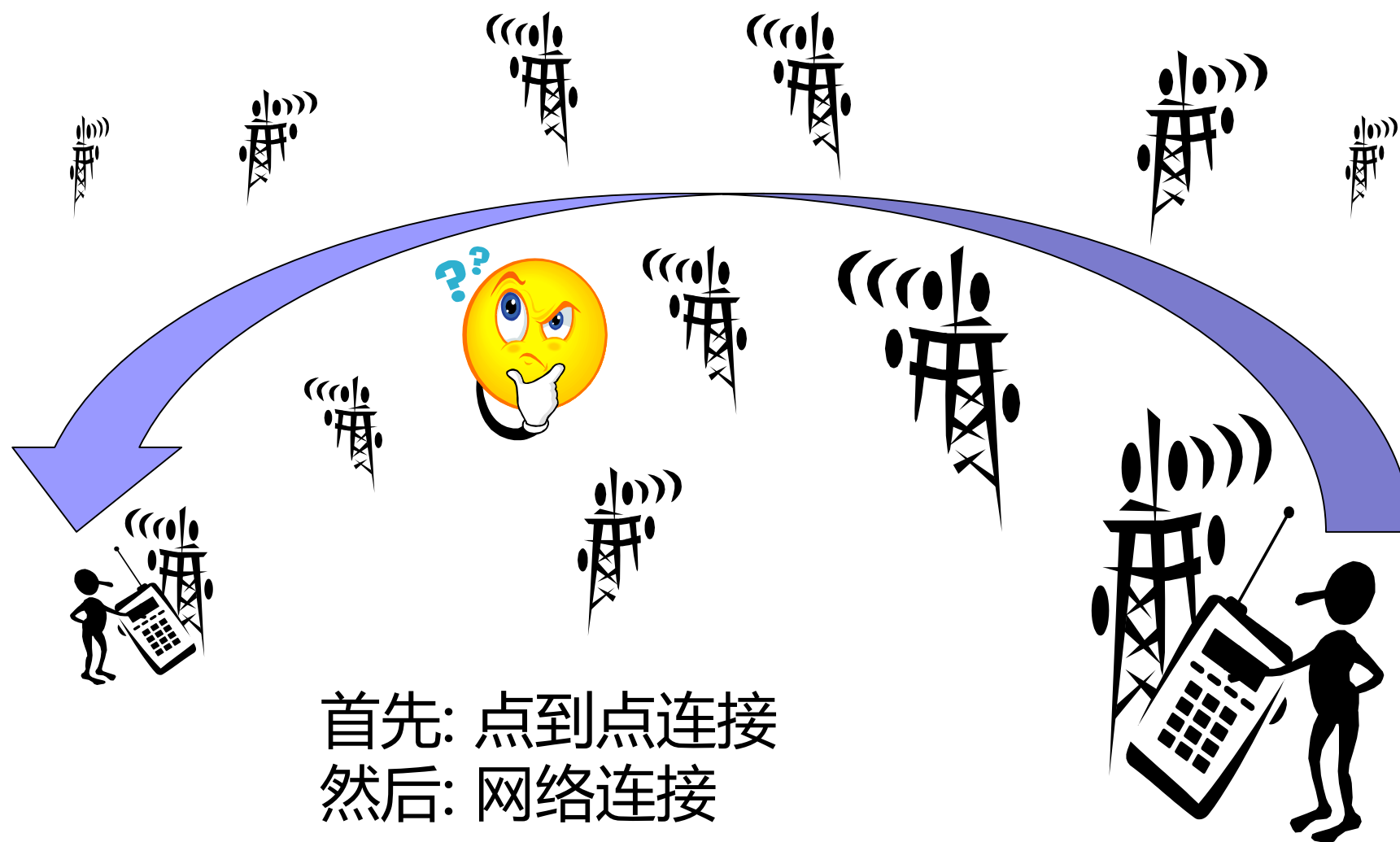


# 基础信息论

## 多符号离散平稳信源

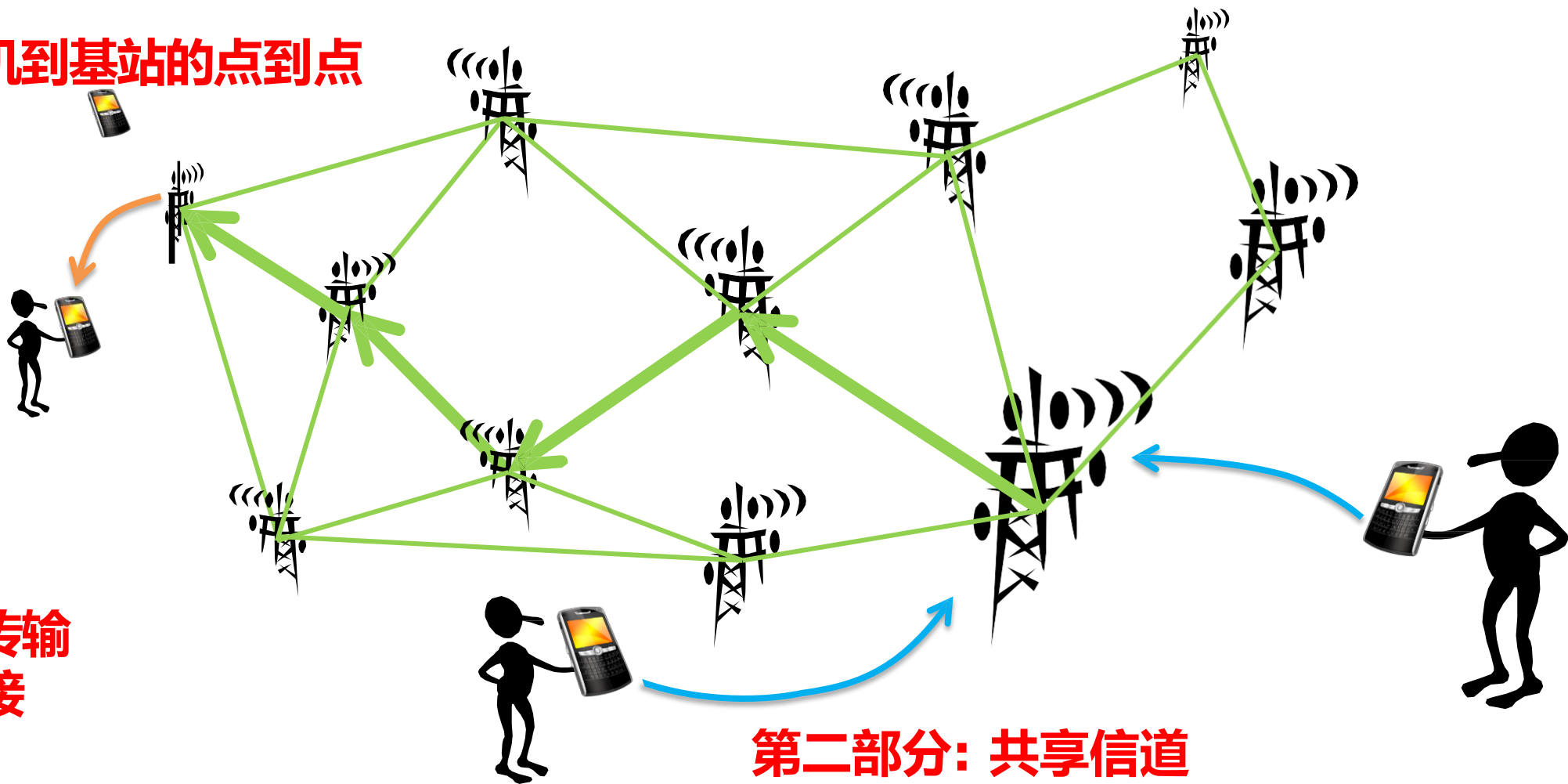
华中科技大学电信学院

# 将信息从一个端点通过无线网络传输到另一个端点



# 怎样通过无线网络传输信息

**第一部分：从手机到基站的点到点连接**

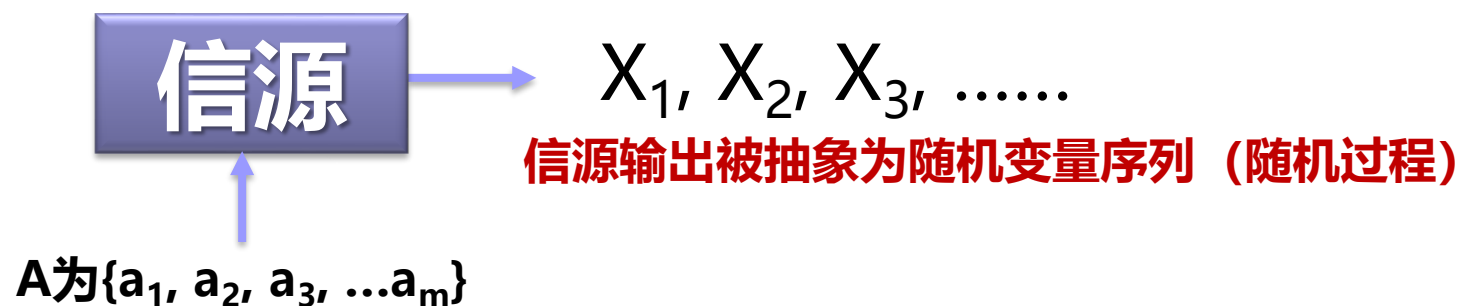


**第三部分：网络传输  
--基站之间的连接**

**第二部分：共享信道  
--允许多个用户同时传输**



# 多符号离散平稳信源



- 输出的随机变量取值于某一离散符号集合，消息在时间和幅值上均是离散的。
- 信源输出的消息常常是以一个个符号的形式出现，例如文字、字母等，这些符号的取值是有限的或可数的。
- 比如：平面图像  $X(x, y)$ 和电报、书信、文稿等等

