

信源是信息的来源，而通信系统的目的就是传输信息，因此对信源所提供信息特性的研究显得非常重要。

信源研究的主要问题包括：

(1) **信源的形态问题**：主要讨论信源的数学建模方法，即如何利用数学方法描述信源和讨论信源性质。

(2) **信源信息的表达问题**：主要讨论信源编码问题，即如何利用消息有效地表达信源所发出的信息。

本章主要介绍信源的形态问题，也就是信源的数学建模方法。如前所述，信源所包含的信息是通过消息来承载的，因此研究信源特性就是研究信源所发送消息的特性。毫无疑问，消息是随机变量，是会随着时间（或者空间）不同而变化的一类随机变量。综合消息的上述特性，可以利用随机过程 $X(t)$ 建模和研究消息，进而得到信源特性。

3.1 随机过程基本知识

研究概率论和离散信息的统计度量时，常使用随机事件这一概念。例如，抛硬币中正面朝上是随机事件，掏球活动中掏出红球是随机事件，等等。

现实的世界以时间和空间为坐标轴，任何事件（包括随机事件）都是在某一时间点的特定空间中发生的，因此描述随机性现象更为一般的概念是**随机过程**。简而言之，随机过程是依赖参数的一族随机变量的全体。由于参数通常是指时间，因此**随机过程**这一概念中有**过程**一词就容易理解了。本节简单回顾随机过程基础知识。

3.1.1 概率空间

随机过程的数学定义建立在**概率空间**概念之上，因此先回顾概率空间的含义。

定义 3.1 概率空间

设随机试验为 E ，其对应的样本空间为 Ω 。 \mathcal{F} 是 Ω 的所有子集组成的集族（ Ω 上所有随机事件（包括空集）组成的事件集合）， P 为事件集合 \mathcal{F} 所对应的概率， $P(A)$ 为事件 A 的概率。称 (Ω, \mathcal{F}, P) 组成**概率空间**。

定义中所述 \mathcal{F} 是 Ω 中所有子集构成的集族, 表明 \mathcal{F} 是 Ω 所能生成的所有事件 (包括空集), 见例 3.1。

例3.1 抛硬币。

小学生春风和同学楠楠抛硬币猜正反面。

求概率空间定义中的集族 \mathcal{F} 。

解: 根据题意, 样本空间 $\Omega = \{H, T\}$ 。 Ω 的所有子集为

$$\emptyset \quad \{H\} \quad \{T\} \quad \{HT\}$$

根据随机事件的定义 (参见定义 2.1), $\emptyset, \{H\}, \{T\}, \{HT\}$ 都是由样本空间 Ω 所派生出的随机事件, 因此集族 $\mathcal{F} = \{\emptyset, \{H\}, \{T\}, \{HT\}\} = \{\emptyset, \{H\}, \{T\}, \Omega\}$ 实际就是由样本空间 Ω 派生的、包括空事件在内的所有事件组成的事件集。

例3.2 掷骰子。

小学生春风和楠楠掷骰子猜点数。

求集族 Ω 。

解: 根据题意, 样本空间 $\Omega = \{\omega_i, i = 1, 2, \dots, 6\}$ 。对应于样本空间 Ω 的基本事件为

$$e_i = \{\omega_i\} = \{i\}$$

根据上述的基本事件, 可以构造以下联合事件:

$$e_{i_1, i_2} = e_{i_1} \cup e_{i_2}$$

$$e_{i_1, i_2, i_3} = e_{i_1} \cup e_{i_2} \cup e_{i_3}$$

$$\vdots$$

$$e_{i_1, \dots, i_6} = e_{i_1} \cup \dots \cup e_{i_6}$$

$$i_1, i_2, \dots, i_6 \in \{1, 2, \dots, 6\}, \text{ 且互不相等}$$

则有

$$\mathcal{F} = \left\{ \emptyset, e_{i_1}, e_{i_1, i_2}, \dots, \underbrace{e_{i_1, i_2, \dots, i_6}}_{\Omega} \right\} = \left\{ \emptyset, \Omega, e_{i_1}, e_{i_1, i_2}, \dots, e_{i_1, i_2, \dots, i_5} \right\}$$

3.1.2 随机过程

1. 随机过程的定义

定义3.2 随机过程

设 (Ω, \mathcal{F}, P) 为一概率空间, T 为一实数参数集。 $X(\omega, t)$ 为定义在 Ω 和 T 上的二元函数。如果对于任意的 $t \in T, \omega \in \Omega$, $X(\omega, t)$ 是 (Ω, \mathcal{F}, P) 上的随机变量, 进而 $\{X(\omega, t), \omega \in \Omega, t \in T\}$ 为一簇随机变量, 则称这一随机变量簇 $\mathbf{X} = \{X(\omega, t), t \in T,$

$\omega \in \Omega$ 为概率空间 (Ω, \mathcal{F}, P) 上的一个随机过程，简记为 $X(t), t \in T$ 或 $X(t)$ 或 $X_t(\omega)$ 。

2. 随机过程的含义

$X(\omega, t)$ 是 (Ω, \mathcal{F}, P) 上的随机变量。根据随机变量的定义（参见定义 2.6），随机变量是样本点 $\omega \in \Omega$ 到实数的一种映射，而 $X(\omega, t)$ 定义为 Ω 和 T 上的二元函数，也是一种映射，只不过自变量是二元的，所以 $X(\omega, t)$ 是一随机变量。

$\{X(\omega, t), \omega \in \Omega, t \in T\}$ 为一簇随机变量。当自变量 ω 固定时，自变量 t 取值于集合 T ， t 值不同，所对应的随机变量 $X(\omega, t)$ 取值不同，构成一簇随机变量，参见图 3.1。

