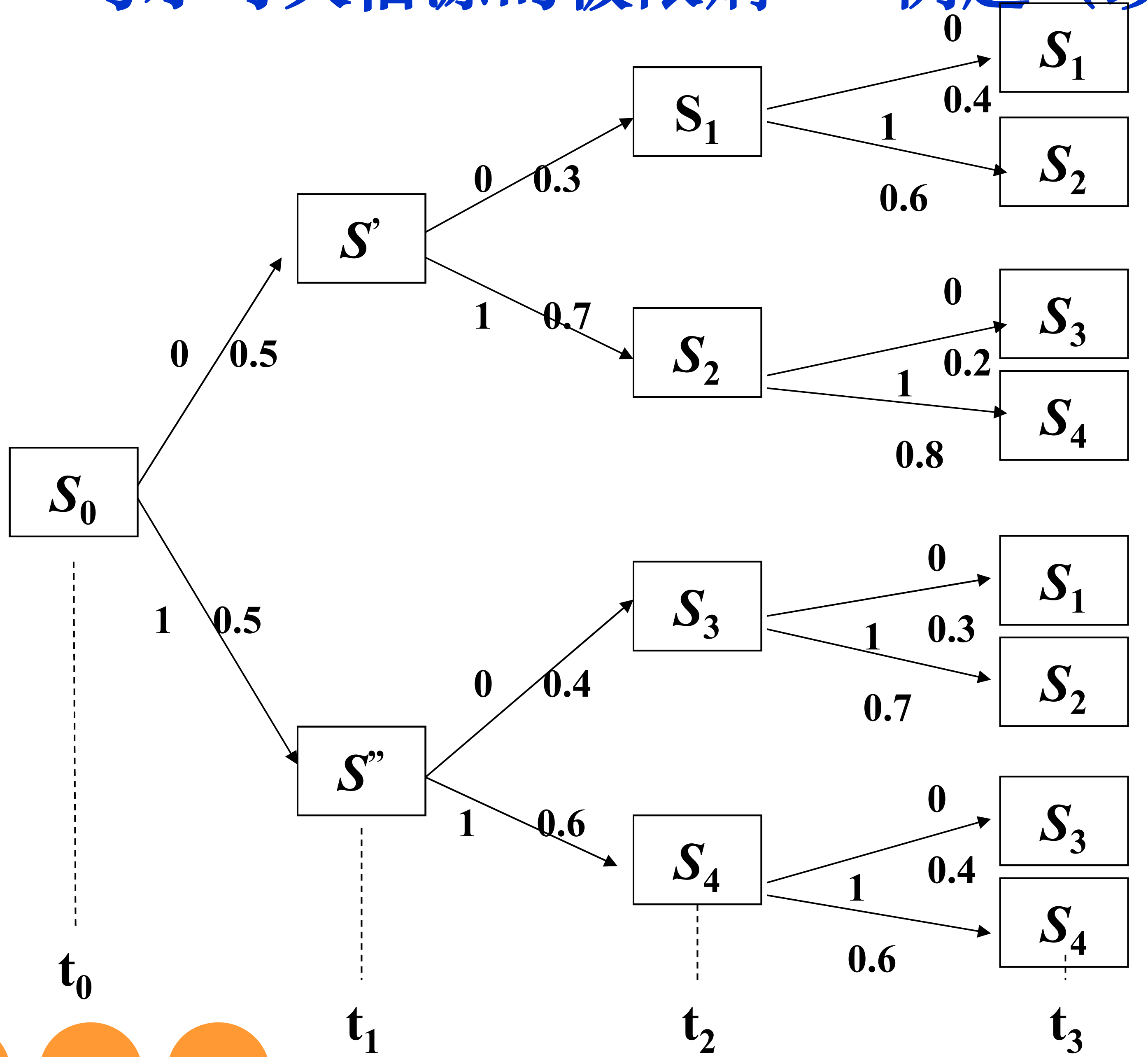
$$p(0|01)=p(0|S_2)=0.2, p(1|01)=p(1|S_2)=0.8,$$