# 学习目标

- 构建多符号离散平稳信源的数学模型
- 计算多符号离散平稳信源的信息熵
- 分析多符号离散平稳信源的信息熵性质

阅读：陈运，信息论与编码（第3版）第2章2.3节

# 多符号离散平稳信源

## 1. 多符号离散信源

- 实际信源输出的消息是**时间**上和**空间**上的一系列符号，每一位出现的符号是随机的，用随机矢量或随机变量序列来表示：

$$X = X_1 X_2 X_3 \cdots X_i \cdots X_j \cdots X_N$$

- **一般地**，不同时刻随机变量的概率分布不同： $P(X_i)$  和  $P(X_{i+r})$  不同，即随机变量的统计特性随着时间的推移而变化。
- 一般前后符号之间是有**统计依赖**关系的。

## 2. 平稳性

随机矢量中的各随机变量的各维联合概率分布均不随时间的推移而变化，信源所发符号序列的概率分布与时间起点无关。

多符号离散信源 + 平稳 → 多符号离散平稳信源

# 离散平稳无记忆信源

## ■ 单符号信源：用一维离散随机变量描述

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

其中  $p(x_i) \geq 0 \quad i = 1, 2, \dots, n$

$$\sum_{i=1}^n p(x_i) = 1$$

例：若  $n = 2$ ，则信源输出符号为“0”和“1”，  
设  $P(X = 0) = p, P(X = 1) = \bar{p}, p + \bar{p} = 1$

此时，信源模型为  $\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ p & \bar{p} \end{bmatrix}$

# 离散平稳无记忆信源

## ■ 定义：

- 假定随机变量序列的长度是有限的，信源输出的消息序列中，符号之间**无相互依赖关系**，这类信源称**离散平稳无记忆信源**
- 又称：**单符号离散平稳无记忆信源的扩展信源**  
序列长度就是**扩展次数**。

## ■ 样例：离散无记忆信源的二次扩展信源

$$X = X_1 X_2$$

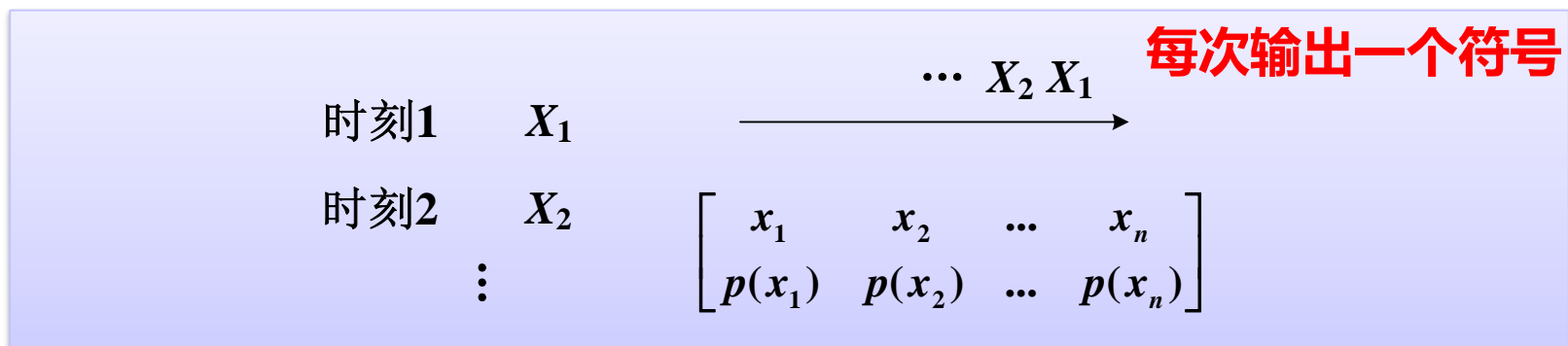
- 输出消息是符号序列，分组发出，每两个符号构成一组  
若单符号信源为二进制信源{0, 1}，经二次扩展，变成了：  
{00, 01, 10, 11}