t 固定和 ω 固定， $X(\omega, t)$ 取固定值，表示一随机事件；

t 固定和 ω 可变， $X(\omega, t)$ 表示样本空间 Ω 上的随机变量；

t 可变和 ω 固定， $X(\omega, t)$ 表示一确定的时间函数；

t 可变和 ω 可变， $X(\omega, t)$ 当然就是随机过程。

状态空间：随机过程 $X(\omega, t)$ 的取值所组成的集合 S 。自然有 $S \subset R$ 。

样本函数： ω 固定情况下的函数 $X(\omega, t)$ ，又称样本轨迹。

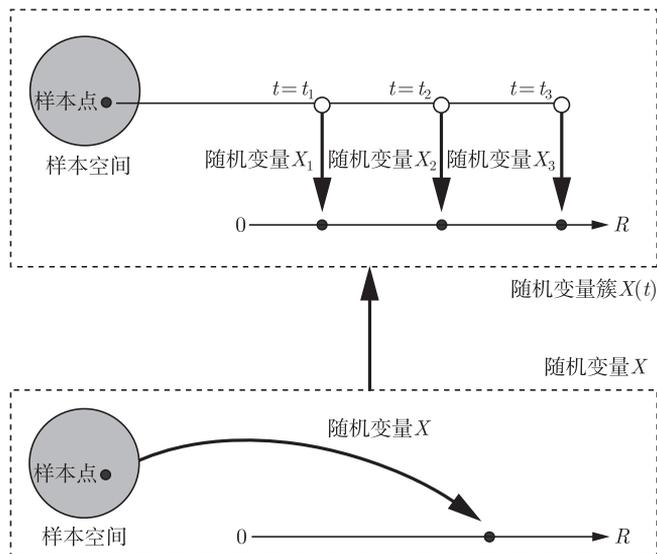


图 3.1 随机变量与随机过程

例 3.3 抛掷硬币。

小学生春风以每隔 $T_0 = 1\text{s}$ 掷一次硬币，并记录抛掷结果。

解：抛掷结果 $X(t)$ 是时间 t 的函数， $X(t)$ 是一随机过程，其映射关系：

样本空间： $\Omega = \{H : \text{正面朝上}, T : \text{反面朝上}\}$

$$\text{随机过程: } X(t) = \begin{cases} -1, & \text{第 } n \text{ 次抛掷结果为 } H \\ 1, & \text{第 } n \text{ 次抛掷结果为 } T \end{cases} \quad (n-1)T_0 \leq t \leq nT_0$$

此随机过程的示意图见图 3.2 (a)。

例3.4 随机过程举例：喜鹊种群。

小学生春风和楠楠接受了一项调查任务：了解小区喜鹊种群数量的变化。

解：种群数量是动态变化的，在某一时间 t ，喜鹊种群数量 $X(t)$ 为一随机变量；如果考察种群数量的变换，则需要从 $t = 0$ 开始每隔一定时间（如 24h）观察喜鹊种群的数量，得到的 $X(t), t = 0, 24, 48, \dots$ 为一簇随机变量，则称这簇随机变量的全体 $\{X(t), t = 0, 24, 48, \dots\}$ 为一随机过程，见图 3.2 (b)。此时有状态空间 $S = \{0, 1, 2, \dots\}$ 和参数集 $T = \{0, 24, 48, \dots\}$ 。

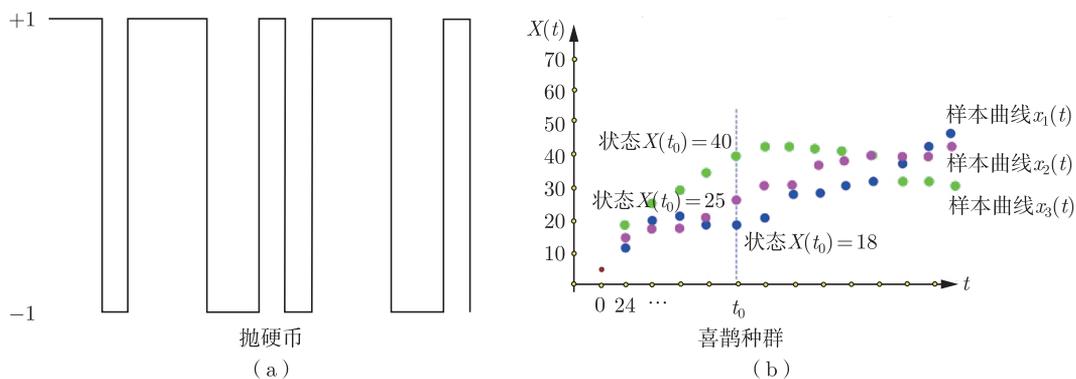


图 3.2 抛掷硬币和喜鹊种群数量

例3.5 随机过程举例：简谐波

小学生春风和楠楠做一实验：漏斗装上沙子并挂起来像钟摆一样摆动，下面放一白纸横向拉动并分析沙子在白纸上留下的痕迹。

解：沙子留下的痕迹为简谐波，钟摆幅度为振幅 A ，做一次实验就得到一个简谐波，但初相位不同，是随机的，得到的痕迹参见图 3.3。此时有参数集 $T = \{-\infty, +\infty\}$ 和状态空间 $S = \{-A, +A\}$ 。

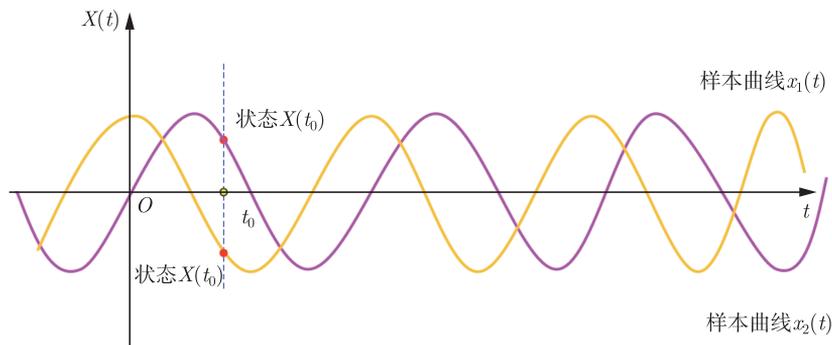


图 3.3 具有随机初相位的简谐波 ($X(t) = A \cos(\omega t + \varphi)$)

3.2 信源的数学模型和分类

如前所述, 信源是信息的来源, 但信息是抽象的, 如何刻画抽象信息的来源? 此时, 就要利用消息。信息加载于消息, 消息是信息的载体, 信源依靠发送消息来传递信息, 所以对信源中信息特性的研究就需要考察信源所发出的消息特性。

消息具有随机性, 因此可以利用随机变量或者随机过程来描述承载信源信息的消息符号, 进而描述信源特性。随机过程是随机变量簇, 是随机变量的时间函数。因此, 随机变量可以用来描述与时间无关的一类信源, 这类信源较为简单, 本章主要讨论此类信源的数学模型; 而随机过程则可以用来描述一些较为复杂的信源。

3.2.1 信源的数学模型

本节仅讨论最简单的一类信源, 可以利用随机变量来描述其输出消息的信源。根据概率论的有关知识, 可以认为随机变量 X 是样本空间 Ω 和实数集的一种映射, 因此利用随机变量的概率空间来描述此随机变量。根据定义 3.1, 概率空间有三要素 (Ω, \mathcal{F}, P) 。为了方便, 本节使用简化的概率空间定义 (见定义 2.13):

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x \\ P(x) \end{bmatrix} = \begin{bmatrix} x \\ p(x) \end{bmatrix} \quad (3.1)$$

式中, x 表示随机变量 X 的取值。

根据 x 的性质, 可以将信源分为离散信源和连续信源两类。

1. 离散信源的数学模型

若式 (3.1) 中 x 取离散值, 则随机变量 X 所表示的信源为离散信源, 意味着信源输出的消息为离散类型的消息符号, 此时消息符号的取值为有限个或者无限可数个。此类信源如书信、文稿、电报和计算机输出代码等。

离散信源的数学模型为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_q \\ p(a_1) & p(a_2) & \cdots & p(a_N) \end{bmatrix} \quad (3.2)$$