$$p(0|10)=p(0|S_3)=0.3, p(1|10)=p(1|S_3)=0.7,$$

$$p(0|11)=p(0|S_4)=0.4, p(1|11)=p(1|S_4)=0.6,$$

4. 从第四个单位时间开始，之后任一时刻的随机变量只与前面两个单位时刻的随机变量之间有依赖关系，与更前面的随机变量无关，其依赖关系均由上述二维条件概率表示。

6.3、马尔可夫信源的极限熵 － 例题（续）



6.3、马尔可夫信源的极限熵 － 例题（续）

- 当马尔可夫信源稳定后，状态S_0,S',S'' 均不可能再重新到达，是过渡态，计算可知具有各态历经性，则：

- 首先**判断**是否存在平稳分布！

$$\begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \end{bmatrix}^T \begin{bmatrix} 0.4 & 0.6 & 0 & 0 \\ 0 & 0 & 0.2 & 0.8 \\ 0.3 & 0.7 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} W_1 \\ W_2 \\ W_3 \\ W_4 \end{bmatrix}^T$$

且

$$\begin{cases} W_1 + W_2 + W_3 + W_4 = 1 \\ W_i > 0 \qquad (i = 1,2,3,4) \end{cases}$$

$$W_1 = \frac{1}{9}, W_2 = \frac{2}{9}, W_3 = \frac{2}{9}, W_4 = \frac{4}{9}$$

$$H_\infty = H_{m+1} = H_3 = \sum_{i=1}^{4} p(S_i) H(X \mid S_i) = 0.896 \quad bit/sym$$

6.3、马尔可夫信源的极限熵 – 例题（续）

- 符号概率分析
  - 起始概率分布给定为：

$$p(X_1=0)=p(X_1=1)=0.5$$

  - 发$\mathbf{X_1}$的条件下，第二单位时间再发$\mathbf{X_2}$的过渡阶段中，{0, 1}符号的概率分布分别为：

$$p(X_2=0)=p(X_1=0)\cdot p(X_2=0\,|\,X_1=0)+p(X_1=1)\cdot p(X_2=0\,|\,X_1=1)$$

$$=0.5\times0.3+0.5\times0.4=0.35$$

$$p(X_2=1)=p(X_1=0)\cdot p(X_2=1\,|\,X_1=0)+p(X_1=1)\cdot p(X_2=1\,|\,X_1=1)$$

$$=0.5\times0.7+0.5\times0.6=0.65$$

6.3、马尔可夫信源的极限熵 － 例题（续）

- 当信源进入稳定状态（**n**时刻）时，符号**{0, 1}**的概率分布分别为：

$$p(X_n=0)=\sum_{i=1}^{4} p(S_i)\cdot p(0\mid S_i)$$

$$=\frac{1}{9}\times 0.4+\frac{2}{9}\times 0.2+\frac{2}{9}\times 0.3+\frac{4}{9}\times 0.4=\frac{1}{3}$$

$$p(X_n=1)=\frac{2}{3}$$

*** 由上可见，在系统进入稳态时，符号{0、1}的概率分布与起始时刻及过渡阶段的概率分布均不相同。**

信源熵的分析历程总结

- 实际信源可能是非平稳有记忆信源
  - 极限熵可能不存在
- 假定为平稳信源
  - 无记忆信源的扩展信源 $H\left(X^N\right)=NH\left(X\right)$
  - 多符号有记忆信源
    - 实际信源符号之间的依赖关系延伸到无穷，应考虑极限熵

$$H_\infty=\lim_{N\to\infty}H_N(X_1X_2\ldots X_N)=\lim_{N\to\infty}H(X_N\mid X_1X_2\ldots X_{N-1})$$

$$\begin{cases}\text{当记忆长度有限时 } \boldsymbol{H}_\infty=\boldsymbol{H}_{m+1} & \\ \text{当记忆长度无限时，需要测定无限阶联合概率分布和条件概率分布，很难计算！} & \end{cases}$$