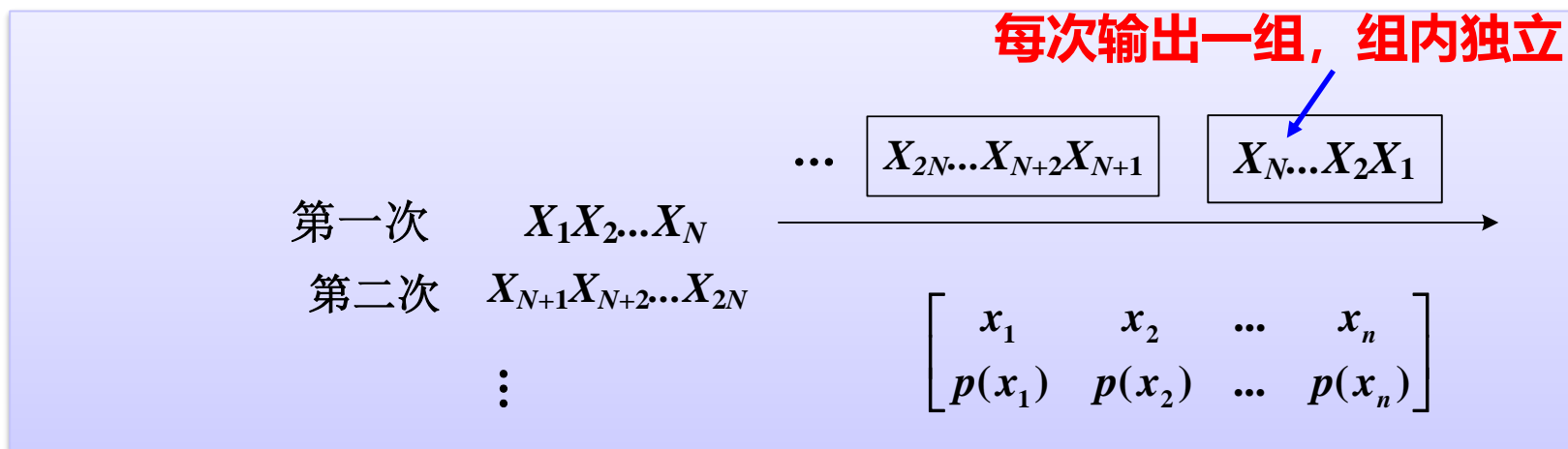
# 离散平稳无记忆信源的N次扩展

离散平稳无记忆信源:

离散单符号信源



以此为基础进行N次扩展, 得:



# ~N次扩展信源的数学模型:

## 单符号信源模型

## 扩展得到多符号信源模型

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \\ p(x_1) & p(x_2) & \dots & p(x_n) \end{bmatrix} \rightarrow \begin{bmatrix} X^N \\ P(X^N) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \dots & a_q \\ p(a_1) & p(a_2) & \dots & p(a_q) \end{bmatrix}$$

模型中:

$$p(a_i) = p(x_{i_1} x_{i_2} \dots x_{i_N})$$

$$i_1, i_2, \dots, i_N \in \{1, 2, \dots, n\}$$

$$p(a_i) = p(x_{i_1}) p(x_{i_2}) \dots p(x_{i_N})$$

无记忆性

存在结论:  $H(X^N) = N \cdot H(X)$  \*

接下来进行证明。

记  $q = n^N$

$$a_1 = x_1 x_1 \dots x_1$$

$$a_2 = x_1 x_1 \dots x_2$$

.....

$$a_n = x_1 x_1 \dots x_n$$

$$a_{n+1} = x_1 x_1 \dots x_2 x_1$$

.....

$$a_{2n} = x_1 x_1 \dots x_2 x_n$$

.....

$$a_q = x_n x_n \dots x_n$$

# N次扩展信源的熵(序列信息的熵)

已知离散平稳无记忆信源的N次扩展信源:

$$\begin{bmatrix} X^N \\ P(X^N) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \dots & a_q \\ p(a_1) & p(a_2) & \dots & p(a_q) \end{bmatrix} \quad \begin{aligned} p(a_i) &= p(x_{i_1} x_{i_2} \dots x_{i_N}) \\ &= p(x_{i_1}) p(x_{i_2}) \dots p(x_{i_N}) \end{aligned}$$

求证:

$$H(X^N) = N \cdot H(X)$$

证: 由信源熵的定义, 可得:

$$\begin{aligned} H(X^N) &= - \sum_{i=1}^q p(a_i) \cdot \log p(a_i) = - \sum_{i_1=1}^n \sum_{i_2=1}^n \dots \sum_{i_N=1}^n p(a_i) \cdot \log p(x_{i_1}) p(x_{i_2}) \dots p(x_{i_N}) \\ &= - \sum_{i_1=1}^n \sum_{i_2=1}^n \dots \sum_{i_N=1}^n p(a_i) \cdot \log p(x_{i_1}) - \dots - \sum_{i_1=1}^n \sum_{i_2=1}^n \dots \sum_{i_N=1}^n p(a_i) \cdot \log p(x_{i_N}) \end{aligned}$$

# 证明 (续) :

以式中的第一项为例进行分析 $X^N$

$$\begin{aligned}
 & - \sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{i_N=1}^n p(a_i) \cdot \log p(x_{i_1}) \\
 &= - \sum_{i_1=1}^n \sum_{i_2=1}^n \cdots \sum_{i_N=1}^n p(x_{i_1}) p(x_{i_2}) \cdots p(x_{i_N}) \cdot \log p(x_{i_1}) \\
 &= - \sum_{i_1=1}^n p(x_{i_1}) \cdot \log p(x_{i_1}) \cdot \sum_{i_2=1}^n p(x_{i_2}) \cdots \sum_{i_N=1}^n p(x_{i_N}) \\
 &= - \sum_{i_1=1}^n p(x_{i_1}) \cdot \log p(x_{i_1}) \cdot 1 \cdots 1 = H(X)
 \end{aligned}$$

同样计算前式中其余 $N-1$ 项, 可得:

$$H(X^N) = H(X) + H(X) + \dots + H(X) = NH(X)$$

即:

$$\therefore H(X^N) = N \cdot H(X)$$

意义:

离散无记忆信源的**N次扩展**信源每输出一个消息符号 (即符号序列) 所提供的信息熵是信源 $X$ 每输出一个消息符号所提供熵的**N倍**。

# 例题1

- 有一离散平稳无记忆信源

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}$$

求该信源二次扩展信源的熵。

解：首先求二次扩展信源的概率分布

$$\begin{bmatrix} X^2 \\ P(X^2) \end{bmatrix} = \begin{matrix} & \begin{matrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 & a_9 \end{matrix} \\ \begin{bmatrix} x_1x_1 & x_1x_2 & x_1x_3 & x_2x_1 & x_2x_2 & x_2x_3 & x_3x_1 & x_3x_2 & x_3x_3 \end{bmatrix} \\ \begin{matrix} 1/4 & 1/8 & 1/8 & 1/8 & 1/16 & 1/16 & 1/8 & 1/16 & 1/16 \end{matrix} \end{matrix}$$

**方法1：为单符号信源的熵的2倍**

$$\begin{aligned} H(X^2) &= 2 \cdot H(X) = 2 \cdot \left[ -\frac{1}{2} \log \frac{1}{2} - 2 \cdot \frac{1}{4} \log \frac{1}{4} \right] \\ &= 3 \text{ 比特/符号} \end{aligned}$$

## 例题1 (续)

方法2: 由定义直接计算多符号信源的熵

$$\begin{aligned} H(X^2) &= -\frac{1}{4} \log \frac{1}{4} - 4 \cdot \frac{1}{8} \log \frac{1}{8} - 4 \cdot \frac{1}{16} \log \frac{1}{16} \\ &= 3 \text{ 比特/符号} \end{aligned}$$

结论:

计算扩展信源的熵时, 不必构造新的信源, 可直接从原信源  $X$  的熵导出, 即一个离散平稳无记忆信源  $X$  的  $N$  次扩展信源的熵等于信源  $X$  熵的  $N$  倍。

注意:

单位中的符号是什么?

扩展后的消息符号, 即符号序列

$x_1 x_1 \quad x_1 x_2 \quad \dots$

## 例题2

- 为使电视图像获得良好清晰度和对比度，每帧需要用50万个像素和10个不同亮度电平。并设每秒要传送 30帧图像，所有像素是独立变化的，且所有亮度电平等概率出现，求传递此图像所需的信息率（比特/秒）。
- 思路：像素点/秒×比特/像素点
- 解：(1) 每秒需要传送的像素点数 =  $30 \text{ 帧/秒} \cdot 5 \times 10^5 \text{ 像素/帧}$   
 $= 1.5 \times 10^7 \text{ 像素/秒}$
- 每像素点提供的信息量 =  $\log 10 \approx 3.32 \text{ 比特/像素}$
- 由于所有像素独立变化，信源为无记忆信源的扩展信源。
- 信息率 =  $1.5 \times 10^7 \text{ 像素/秒} \cdot 3.32 \text{ 比特/像素} \approx 4.98 \times 10^7 \text{ 比特/秒}$

## 例题2(续)

- (2) 设某彩电系统，除了满足对于黑白电视系统的上述要求外，还必须有30个不同的色彩度，再求传递图像所需的信息率。
- 每像素点提供的信息量 =  $\log 30 + \log 10 \approx 8.23$  比特/像素
- 信息率 =  $1.5 \times 10^7 \frac{\text{像素}}{\text{秒}} \cdot 8.23 \frac{\text{比特}}{\text{像素}} \approx 12.35 \times 10^7$  比特/秒



# 离散平稳信源的数学模型

# 离散平稳信源

- 实际信源常不是简单无记忆信源，符号之间具有**统计关联关系**，常用**联合概率**来描述符号间的相互依存关系。下面引入**平稳信源**概念。
- 定义：**若信源产生的随机序列 $X_i, i = 1, 2, \dots$  满足：

1. 所有 $X_i, i = 1, 2, \dots$ 都取值于**有限**的信源符号集

$$A = (x_1, x_2, \dots, x_n)$$

2. 随机序列是**平稳**的，即对所有非负整数 $i_1, i_2, \dots, i_N, h$ 有

$$P\{X_{i_1}, X_{i_2}, \dots, X_{i_N}\} = P\{X_{i_1+h}, X_{i_2+h}, \dots, X_{i_N+h}\}$$

则称此信源为**离散平稳信源**。

平稳信源发出的符号序列的**概率分布与时间起点无关**，即，平稳信源发出的符号序列的概率分布函数可以平移。

# 一维平稳和二维平稳

- 对于随机变量序列

$$\mathbf{X} = X_1 X_2 \dots X_N$$

- 一维平稳

- 任意两不同时刻，信源发出的消息概率分布完全相同

$$P(X_i = x_1) = P(X_j = x_1) = p(x_1)$$

$$P(X_i = x_2) = P(X_j = x_2) = p(x_2)$$

...

$$P(X_i = x_n) = P(X_j = x_n) = p(x_n)$$

- 一维平稳信源任何时刻均以相同概率分布发出符号

- 二维平稳

- 一维平稳信源，如果二维联合概率分布也**与时间起点无关**，即

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

- 二维平稳信源在任何时刻发出的两个符号的概率完全相同

# 离散平稳信源

- 如果各维联合概率分布均与时间起点无关，即对两个不同的时刻*i*和*j*，有

- 一维平稳

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

- 二维平稳

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

...