式中: $a_i (i = 1, 2, \dots, N)$ 为随机变量 X 的取值, 即信源 X 所发出的 N 个消息符号; $p(a_i)$ 为信源发出消息符号 a_i 的概率, 即 $p(a_i) = P(X = a_i)$, $p(a_i) \geq 0, i = 1, 2, \dots, N$, $\sum_{i=1}^N p(a_i) = 1$, N 为有限正整数, 或者可数无穷; 离散信源发出的消息在时间和幅值上均是离散的。

例3.6 二元信源。

已知二元信源 X 的符号集为 $\{0, 1\}$, 先验概率 $p(0) = 0.4, p(1) = 0.6$, 其数学模型为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0.4 & 0.6 \end{bmatrix}$$

二元信源又称为二进制信源。

2. 连续信源的数学模型

若式 (3.1) 中 x 取值连续, 或者取值数目是不可数的无穷值 (如某一实数区间), 则信源为连续信源, 意味着信源输出的消息为连续类型的消息符号, 此时消息符号的取值为无限不可数个。此类信源如语音、视频等。连续信源的数学模型为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} (a, b) \\ p(x) \end{bmatrix} \quad (3.3)$$

式中, $p(x)$ 为连续随机变量 X 的概率密度函数, $p(x) \geq 0$, $\int_a^b p(x)dx = 1$; (a, b) 为随机变量 X 的取值区间。

3.2.2 信源分类

信源作为信息的来源, 具有多种多样的形态, 如人、物、机器等, 尤其在万物互联时代, 世间万物都可成为信源。但不同类型的信源又表现出较大的差异, 仅以最直观的信源人为例, 智力正常的人与智力异常的人在信息表达上存在差异, 成人与婴儿和孩童等也有明显的差异, 小学生春风与中学生春风和大学生春风也存在差异, 等等。

对形态各异的信源进行分类研究是有必要的, 这样可以集中研究特定类型信源的共有特性, 而不必具体到万物中的某个“物”。不同的分类标准会得到不同的类别, 本章仅介绍几种常见的分类标准及其对应的类别。

1. 随机变量的连续性

在介绍信源的数学模型 (3.2.1节) 时, 已经对信源进行了初步分类: 信源可以建模为随机变量, 随机变量的取值可简单分为离散和连续两种类型, 因此信源也可以根据取值是否连续分为离散信源和连续信源两种类型。这是根据随机变量取值的连续性进行分类的。

2. 随机变量的关联性

信源不可能只发送一个消息符号, 它总是不停地发送一系列的消息符号, 可以根据消息符号之间的关联性对信源进行分类。若相邻消息之间统计独立, 则信源是无记忆信源; 若相邻消息不是统计独立的, 则信源为有记忆信源。

对于有记忆信源, 根据相邻符号之间关联的程度, 又进一步分为有限记忆信源和无限记忆信源两类。若消息符号之间的关联性只存在于相邻的有限个符号之间, 则信源称为有限记忆信源; 若这种关联性在所有消息符号之间都存在, 则信源称为无限记忆信源。

以离散信源为例, 前述的消息符号之间的关联程度可以利用联合概率来刻画。假设当前信源输出消息为 X_i (下标 i 表示信源所发出的消息符号的序号, 即当前信源发出了第 i 个消息符号 X_i), 则此消息符号发生的条件概率 $P[X_i|X_{i-1}X_{i-2}\cdots X_1]$ 可分为下面三种情况。

(1) $P[X_i|X_{i-1}X_{i-2}\cdots X_1] = P[X_i]$: 当前消息 X_i 与所有的历史消息 $X_{i-1}X_{i-2}\cdots X_1$ 无关, 消息之间统计独立, 此即为无记忆信源。



视频

(2) $P[X_i|X_{i-1}X_{i-2}\cdots X_1] = P[X_i|X_{i-1}X_{i-2}\cdots X_{i-M}]$: 当前消息 X_i 仅与 M 个历史消息符号 $X_{i-1}X_{i-2}\cdots X_{i-M}$ 有关, 而与更早的历史消息无关, 此即为有限记忆信源。

(3) $P[X_i|X_{i-1}X_{i-2}\cdots X_1] = P[X_i|X_{i-1}X_{i-2}\cdots X_1]$: 当前消息 X_i 与所有的历史消息 $X_{i-1}X_{i-2}\cdots X_1$ 有关, 此即为无限记忆信源。

3. 随机变量的平稳性

根据随机变量的平稳性(随机变量的统计特性是否随时间而变化), 可以将信源分为以下两种类型。

(1) **平稳信源**: 信源符号的概率分布不随时间而变化。

(2) **非平稳信源**: 信源符号的概率分布随时间而变化。比如, 语音信号可以建模为非平稳随机过程, 是非平稳信源。

4. 消息符号个数

信源之所以存在, 就是要孕育信息并将信息加载于消息之上以达到信息传输目的。因此, 就其物理本质而言, 信源会源源不断地发送消息。从数学建模角度, 信源所发送的消息符号是一无穷序列。但是, 将无穷序列作为一个随机变量进行研究较为复杂, 常将信源所发出的无穷序列简化为一系列长度为 N 的随机变量序列。根据 N 的取值, 可以将信源划分为以下两种类型。

(1) **单符号信源**: $N = 1$, 此类信源表示将信源发出的单个符号作为研究对象并建模为一个随机变量, 这是最简单的一类信源。

(2) **多符号信源**: $N > 1$, 此类信源表示将信源发出的 N 个符号作为研究对象并将其建模为一随机矢量。

5. 标准组合

前面根据随机变量的连续性、关联性等性质对信源进行了简单划分, 也可以看出分类标准其实反映的就是信源某些方面的性质, 因此也可以将不同的分类标准(或者性质)叠加融合, 从而得到更有针对性、更具体的一些信源。