二维平稳信源例

信源熵的分析历程总结（续）

- 实际信源常常是非平稳、有记忆信源，极限熵不存在，为了便于分析，常常做如下近似：

  （1）在一个较短的时间窗类，可近似为平稳。

  （2）实际信源的记忆长度可追溯到起始时刻，但是通常相隔较远的两个符号，相关性已变得很弱，因此可以限制其记忆长度。

- 在做了上述假定之后，即可把实际信源当作马尔可夫信源进行研究。

信源熵的分析历程总结（续）

- 进一步假设为m阶马尔可夫信源
  - 信源的随机状态序列构成马尔可夫链，当此马氏链满足各态历经的条件时，即具有稳态分布。

$$H_{\infty}=H_{m+1}=\sum_{i}p(S_i)H(X\mid S_i)$$

  - 只需知道和前m个符号有关的条件概率既可计算信息熵
- 一般平稳信源可以用m阶马尔可夫信源来近似（有效地限定记忆长度）。

本章主要内容

- 信源的数学模型及分类
- 离散信源的信息熵
- 熵的基本性质
- 离散无记忆扩展信源
- 离散平稳信源
- 马尔可夫信源
- 信源冗余度与自然语言的熵
  - 信源相关性
  - 英语信源的熵
  - 信源冗余度分析
- 加权熵

7.1、信源的相关性

- 信源的相关性就是信源符号间的依赖程度。
- 设信源有m个符号，那么不同情况信源熵的计算如下：（对于马氏信源，则？）

$H_0 = \log m$ （独立等概同分布）

$H_1 = H(X_1)$ （独立同分布平稳信源）

$H_2 = H(X_2 \mid X_1)$ （记忆长度为2的平稳信源）

$H_n = H(X_n \mid X_1 X_2 \cdots X_{n-1})$ （记忆长度为n的平稳信源）

$H_\infty = \lim\limits_{n\to\infty} H(X_n \mid X_1 X_2 \cdots X_{n-1})$ （记忆长度无限的平稳信源）

7.1、信源的相关性

- 相关性与信息熵的关系
  - 对同一信源，采用不同的模型，计算得到的熵的关系为

$$H_0 \geq H_1 \geq H_2 \geq \cdots \geq H_{m+1} \geq \cdots \geq H_\infty$$

  - 结论：符号间相关性越大，信源熵越小。

7.2、英语信源的熵

① 把英语看成是离散无记忆信源

- 英语信源的最大熵(等概率) $H_0=\log_2 27=4.76$(比特/符号)
- 英语字母并非等概率出现，字母之间有严格的依赖关系。表中是对27个符号出现的概率统计结果。

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

- 如果不考虑符号间的依赖关系，近似认为信源是离散无记忆的，则

$$\boxed{H_1 = -\sum_{i=1}^{27} p(e_i)\log_2 p(e_i) = 4.03(\text{比特}/\text{符号})}$$

- 按表中的概率分布，随机地选择英语字母并排列起来，得到一个输出序列：
- **AI_NGAE_ITE_NNR_ASAEV_OTE_BAINTHA_HYROO_PORE_SETRYGAIETRWCO_EHDUARU_EUEU_C_FT_NSREM_DIY_EESE_F_O_SRIS_R_UNNASHOR…**
- 有点像英语，但不是。
- 实际英语<u>字母之间有强烈的依赖性</u>。例如 ***T*** 后面出现 ***H***, ***R*** 的可能性较大，出现 ***J***, ***K***, ***M***, ***N*** 的可能性极小，而根本不会出现 ***Q***, ***F***, ***X***。

② 把英语看成马尔可夫信源

- 为了进一步逼近实际情况，可把英语信源近似看做**1阶**，**2阶**，…∞阶马尔可夫信源，它们的熵为

$$H_2=3.32(\text{比特/符号})$$

$$H_3=3.1(\text{比特/符号})$$

- 若把英语信源近似成**2**阶马尔可夫信源，可得到某个输出序列：
- **IANKS_CAN_OU_ANG_RLER_THTTED_OF_TO_SHOR_OF_TO_HAVEMEM_A_I_MAND_AND_BUT_WHISS_ITABLY_THERVEREER…**

- 二阶条件下：被空格分开的两字母或三字母，组成的大都是有意义的英语单词，而四个以上字母组成的“单词”，很难从英语词典中查到。因为该序列仅考虑了3个以下字母之间的依赖关系。实际英语字母之间的关系延伸到更多的符号，单词之间也有依赖关系。
- 有依赖关系的字母数越多，即马尔可夫信源的阶数越高，输出的序列就越接近于实际情况。当依赖关系延伸到无穷远时，信源输出的就是真正的英语，此时可求出马尔可夫的极限熵 H_∞ =1.4(比特/符号)。
- 汉语：复杂多了！
  - 汉字符号集很大；
  - 分析思路：约定汉字集 – 等概出现 – 不同出现频率集 – 字、词、句之间的相关性 – 语法规则 - ……