N+1维平稳

$$P(X_i X_{i+1} X_{i+2} \dots X_{i+N}) = P(X_j X_{j+1} X_{j+2} \dots X_{j+N})$$

...

各维联合概率均与时间起点无关的完全平稳信源成为**离散平稳信源**。

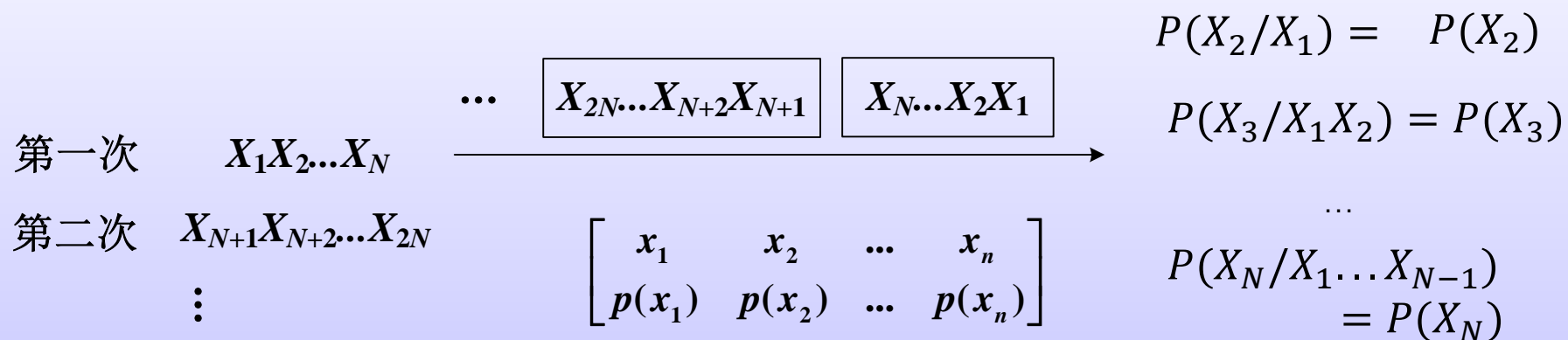
# 离散平稳信源的信息熵和极限熵

# 离散平稳有记忆信源

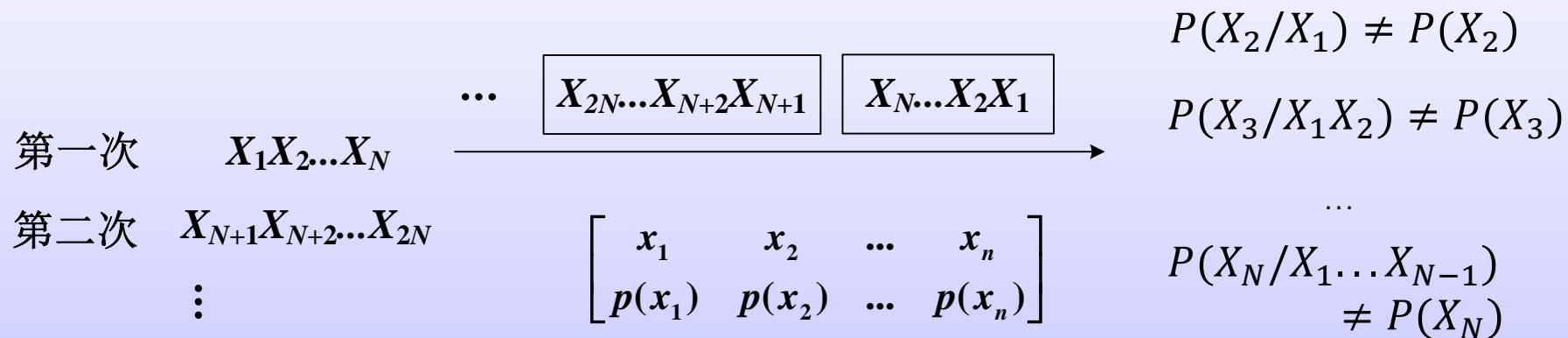
- 离散平稳信源一般是**有记忆**信源，即发出的各个符号之间具有统计关联关系的一类信源
- **统计关联性**（反映信源记忆特性）可用两种方式表示：
  1. 用**联合概率**反映信源记忆特性**有记忆信源(2.3.3)**  
信源发出的一个符号序列的**整体概率**(即**N个符号的联合概率**)
  2. 用**条件概率**反映信源记忆特性**马尔可夫信源(2.3.4)**  
用信源发出符号序列中**各个符号之间的条件概率**

# 离散平稳有记忆信源

~的 $N$ 次扩展信源:



有记忆信源:



# 平稳信源的熵-联合熵

- 以最简单的二维平稳信源为例,它是N长为2的有记忆平稳信源。
- 假设信源符号序列组之间相互独立
- 信源  $X = X_1X_2$ , 数学模型为

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} x_1x_1 & x_1x_2 & \dots & x_nx_n \\ p(x_1x_1) & p(x_1x_2) & \dots & p(x_nx_n) \end{bmatrix}$$