(1) **单符号离散无记忆信源**: 建模为取值离散的随机变量, 相邻随机变量相互独立。

(2) **离散有记忆信源**: 取值离散的、相互之间并不独立的随机变量。

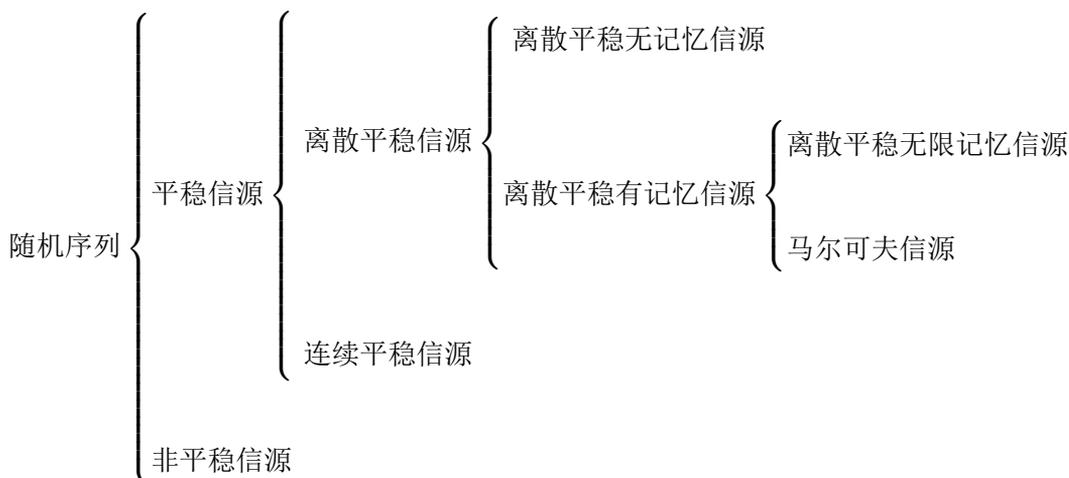
6. 简单总结

由前面的介绍可知, 信源的分类方法很多, 信源的常见分类如表 3.1 所示。

实际信源的统计特性往往是相当复杂的, 要想找到精确的数学模型很困难, 实际应用时常利用一些可以处理的数学模型来近似。例如, 语音信号是非平稳随机过程, 但常用平稳随机过程来描述。平稳随机过程抽样后就是平稳随机序列。数学上, 随机序列是随机过程的一种, 是时间参数离散的随机过程。随机序列(特别是离散平稳随机序列)是我们研究的主要内容, 单列出来, 将其分类如下:

表 3.1 信源的常见分类

时间（空间）	取值	信源种类	举例	数学描述
离散	离散	离散信源 (数字信源)	文字、数据 离散化图像	离散随机变量序列 $P[\mathbf{X}] = P[X_1 X_2 \cdots X_N]$
离散	连续	连续信源	抽样后的语音信号	连续随机变量序列 $P[\mathbf{X}] = P[X_1 X_2 \cdots X_N]$
连续	连续	波形信源 (模拟信源)	语音、热噪声 图形、图像	随机过程 $\{X(\omega, t)\}$
连续	离散		不常见	



3.3 离散单符号信源

3.3.1 离散单符号信源的概念和数学模型

若信源 X 符号集 $\mathcal{A} = \{a_1, a_2, \cdots, a_N\}$ (N 为符号集大小), 信源符号的概率分布为 $P[a_n] (n = 1, 2, \cdots, N)$, 则此信源为**离散单符号信源**。其数学模型为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_N \\ p(x_1) & p(x_2) & \cdots & p(x_N) \end{bmatrix} \quad (3.4)$$

根据定义可知, 离散单符号信源对信源中消息符号的个数并无限制, 因此 $N \rightarrow +\infty$ 也是可以的。也就是说, 离散单符号信源所发出的消息符号既可以取自有限符号集, 也可以取自无限符号集。

根据离散单符号的数学模型可以发现离散单符号信源没有考虑符号之间的关联关系, 即隐含信源所发出的符号是相互独立的, 因此离散单符号信源也是一类**离散无记忆信源**。

离散单符号信源是最简单最基本的信源, 是组成实际信源的基本单元, 可利用离散随机变量表示。

例3.7 二进制信源。

已经在例 3.6 中接触过二进制信源，其更为一般的模型是一个单符号离散无记忆信源，信源的每个符号取自于同一符号集 $\mathcal{A} = \{0, 1\}$ 。由于符号集中只有两个符号，所以称为二进制信源，又称为二元信源。二进制信源的数学模型可以表示为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ p & 1-p \end{bmatrix} \quad (3.5)$$

3.3.2 离散单符号信源的信息度量

离散单符号信源所发出的消息符号是随机的，信源发出符号 x_i 也可以认为是随机事件 x_i 发生，因此根据信息度量的性质，消息符号 x_i 含有自信息量 $I(x_i)$ ；信源 X 则会包含平均自信息量，即信息熵 $H(X)$ ：

$$I(x_i) = -\log p(x_i), \quad H(X) = \sum_{i=1}^N p(x_i) I(x_i) = -\sum_{i=1}^N p(x_i) \log p(x_i) \quad (3.6)$$

分析式 (3.6) 可以发现，这与定义 2.12 和定义 2.14 中所定义的自信息量和信息熵是完全一致的。考虑到 X 为离散单符号信源， x_i 为消息符号这些具体对象，可以认为前面所介绍的随机事件的自信息量和信息熵等是定义在离散单符号信源上的相关概念。有鉴于此，若信息熵 $H(X)$ 中 X 特指信源，则 $H(X)$ 称为信源熵。

在 X 为离散单符号信源的情况下，信源熵 $H(X)$ 继承了信息熵 $H(X)$ 所具有的一切性质，如极值性和可加性等，不再赘述。

3.4 离散多符号信源

前面介绍的单符号信源是最简单的信源模型，可以利用离散随机变量表示。实际信源输出的往往是符号序列，称为离散多符号信源，通常用离散随机变量序列（随机矢量）来表示， $\mathbf{X} = X_1 X_2 \cdots$ 。例如，电报系统发出的是一串脉冲信号（1 表示有脉冲，0 表示无脉冲），因此电报系统是离散多符号二进制信源。

为了简单起见，本书只研究离散平稳信源，即统计特性不随时间改变的信源。

定义3.3 离散平稳信源

若信源发出的消息为一离散随机序列 $\mathbf{X} = X_1 X_2 X_3 \cdots$ ，且满足：

- (1) **离散性**：随机序列中的 $X_i (i = 1, 2, \cdots)$ 均取值于有限符号集 $\mathcal{A} = \{a_1, a_2, \cdots, a_Q\}$ 。

(2) **平稳性**: 对任意的非负整数 $N, i_1, i_2, \dots, i_N, h$ 以及 $x_1, x_2, \dots, x_N \in \mathcal{A}$, 有

$$P[X_{i_1} = x_1, X_{i_2} = x_2, \dots, X_{i_N} = x_N] = P[X_{i_1+h} = x_1, X_{i_2+h} = x_2, \dots, X_{i_N+h} = x_N]$$

3.4.1 离散平稳信源的性质

根据离散平稳信源的定义可以得到离散平稳信源的一些有用性质, 具体如下:
任意时刻单符号信源的信源符号分布相同:

$$P[X_i] = P[X_j] \leftarrow N = 1 \quad (3.7)$$

任意时刻二次扩展信源的信源符号分布相同:

$$P[X_i X_{i+1}] = P[X_j X_{j+1}] \leftarrow N = 2, i_1 = i, i_2 = i_1 + 1, i_1 + h = j \quad (3.8)$$

任意时刻三次扩展信源的信源符号分布相同:

$$P[X_i X_{i+1} X_{i+2}] = P[X_j X_{j+1} X_{j+2}] \leftarrow N = 3, i_1 = i, i_2 = i_1 + 1, i_3 = i_2 + 1, i_1 + h = j \quad (3.9)$$

任意时刻 $N + 1$ 次扩展信源的信源符号分布相同:

$$P[X_i X_{i+1} \cdots X_{i+N}] = P[X_j X_{j+1} \cdots X_{j+N}]$$

$$i_1 = i, i_2 = i_1 + 1, i_3 = i_2 + 1, \dots, i_N = i_{N-1} + 1, i_1 + h = j \quad (3.10)$$