7.3、信源的冗余度分析 - 1

- 信息传输手段的浪费
  - 对一般离散平稳信源，H_∞就是实际信源熵。理论上只要有传送H_∞的手段，就能把信源包含的信息全部发送出去。
- 冗余度定义及意义 $$R = 1 - \frac{H_\infty}{H_0}$$ ★
  - 信源熵的相对率η：信源实际的信息熵与同样符号数的最大熵的比值：$\eta = H_\infty / H_0$
- 信源的实际熵应为H_∞，但H_∞很难得到，于是用H_0来表达信源。两者之差代表了语言结构确定的信息。差值越大，冗余度越大。英语信源剩余度为 R=(4.76-1.4)/4.76=0.71

7.3、信源的冗余度分析 - 2

- 重要结论
  - 例如：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（*理想化*）。信息的冗余度表示信源可压缩的程度。
  - 从提高传输效率的观点出发，总是希望减少或去掉冗余度。
  - 冗余度大的消息抗干扰能力强。能通过前后字之间的关联纠正错误。

本章主要内容

- 信源的数学模型及分类
- 离散信源的信息熵
- 熵的基本性质
- 离散无记忆扩展信源
- 离散平稳信源
- 马尔可夫信源
- 信源冗余度与自然语言的熵
- 加权熵

8、加权熵

- 引入事件的权重，度量事件的重要性或主观价值

$$\begin{bmatrix} X \\ P \\ W \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ p(x_1) & p(x_2) & \dots & p(x_n) \\ w_1 & w_2 & \dots & w_n \end{bmatrix}$$

- 加权熵定义

$$H_W(X) = -\sum_{i=1}^{n} w_i p(x_i) \log p(x_i) \qquad w_i \geq 0$$

8、加权熵 – 基本性质

性质1：非负性

性质2：等重性：若权重相等，则加权熵与香农熵类似

性质3：无重性：若权重均为零，则加权熵为零

性质4：确定性

性质5：非容性：若样本空间由I, J构成，即

$$I \cup J = (1, 2, \cdots n) \qquad I \cap J = \wedge$$

若 $\forall i \in I$， 有$p_i = 0, w_i \neq 0$

且 $\forall j \in J$， 有$p_j \neq 0, w_i = 0$

则 $H_W(X) = 0$

$$\boxed{w_i \geq 0}$$

$$\boxed{H_W(X) = -\sum_{i=1}^{n} w_i p(x_i) \log p(x_i)}$$

8、加权熵 – 基本性质（续）

性质6：扩展性

$$H_w^{q+1}\left(w_1,\cdots,w_q,w_{q+1},p_1,\cdots,p_q,p_{q+1}=0\right)$$
$$=H_w^{q}\left(w_1,\cdots,w_q,p_1,\cdots,p_q\right)$$

性质7：线性叠加性

$$H_w\left(\lambda w_1,\lambda w_2\cdots,\lambda w_q,p_1,\cdots,p_q\right)$$
$$=\lambda H_w\left(w_1,\cdots,w_q,p_1,\cdots,p_q\right)$$

性质8：递增性

$$H_w^{n+1}\left(w_1,\cdots,w_{n-1},w',w'',p_1,\cdots,p_{n-1},p',p''\right)$$
$$=H_w^{n}\left(w_1,\cdots,w_n,p_1,\cdots,p_n\right)+p_nH_w^2\left(w',w'',\frac{p'}{p_n},\frac{p''}{p_n}\right)$$

其中：

$$p_n=p'+p'' \qquad w_n=\frac{p'w'+p''w''}{p'+p''}$$

本章总结

- 信源的数学模型及分类
- 离散信源的信息熵
- 熵的基本性质
- 离散无记忆扩展信源
- 离散平稳信源
- 马尔可夫信源
- 信源冗余度与自然语言的熵
- 加权熵