- 由熵的定义, 可以得到信源的联合熵为

$$H(X) = H(X_1X_2) = - \sum_{i=1}^n \sum_{j=1}^n p(x_ix_j) \log p(x_ix_j)$$



# 平稳信源的熵-熵的可加性

- 一般信源发出的符号序列中，前后符号之间总是存在着依赖关系，此时，可证信息熵存在关系式：
$$H(X_1, X_2) = H(X_1) + H(X_2|X_1)$$
- 熵的可加性

## 说明：

- 两个有相互依赖关系的随机变量  $X_1$  和  $X_2$  所组成的随机矢量  $X = X_1X_2$  的联合熵  $H(X)$  等于第一个随机变量的熵  $H(X_1)$  与第一个随机变量  $X_1$  已知的前提下，第二个随机变量  $X_2$  的条件熵  $H(X_2 / X_1)$  之和。

- 当前后序列符号没有依赖关系时（即  $X_1$  和  $X_2$  相互独立）有：

$$H(X_1, X_2) = H(X_1) + H(X_2)$$

# 熵可加性的分析

观察

$$H(X) = H(X_1) + H(X_2/X_1)$$

□ 由于条件熵小于无条件熵  $H(X_2/X_1) \leq H(X_2)$

□ 所以

$$H(X) \leq H(X_1) + H(X_2)$$

上式说明：二维离散平稳**有记忆**信源的熵**小于等于**二维离散平稳**无记忆**信源的熵

- 对于二维离散平稳无记忆信源， $X_1$ 对 $X_2$ 不产生任何影响。
- 对于二维离散平稳有记忆信源，由于 $X_1$ 和 $X_2$ 有统计依赖关系， $X_1$ 的发生会提供 $X_2$ 的部分相关信息。

# 二维平稳信源 vs 二次扩展信源

- 若 $X_1$ 和 $X_2$ 取值于同一集合:

$$H(X_1) = H(X_2) = H(X)$$

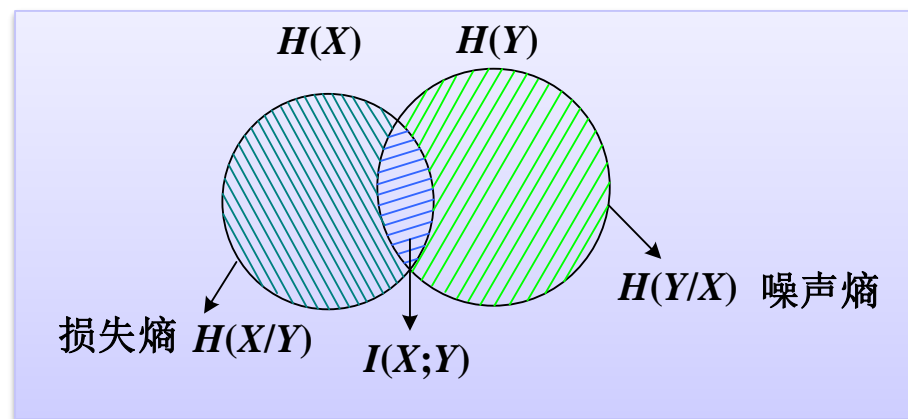
- ~的2次扩展信源:

$$H(X_1X_2) = H(X_1) + H(X_2) = 2H(X)$$

- 有记忆信源:

$$H(X_1X_2) = H(X_1) + H(X_2/X_1) \leq H(X_1) + H(X_2)$$

- 维拉图



- 物理意义

- 分组长度为2的有记忆信源, 其信息量等于第一个符号单独提供的信息量, 加上已知第一个符号时第二个符号又提供的信息量。
- 由于第一个符号多少提供了关于第二个符号的一些信息, 因此其小于无记忆信源的熵。

# 例题1

设某二维离散信源 $X_1X_2$   
的原始信源 $X$ 为：

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{Bmatrix} x_1 & x_2 & x_3 \\ \frac{1}{4} & \frac{4}{9} & \frac{11}{36} \end{Bmatrix}$$

前后两个符号的条件概率如右表所示：

试计算 $H(X_1X_2)$ ，并 $H(X^2)$ 进行比较。

| $P(X_2/X_1) \backslash X_2$<br>$X_1$ | $x_1$ | $x_2$ | $x_3$ |
|--------------------------------------|-------|-------|-------|
| $x_1$                                | 7/9   | 2/9   | 0     |
| $x_2$                                | 1/8   | 3/4   | 1/8   |
| $x_3$                                | 0     | 2/11  | 9/11  |

解：  $H(X_1X_2) = H(X_1) + H(X_2/X_1)$

$$H(X_1) = -\frac{1}{4} \cdot \log \frac{1}{4} - \frac{4}{9} \cdot \log \frac{4}{9} - \frac{11}{36} \cdot \log \frac{11}{36}$$

$$= 1.542 \text{ 比特/符号}$$

# 例题1(续)

计算联合概率:

$$\begin{aligned} p(x_1x_1) &= p(x_1) \cdot p(x_1/x_1) &= \frac{1}{4} \cdot \frac{7}{9} = \frac{7}{36} \\ p(x_1x_2) &= p(x_1) \cdot p(x_2/x_1) &= \frac{1}{4} \cdot \frac{2}{9} = \frac{1}{18} \end{aligned}$$

.....

$$\begin{aligned} H(X_2/X_1) &= -p(x_1x_1) \log p(x_1/x_1) - \dots - p(x_3x_3) \log p(x_3/x_3) \\ &= 0.87 \quad \text{比特/符号} \end{aligned}$$

$$\therefore H(X_1X_2) = H(X_1) + H(X_2/X_1) = 1.542 + 0.87 = 2.412 \quad \text{比特/符号}$$

$$H(X^2) = 2 \cdot H(X_1) = 2 \cdot 1.542 = 3.084 \quad \text{比特/符号}$$

$x_1x_1$   
 $x_1x_2$   
...