上面的性质说明, 离散平稳信源的各维联合概率分布与时间起点 (前述性质中的下标 i, j) 无关。

根据性质各维联合分布均与时间起点无关可得到下述性质:

离散平稳信源的条件概率分布均与时间起点无关, 只与关联长度 N 有关:

$$P[X_{i+1}|X_i] = P[X_{j+1}|X_j] \quad (3.11)$$

$$P[X_{i+2}|X_{i+1}X_i] = P[X_{j+2}|X_{j+1}X_j] \quad (3.12)$$

$$\vdots \quad (3.13)$$

$$P[X_{i+N}|X_{i+N-1} \cdots X_i] = P[X_{j+N}|X_{j+N-1} \cdots X_j] \quad (3.14)$$

根据这一性质可知, 平稳信源不一定是无记忆信源, 因为 $P[X_{i+1}|X_i] = P[X_{j+1}|X_j]$, 并不意味着 $P[X_{i+1}|X_i] = P[X_i]$ 。

根据前述的性质, 即联合概率分布和条件概率分布均与时间起点无关, 可得

$$H(X_1) = H(X_2) = \cdots = H(X_N) \quad (3.15)$$

$$H(X_2|X_1) = H(X_3|X_2) = \cdots = H(X_N|X_{N-1}) \quad (3.16)$$

$$H(X_3|X_2X_1) = H(X_4|X_3X_2) = \cdots = H(X_N|X_{N-1}X_{N-2}) \quad (3.17)$$

⋮

3.4.2 常见的离散平稳信源

根据前述的离散平稳信源的定义, 要求对任意的非负整数 N , 都需要满足离散平稳信源所要求的平稳性。数学上这是一个很理想的条件, 因此满足这个要求的离散信源称为**严格平稳离散信源**; 这也是一个很苛刻的条件, 实际信源很难满足这个要求, 因此在建模实际信源时, 往往对平稳性条件进行松弛, 从而定义了如下几种满足实际情况的平稳信源。

1. 一维离散平稳信源

当前述的信源符号数 $N = 1$ 时, 平稳性退化为

$$P[X_i] = P[X_j] \quad (3.18)$$

此时信源为单符号离散信源, 信源符号的一维概率 (先验概率) 与时间无关, 因此这样的信源称为**一维离散平稳信源**。

2. 二维离散平稳信源

当前述的信源符号数 $N = 2$ 时, 平稳性退化为

$$P[X_i] = P[X_j]$$

$$P[X_iX_{i+1}] = P[X_jX_{j+1}]$$

因为此时信源符号的二维联合概率分布与时间起点无关, 故称为**二维离散平稳信源**。实际中, 常使用下面的等价公式表征二维离散平稳信源:

$$P[X_i] = P[X_j] \quad (3.19)$$

$$P[X_{i+1}|X_i] = P[X_{j+1}|X_j] \quad (3.20)$$

这说明二维离散平稳信源的条件概率与时间起点无关。

3. 有限维离散平稳信源

一般地, 当前述的信源符号数 N 为有限的正整数 m 时, 平稳性退化为

$$P[X_i] = P[X_j] \quad (3.21)$$

$$P[X_{i+1}|X_i] = P[X_{j+1}|X_j] \quad (3.22)$$

$$P[X_{i+2}|X_{i+1}X_i] = P[X_{j+2}|X_{j+1}X_j] \quad (3.23)$$

$$\vdots \quad (3.24)$$

$$P[X_{i+m-1}|X_{i+m-2} \cdots X_i] = P[X_{j+m-1}|X_{j+m-2} \cdots X_j] \quad (3.25)$$

因为此时信源符号的有限维联合概率分布与时间起点无关，所以称为**有限维离散平稳信源**，或者更具体地称为 **m 维离散平稳信源**。 m 维离散平稳信源的 $m-1$ 个条件概率与时间起点无关。

3.4.3 离散多符号信源的信源熵

用信息熵（或信源熵）表示离散单符号信源的平均不确定程度。究其本质，信息熵表示随机变量集合中单个随机事件所包含的平均信息量，若所考察的随机变量集合特指信源，则信息熵（在这种情况下也可称为信源熵）是指单个信源符号所包含的平均信息量。

离散多符号信源所发出的是一个 N 长序列，因此 $H(\mathbf{X})$ 表示离散多符号信源 X 中单个符号序列 $\mathbf{X} = X_1X_2\cdots X_N$ 所包含的平均信息量。为了衡量通常意义上信源单个符号所包含的平均信息量，特引入**平均符号熵**这一概念，表示离散多符号信源所输出的符号序列中平均每个符号所携带的信息量。

定义3.4 平均符号熵

N 长随机变量序列 $\mathbf{X} = X_1X_2\cdots X_N$ 中，平均符号熵定义为

$$H_N(\mathbf{X}) = \frac{1}{N}H(X^N) = \frac{1}{N}H(X_1X_2\cdots X_N) \quad (3.26)$$

当平均符号熵中符号序列长度 N 趋于无穷时，可以得到**极限熵** H_∞ 。

定义3.5 极限熵

N 长随机变量序列 $\mathbf{X} = X_1X_2\cdots X_N$ 中，当 $N \rightarrow \infty$ 时，平均符号熵 $H_N(X)$ 定义为极限熵

$$H_\infty = \lim_{N \rightarrow \infty} H_N(X) = \lim_{N \rightarrow \infty} \frac{1}{N}H(X^N) \quad (3.27)$$

极限熵又称为**熵率**。

3.4.4 离散平稳无记忆信源

一般情况下，信源输出序列 $\mathbf{X} = X_1X_2\cdots$ 中第 n 个元素 X_n 的取值是随机的，但是前后符号的取值有一定的统计关系。简单起见，假设消息符号序列 $\mathbf{X} = X_1X_2\cdots$ 中前后符号的取值是无关的，即首先讨论**无记忆信源**这一特殊类型的信源；同时假设信源具有平稳性，故首先讨论**离散平稳无记忆信源**。

离散平稳无记忆信源输出的符号序列是平稳随机序列，并且符号序列中的各个符号相互独立。假定信源每次输出的是 N 长符号序列，这可以看作一个新的信源 Y 。这个新的信源输出的符号是 $X_1X_2\cdots X_N$ ，由于这 N 个符号相互独立，并且来自同一个符号集 $\mathcal{A} = \{a_1, a_2, \cdots, a_N\}$ ，因此离散平稳无记忆信源可以看成离散单符号无记忆信源的 N 次扩展信源。