练习 2.2 – 解答

返回

例：掷一个均匀的硬币，直到出现“正面”为止，令X表示所需掷的次数，求熵$H(X)$。

解：X表示所需掷的次数，则信源X可能发出的符号是表示次数的数字，X：{1, 2, 3, ...}。第n次出现正面的概率：$p=2^{-n}$。所以，信源X的信源空间可表示为：

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 1, & 2, & 3, & \cdots \\ 2^{-1}, & 2^{-2}, & 2^{-3}, & \cdots \end{bmatrix}$$

$$H(X) = -\sum_{n=1}^{\infty} p(X=n)\cdot \log p(X=n)$$

$$= -\sum_{n=1}^{\infty} 2^{-n}\cdot \log 2^{-n} = \sum_{n=1}^{\infty} n\cdot 2^{-n} = 2 \quad bit/sym$$

上下凸函数的数学定义

返回

- 凸函数定义：参见附录A
- 拐点：设曲线$y=f(x)$在点a的一边为上凸，一边为下凸，则称该点a为曲线的拐点。
- 定理（凸函数与二阶导数的关系）：
  - 设$f(x)$在(a, b)二阶可导，则
  1. 若$f(x)$的二阶导数在(a, b)小于0，则函数在此区间上凸;
  2. 若$f(x)$的二阶导数在(a, b)大于0，则函数在此区间下凸;
- 拐点必要条件
  - 若函数在点a为拐点，则要么（1）在此点的二阶导数为0；要么（2）函数在此点不可导。

对于f(x)的凸性判断

返回

$$f(x) = -x\log x$$

$$f'(x) = -\left(1*\log x + x*\log e*\frac{1}{x}\right) = -(\log x + \log e)$$

$$f''(x) = -\log e*\frac{1}{x} = -\frac{\log e}{x} < 0 \qquad \mathrm{x} \in (0, \quad +\infty)$$

由于有对数函数，x的取值空间是正数，

则函数的二阶导数为负数，故该函数为上凸函数。

$$H\left(X^N\right)=NH\left(X\right)$$

$$H(\mathbf{X})=H(X^N)=-\sum_{X^N}P(\mathbf{X})\log P(\mathbf{X})=-\sum_{X^N}p(\alpha_i)\log p(\alpha_i)$$

$$H\left(X^N\right)=\sum_{X^N}p\left(\alpha_i\right)\log\frac{1}{p_{i_1}p_{i_2}\cdots p_{i_N}} \qquad p(\alpha_i)=p(x_{i_1})p(x_{i_2})\cdots p(x_{i_N})=\prod_{k=1}^{N}p_{i_k}$$

$$=\sum_{X^N}p\left(\alpha_i\right)\log\frac{1}{p_{i_1}}+\sum_{X^N}p\left(\alpha_i\right)\log\frac{1}{p_{i_2}}+\cdots+\sum_{X^N}p\left(\alpha_i\right)\log\frac{1}{p_{i_N}}$$

其中一项：$\displaystyle\sum_{X^N}p\left(\alpha_i\right)\log\frac{1}{p_{i_1}}=\sum_{X^N}p_{i_1}p_{i_2}\cdots p_{i_n}\log\frac{1}{p_{i_1}}$

$$=\sum_{i_1=1}^{q}p_{i_1}\log\frac{1}{p_{i_1}}\sum_{i_2=1}^{q}p_{i_2}\sum_{i_3=1}^{q}p_{i_3}\cdots\sum_{i_N=1}^{q}p_{i_N} \qquad \sum_{i_k=1}^{q}p_{i_k}=1$$

$$=\sum_{i_1=1}^{q}p_{i_1}\log\frac{1}{p_{i_1}}=\sum_{X}p_{i_1}\log\frac{1}{p_{i_1}}=H\left(X\right)$$

$$H\left(X^N\right)=H\left(X\right)+H\left(X\right)+\cdots+H\left(X\right)=NH\left(X\right)$$

返回

条件熵随N的增加是非递增的

返回

$$\boxed{H\left(X_N \middle| X_1 X_2 \cdots X_{N-1}\right) \le H\left(X_{N-1} \middle| X_1 X_2 \cdots X_{N-2}\right)}$$