# N维平稳有记忆信源的熵

可将上述结论推广到分组长度为 $N$ 的任意情况：

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

证明：为方便，记  $Y_1 = X_1 X_2 \cdots X_{N-1}$        $Y_2 = X_1 X_2 \cdots X_{N-2}$

$$\begin{aligned} & \dots\dots\dots \\ H(X) &= H(Y_1 X_N) = H(Y_1) + H\left(\frac{X_N}{Y_1}\right) = \underbrace{H(X_1 X_2 \cdots X_{N-1})} + H(X_N/X_1 X_2 \cdots X_{N-1}) \\ &= [H(Y_2) + H(X_{N-1}/Y_2)] + H(X_N/X_1 X_2 \cdots X_{N-1}) \\ &= \underbrace{H(X_1 X_2 \cdots X_{N-2})} + H(X_{N-1}/X_1 X_2 \cdots X_{N-2}) + H(X_N/X_1 X_2 \cdots X_{N-1}) \\ &= \dots\dots\dots \\ &= H(X_1) + H(X_2/X_1) + H(X_3/X_1 X_2) + \cdots + H(X_N/X_1 X_2 \cdots X_{N-1}) \end{aligned}$$

## 说明:

1. 多符号离散平稳有记忆信源的熵是X起始时刻随机变量X1的熵与各阶条件熵之和。
2. 由于信源的平稳，信源熵与起始时刻无关，是一个**固定不变**的值。
3. 可以证明，条件熵是非递增的，即

$$\begin{aligned} H(X_N | X_1 X_2 \cdots X_{N-1}) &\leq H(X_{N-1} | X_1 \cdots X_{N-2}) \\ &\leq \cdots \leq H(X_2 | X_1) \leq H(X_1) \end{aligned}$$

当 $N = 3$ 时有,  $H(X_3 | X_1 X_2) \leq H(X_2 | X_1) \leq H(X_1)$

# 平均符号熵与极限熵

- 信源的矢量熵（或联合熵）： $H(X_1 X_2 \cdots X_N)$

信源平均每发出一个消息（N个符号）所提供的信息量

- 平均符号熵：  
每组符号中，平均每个符号所提供信息量。

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

- 极限熵：  
当分组长度N趋于无穷大时的平均符号熵。

$$H_\infty = \lim_{N \rightarrow \infty} H_N = \lim_{N \rightarrow \infty} \frac{1}{N} \cdot H(X_1 X_2 \cdots X_N)$$



# 极限熵的意义

**问题1：**为什么要引出极限熵的概念？

**回答：**

- 多符号信源**假设**前后组之间**独立**，与实际信源的情况不相符合。
- 信源在不断地发出符号，符号之间的统计关联关系也并不仅限于长度 $N$ ，而是伸向无穷远。**极限熵**最真实地反映了信源的实际情况。
- 故研究实际信源，必须求出信源的极限熵，能表示多符号离散平稳有记忆信源**平均每发一个符号的信息量**。

# 极限熵的计算

**问题2：**极限熵是否一定存在？

**回答：**可以证明，对任意离散**平稳**信源，极限熵是存在的。

$$H_{\infty} = \lim_{N \rightarrow \infty} H_N(X) = \lim_{N \rightarrow \infty} H(X_N | X_1, X_2, \dots, X_{N-1})$$

**问题3：**如何才能求出极限熵？

**回答：**

1. 必须求出信源的**无穷维**的联合概率和条件概率的分布。
2. 相当困难。

# 极限熵的计算

转化为:  $H_{\infty} = \lim_{N \rightarrow \infty} H(X_N / X_1 X_2 \cdots X_{N-1})$

但实际上对上述条件熵的计算仍需要得到无穷维的联合概率分布, 问题并未得到简化。

将实际信源近似为  $N$  为有限值的多符号信源:

只需要得到有限维的联合概率分布即可。

某些类型信源, 实际记忆长度为有限值:

可能得到极限熵的真实值。如: 马尔可夫信源

# 总结

- 通过扩展单符号信源，可分析多符号离散平稳信源
- 极限熵刻画多符号离散平稳有记忆信源平均每发一个符号的信息量

谢谢!

黑晓军

华中科技大学

电子信息与通信学院

Email: [heixj@hust.edu.cn](mailto:heixj@hust.edu.cn)

网址: <http://eic.hust.edu.cn/aprofessor/heixiaojun>

## 参考资料

- 陈运, 信息论与编码 (第3版) 第2章2.3节, 电子工业出版社出版, 2015