1. 扩展信源的含义

扩展信源是从单符号信源扩展而成的新信源。下面举例说明信源是如何扩展的：

二次扩展信源是一个二维随机矢量，此二维随机矢量中的两个元素取自同一个信源符号集 \mathcal{A} ；

三次扩展信源是一个三维随机矢量，此三维随机矢量中的三个元素取自同一个信源符号集 \mathcal{A} ；

N 次扩展信源是一个 N 维随机矢量，此 N 维随机矢量中的 N 个元素取自同一个信源符号集 \mathcal{A} 。

所述的同一个信源符号集是指扩展信源中所有的符号来自一个样本空间。

定义 3.6 N 次扩展信源

设信源 X 的信源符号 x 为 N 维随机矢量 $x = (X_1 X_2 \cdots X_N)$ ，并且随机矢量中的元素取自同一个符号集，即 $X_1 = x_1 \in \mathcal{A}, X_2 = x_2 \in \mathcal{A}, \cdots, X_N = x_N \in \mathcal{A}$ ，则此类信源称为 N 次扩展信源。

其中： X_1 表示信源符号 X 中的第一个符号，为一随机变量； X_2 表示信源符号 X 中的第二个符号，为一随机变量； X_N 表示信源符号 X 中的第 N 个符号，为一随机变量； x_1 表示随机变量 X_1 的取值； x_2 表示随机变量 X_2 的取值； x_N 表示随机变量 X_N 的取值； \mathcal{A} 表示信源符号集合。

扩展的含义：信源中的所有符号来自同一个符号集 \mathcal{A} ，即信源 X 是由一个取值为 \mathcal{A} 中元素的单符号信源扩展了 N 次得到的（信源 X 中的一个信源符号由 N 个 \mathcal{A} 中的符号组成）。

2. 离散无记忆二进制信源的二次扩展信源

若信源是由离散无记忆二进制信源经过二次扩展而来，则此信源称为离散无记忆二进制信源的二次扩展信源。

根据前面的介绍，可以得到离散无记忆二进制信源的二次扩展信源的数学模型为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ p(\alpha_1) & p(\alpha_2) & p(\alpha_3) & p(\alpha_4) \end{bmatrix}$$

信源有 4 个信源符号，分别为 α_1 、 α_2 、 α_3 、 α_4 。

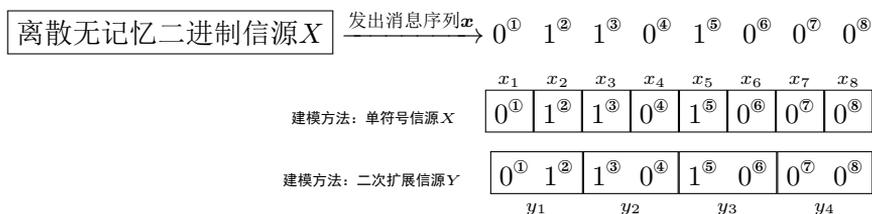
$\alpha_1 = 00, \alpha_2 = 01, \alpha_3 = 10, \alpha_4 = 11$ ，从二进制信源扩展而来。

离散无记忆性：

$$p(\alpha_1) = p(00) = p(0)p(0), \quad p(\alpha_2) = p(01) = p(0)p(1)$$

$$p(\alpha_3) = p(10) = p(1)p(0), \quad p(\alpha_4) = p(11) = p(1)p(1)$$

二次扩展信源示意:

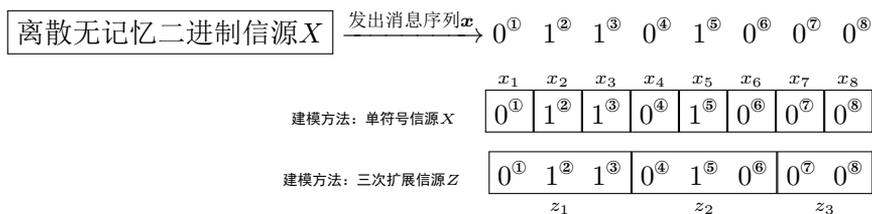


符号右上角的带圈数字表示消息符号发出的顺序, 例如 $1^{⑤}$ 表示信源 X 发出的第 5 个消息符号是 1; 符号 $x_i (i = 1, 2, \dots, 8)$ 表示单符号信源 X 发出的第 i 个消息符号; 符号 $y_i (i = 1, 2, \dots, 4)$ 表示二次扩展信源 Y 发出的第 i 个消息符号, 每个消息符号 y_i 包含两个取自基本符号集 $A = \{0, 1\}$ 的基本符号。由于此时一个信源符号含有两个基本符号, 所以消息序列 \mathbf{x} 由 4 个消息符号 y_1 、 y_2 、 y_3 、 y_4 组成。

由于二次扩展信源的符号来自同一个符号集 \mathcal{A} , 所以二次扩展信源也可以表示为 $X^2 = (X_1, X_2)$ 。

3. 离散无记忆二进制信源的三次扩展信源

离散无记忆二进制信源的三次扩展信源的生成如下:



三次扩展信源中每个信源符号有 3 个基本符号, 这 3 个基本符号取自同一符号集 $\mathcal{A} = \{0, 1\}$ 。

三次扩展信源 Z 所发出的第三个信源符号应该有三个符号, 可表示为 $z_3 = 00*$ 。其中, “*” 表示下一步符号要么是 0, 要么是 1。

通过前述的扩展信源生成过程, 可以得到离散无记忆二进制信源的三次扩展信源的数学模型为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 & \alpha_7 & \alpha_8 \\ p(\alpha_1) & p(\alpha_2) & p(\alpha_3) & p(\alpha_4) & p(\alpha_5) & p(\alpha_6) & p(\alpha_7) & p(\alpha_8) \end{bmatrix} \quad (3.28)$$

信源符号:

$$\begin{aligned} \alpha_1 &= 000, & \alpha_2 &= 001, & \alpha_3 &= 010, & \alpha_4 &= 011 \\ \alpha_5 &= 100, & \alpha_6 &= 101, & \alpha_7 &= 110, & \alpha_8 &= 111 \end{aligned}$$

信源符号发生的概率 (先验概率):

$$\begin{aligned}
 p(\alpha_1) &= p(000) = p(0)p(0)p(0), & p(\alpha_2) &= p(001) = p(0)p(0)p(1) \\
 p(\alpha_3) &= p(010) = p(0)p(1)p(0), & p(\alpha_4) &= p(011) = p(0)p(1)p(1) \\
 p(\alpha_5) &= p(100) = p(1)p(0)p(0), & p(\alpha_6) &= p(101) = p(1)p(0)p(1) \\
 p(\alpha_7) &= p(110) = p(1)p(1)p(0), & p(\alpha_8) &= p(111) = p(1)p(1)p(1)
 \end{aligned}$$

4. 离散无记忆信源的 N 次扩展信源

上面分析了离散无记忆二进制信源的扩展信源及其含义，下面将扩展信源的内涵推广到更一般的情况，即离散无记忆信源的 N 次扩展信源。

定义 3.7 离散无记忆信源的 N 次扩展信源

X 为一离散无记忆信源，其数学模型为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_Q \\ p_1 & p_2 & \cdots & p_Q \end{bmatrix}$$