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

$$\Rightarrow H(Z \mid XY) \le H(Z \mid X)$$

$$H(X_N \mid X_1 X_2 \cdots X_{N-1}) \le H(X_N \mid X_2 \cdots X_{N-1})$$

由序列的平稳性可得:

$$H(X_N \mid X_2 \cdots X_{N-1}) = H(X_{N-1} \mid X_1 X_2 \cdots X_{N-2})$$

返回

$$H_N(X) \geq H(X_N \mid X_1 X_2 \cdots X_{N-1})$$

$$\begin{aligned} NH_N(X) &= H(X_1 X_2 \cdots X_{N-1} X_N) \\ &= H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_2 X_1) \\ &+ \cdots + H(X_N \mid X_1 X_2 \cdots X_{N-1}) \\ &= H(X_N) + H(X_N \mid X_{N-1}) + H(X_N \mid X_{N-1} X_{N-2}) \\ &+ \cdots + H(X_N \mid X_1 X_2 \cdots X_{N-1}) \\ &\geq NH(X_N \mid X_1 X_2 \cdots X_{N-1}) \end{aligned}$$

$$\boxed{\because \quad H(X_N \mid X_1 X_2 \cdots X_{N-1}) \leq H(X_{N-1} \mid X_1 X_2 \cdots X_{N-2})}$$

返回

$$\boxed{H_N(X) \le H_{N-1}(X)}$$

$$\begin{aligned}
NH_N(X) &= H(X_1X_2\cdots X_{N-1}X_N) \\
&= H(X_N \mid X_1X_2\cdots X_{N-1}) + H(X_1X_2\cdots X_{N-1}) \\
&= H(X_N \mid X_1X_2\cdots X_{N-1}) + (N-1)H_{N-1}(X) \\
&\le H_N(X) + (N-1)H_{N-1}(X)
\end{aligned}$$

$$\therefore \quad (N-1)H_N(X) \le (N-1)H_{N-1}(X)$$

状态转移矩阵的变化过程

$$P=\begin{bmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}$$

0	0.5	0	0.5
0.5	0	0.5	0
0	0.5	0	0.5
0.5	0	0.5	0

0.5	0	0.5	0
0	0.5	0	0.5
0.5	0	0.5	0
0	0.5	0	0.5

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

0.5000	0.5000	0
0.2500	0.1944	0.5556
0.2500	0.1667	0.5833

0.2500	0.1667	0.5833
0.3750	0.3426	0.2824
0.3750	0.3472	0.2778

0.3750	0.3472	0.2778
0.3125	0.2554	0.4321
0.3125	0.2546	0.4329

0.3125	0.2546	0.4329
0.3438	0.3012	0.3551
0.3438	0.3013	0.3549

0.3438	0.3013	0.3549
0.3281	0.2779	0.3939
0.3281	0.2779	0.3940

0.3281	0.2779	0.3940
0.3359	0.2896	0.3744
0.3359	0.2896	0.3744

0.3359	0.2896	0.3744
0.3320	0.2838	0.3842
0.3320	0.2838	0.3842

0.3320	0.2838	0.3842
0.3340	0.2867	0.3793
0.3340	0.2867	0.3793

状态转移矩阵的变化过程（续）

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

0.3320	0.2838	0.3842
0.3340	0.2867	0.3793
0.3340	0.2867	0.3793

$$\begin{cases} \sum_{i=1}^{3} W_i p_{ij} = W_j \\ \sum_{j=1}^{3} W_j = 1 \end{cases} \Rightarrow \begin{cases} 0\cdot W_1 + 1/2\cdot W_2 + 1/2\cdot W_3 = W_1 \\ 0\cdot W_1 + 1/3\cdot W_2 + 1/2\cdot W_3 = W_2 \\ 1\cdot W_1 + 1/6\cdot W_2 + \quad 0\cdot W_3 = W_1 \\ W_1 + \quad W_2 + \quad W_3 = 1 \end{cases}$$

$$W_1 = 1/3 \qquad W_2 = 2/7 \qquad W_3 = 8/21$$

$$W_1 = 0.333 \qquad W_2 = 0.2857 \qquad W_3 = 0.3810$$

返回

切普曼-柯尔莫哥洛夫方程 – 证明

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

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

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

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

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

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

返回

本章作业

- 作业要求：先定义概率空间；必须有过程
- 作业题
  1. 课件上补充的三道练习题、2、4
  2. 5、7、13
  3. 17、18、19、20