X 的 N 次扩展信源 \mathbf{X} (或记为 X^N) 是具有 Q^N 个消息符号的离散多符号信源，其数学模型为

$$\begin{bmatrix} X^N \\ P \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{Q^N} \\ p(\alpha_1) & p(\alpha_2) & \cdots & p(\alpha_{Q^N}) \end{bmatrix} \quad (3.29)$$

其中， Q 表示信源 X 中的符号个数； $p_i = P[X = a_i] (i = 1, 2, \dots, Q)$ 。

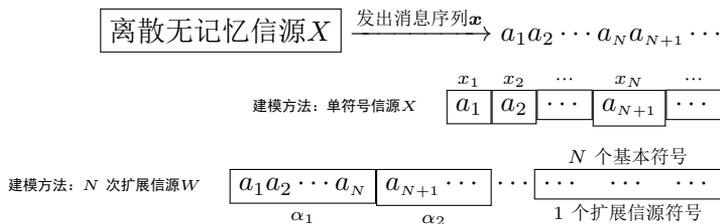
扩展信源： $X^N = (X_1, X_2, \dots, X_N)$ 。

扩展信源符号 α_i ： $\alpha_i = a_{i_1}, a_{i_2}, \dots, a_{i_N} (i = 1, 2, \dots, Q^N)$ 。

扩展信源的体现： $a_{i_k} \in \mathcal{A} = \{a_1, a_2, \dots, a_Q\} (k = 1, 2, \dots, N)$ 。

无记忆信源： $p(\alpha_i) = P[X^N = \alpha_i] = \prod_{k=1}^N p_{i_k}$ 。

N 次扩展信源生成过程：



5. 扩展信源的信源熵

前面还定义信源熵 $H(X) = \sum_{n=1}^N p(x_n) \log p(x_n)$ 。结合本章介绍的扩展信源的含义可以发现，此处的信源 X 实际为最简单的单符号离散无记忆信源。可以对此处定义的信源熵进

行推广，将信源类型从单符号离散无记忆信源推广到扩展信源。根据信源熵的含义，扩展信源的信源熵可以定义如下：

定义3.8 扩展信源的信源熵

设一离散信源 X ，其符号集为 $\mathcal{A} = \{a_1, a_2, \dots, a_Q\}$ 。信源 X 的 N 次扩展信源 X^N 的信源熵定义为

$$H(X^N) = H(X_1 X_2 \cdots X_N) = \sum_{X^N} p(\alpha_i) \log p(\alpha_i) = - \sum_{i=1}^{Q^N} p(\alpha_i) \log p(\alpha_i) \quad (3.30)$$

其中， α_i 为扩展信源 X^N 的第 i 个信源符号 ($i = 1, 2, \dots, Q^N$)。

根据扩展信源信源熵的定义来计算最简单的扩展信源（离散无记忆信源的 N 次扩展信源）所具有的信源熵，此类信源的信源熵满足下面的定理：

定理3.1 离散无记忆信源的 N 次扩展信源的信源熵

离散无记忆信源 X 的 N 次扩展信源 X^N 的信源熵等于信源 X 的信源熵的 N 倍，即

$$H(X^N) = NH(X) \quad (3.31)$$

证明： 假设一离散无记忆信源的数学模型为

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_Q \\ p_1 & p_2 & \cdots & p_Q \end{bmatrix}$$

根据扩展信源的信源熵定义，扩展信源 X^N 的信源熵为

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

证明方法一： 由于 X^N 为离散无记忆信源 X 的扩展信源，所以信源 X_1, X_2, \dots, X_N 是相互独立的 N 个信源，且 $H(X_1) = H(X_2) = \cdots = H(X_N) = H(X)$ 。根据熵的可加性（定理 2.1），有

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

证明方法二：

$$H(X^N) = - \sum_{X^N} p(\alpha_i) \log p(\alpha_i) = - \sum_{i=1}^{Q^N} p(\alpha_i) \log(\alpha_i)$$

$$\begin{aligned}
& \text{离散无记忆信源: } p(\alpha_i) = \prod_{k=1}^N p(a_{i_k}) \\
& \dots\dots\dots \\
& = - \sum_{i_1, i_2, \dots, i_N=1}^Q p(a_{i_1})p(a_{i_2}) \cdots p(a_{i_N}) \left[\log p(a_{i_1}) + \log p(a_{i_2}) + \cdots + \log p(a_{i_N}) \right] \\
& = - \sum_{i_1, i_2, \dots, i_N=1}^Q p(a_{i_1})p(a_{i_2}) \cdots p(a_{i_N}) \log p(a_{i_1}) \\
& \quad \dots \\
& \quad - \sum_{i_1, i_2, \dots, i_N=1}^Q p(a_{i_1})p(a_{i_2}) \cdots p(a_{i_N}) \log p(a_{i_N}) \\
& = - \sum_{i_1=1}^Q p(a_{i_1}) \log p(a_{i_1}) \underbrace{\sum_{i_2=1}^Q \cdots \sum_{i_N=1}^Q p(a_{i_2}) \cdots p(a_{i_N})}_1 \\
& \quad \dots \\
& \quad - \sum_{i_N=1}^Q p(a_{i_N}) \log p(a_{i_N}) \underbrace{\sum_{i_1=1}^Q \cdots \sum_{i_{N-1}=1}^Q p(a_{i_1}) \cdots p(a_{i_{N-1}})}_1 \\
& = - \underbrace{\sum_{i_1=1}^Q p(a_{i_1}) \log p(a_{i_1})}_{H(X_1)=H(X)} - \cdots - \underbrace{\sum_{i_N=1}^Q p(a_{i_N}) \log p(a_{i_N})}_{H(X_N)=H(X)} = NH(X)
\end{aligned}$$

推论: 根据定义 3.4 可以求得离散无记忆信源的 N 次扩展信源的平均符号熵为

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

根据定义 3.5 可以求得离散无记忆信源的 N 次扩展信源的熵率为

$$H_\infty = \lim_{N \rightarrow \infty} H_N(X^N) = \lim_{N \rightarrow \infty} NH(X) = H(X)$$

例 3.8 红绿灯。

已知小学生春风和楠楠庆祝完了 12 岁生日, 根据交通法规, 可以骑自行车了。两人高兴地骑自行车去兜风, 并记录了骑行路上所遇到的交通信号灯情况, 得到信号灯序列 X :

$$X = \text{红}^{\textcircled{1}} \text{红}^{\textcircled{2}} \text{绿}^{\textcircled{3}} \text{红}^{\textcircled{4}} \text{红}^{\textcircled{5}} \text{绿}^{\textcircled{6}} \text{绿}^{\textcircled{7}} \text{红}^{\textcircled{8}}$$

其中, 右上标的数字表示序号, 例如 $\text{红}^{\textcircled{8}}$ 表示第 8 个路口遇到的是红灯。

同时做如下假设:

- (1) 相邻信号灯之间没有任何关联;
- (2) 遇到红灯和绿灯的概率相等。

将信号灯序列 X 建模为信源, 求下列建模方法所对应的信号灯序列 X 的自信息量:

- (1) 单符号信源;
- (2) 两符号信源;
- (3) 四符号信源;
- (4) 八符号信源。

解: 根据已知条件, 可设基本符号 $a_1 = \text{红}$ 和 $a_2 = \text{绿}$, 且有

$$P[\text{红}] = 0.5, \quad P[\text{绿}] = 0.5$$

在 $N = 1$ 的情况下, 每个基本符号 (红灯或者绿灯) 都是一个随机事件, 所求序列 \mathbf{X} 则为 8 个随机事件组成的联合事件, 其发生概率为

$$\begin{aligned} P[\mathbf{X}] &= P[\text{红}^{\textcircled{1}}]P[\text{红}^{\textcircled{2}}]P[\text{绿}^{\textcircled{3}}]P[\text{红}^{\textcircled{4}}]P[\text{红}^{\textcircled{5}}]P[\text{绿}^{\textcircled{6}}]P[\text{绿}^{\textcircled{7}}]P[\text{红}^{\textcircled{8}}] \\ &= 0.5^8 \end{aligned}$$

则 \mathbf{X} 的自信息量 $I_1(\mathbf{X}) = -\log_2 P[\mathbf{X}] = 8\text{bit}$

由于此序列含有 8 个基本符号, 平均每个基本符号所含有的自信息量 $I_1^{\text{平均}} = \frac{8\text{bit}}{8} = 1\text{bit}$

在 $N = 2$ 的情况下, 一个信源符号中含有 2 个基本符号 (红灯或者绿灯), 此时信源符号为一个二维随机矢量 $\mathbf{x} = (x_1 x_2)$ 。则序列 \mathbf{X} 由 4 个信源符号组成:

$$\mathbf{X} = \underbrace{\text{红}^{\textcircled{1}} \text{红}^{\textcircled{2}}}_{\mathbf{x}^1} \quad \underbrace{\text{绿}^{\textcircled{3}} \text{红}^{\textcircled{4}}}_{\mathbf{x}^2} \quad \underbrace{\text{红}^{\textcircled{5}} \text{绿}^{\textcircled{6}}}_{\mathbf{x}^3} \quad \underbrace{\text{绿}^{\textcircled{7}} \text{红}^{\textcircled{8}}}_{\mathbf{x}^4}$$

每个信源符号的发生概率为

$$P[\mathbf{x}^1] = P[\text{红}] \times P[\text{红}] = 0.5^2, \quad P[\mathbf{x}^2] = P[\text{绿}] \times P[\text{红}] = 0.5^2$$

$$P[\mathbf{x}^3] = P[\text{红}] \times P[\text{绿}] = 0.5^2, \quad P[\mathbf{x}^4] = P[\text{绿}] \times P[\text{红}] = 0.5^2$$

符号序列 \mathbf{X} 的发生概率为

$$P[\mathbf{X}] = P[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4] = P[\mathbf{x}_1] \times P[\mathbf{x}_2] \times P[\mathbf{x}_3] \times P[\mathbf{x}_4] = 0.5^8$$

序列 \mathbf{X} 所包含的自信息量 $I_2(\mathbf{X}) = -\log_2 P[\mathbf{X}] = 8\text{bit}$, 平均每个信源符号所携带的自信息量 $I_2(\mathbf{x}) = \frac{8\text{bit}}{4} = 2\text{bit}$, 平均每个基本符号所携带的自信息量 $I_2^{\text{平均}} = \frac{2\text{bit}}{2} = 1\text{bit}$ 。

在 $N = 4$ 的情况下, 一个信源符号中含有 4 个基本符号 (红灯或者绿灯), 此时信源符号为一个四维随机矢量 $\mathbf{x} = (x_1 x_2 x_3 x_4)$, 则序列 \mathbf{X} 由 2 个信源符号组成:

$$\mathbf{X} = \underbrace{\text{红}^{\textcircled{1}} \text{红}^{\textcircled{2}} \text{绿}^{\textcircled{3}} \text{红}^{\textcircled{4}}}_{\mathbf{x}^1} \quad \underbrace{\text{红}^{\textcircled{5}} \text{绿}^{\textcircled{6}} \text{绿}^{\textcircled{7}} \text{红}^{\textcircled{8}}}_{\mathbf{x}^2}$$

每个信源符号的发生概率为

$$P[\mathbf{x}^1] = P[\text{红}] \times P[\text{红}] \times P[\text{绿}] \times P[\text{红}] = 0.5^4$$

$$P[\mathbf{x}^2] = P[\text{红}] \times P[\text{绿}] \times P[\text{绿}] \times P[\text{红}] = 0.5^4$$

符号序列 \mathbf{X} 的发生概率为

$$P[\mathbf{X}] = P[\mathbf{x}^1, \mathbf{x}^2] = P[\mathbf{x}^1] \times P[\mathbf{x}^2] = 0.5^8$$

符号序列 \mathbf{X} 所携带的自信息量 $I_4(\mathbf{X}) = -\log_2 P[\mathbf{X}] = 8\text{bit}$

平均每个信源符号所携带的自信息量 $I_4^2 = \frac{8\text{bit}}{2} = 4\text{bit}$

平均每个基本符号所携带的自信息量 $I_4^4 = \frac{4\text{bit}}{4} = 1\text{bit}$

在 $N = 8$ 的情况下, 一个信源符号中含有 8 个基本符号 (红灯或者绿灯), 此时信源符号为一个八维随机矢量 $\mathbf{x} = (x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8)$ 。则序列 \mathbf{X} 由一个信源符号组成:

$$\mathbf{X} = \underbrace{\text{红}^{\textcircled{1}} \text{红}^{\textcircled{2}} \text{绿}^{\textcircled{3}} \text{红}^{\textcircled{4}} \text{红}^{\textcircled{5}} \text{绿}^{\textcircled{6}} \text{绿}^{\textcircled{7}} \text{红}^{\textcircled{8}}}_{\mathbf{x}^1}$$

每个信源符号的发生概率为

$$P[\mathbf{x}^1] = P[\text{红}]P[\text{红}]P[\text{绿}]P[\text{红}]P[\text{红}]P[\text{绿}]P[\text{绿}]P[\text{红}] = 0.5^8$$

符号序列 \mathbf{X} 的发生概率为

$$P[\mathbf{X}] = P[\mathbf{x}^1] = 0.5^8$$

符号序列 \mathbf{X} 所携带的自信息量 $I_8(\mathbf{X}) = -\log_2 P[\mathbf{X}] = 8\text{bit}$

平均每个信源符号所携带的自信息量 $I_8^1 = \frac{8\text{bit}}{1} = 8\text{bit}$

平均单个基本符号所携带的自信息量 $I_8^8 = \frac{8\text{bit}}{8} = 1\text{bit}$

3.4.5 离散平稳有记忆信源

前面介绍了离散平稳信源中最简单的离散平稳无记忆信源, 而实际信源往往是有记忆信源。假定信源输出 N 长符号序列, 则它的数学模型是 N 维随机变量序列 (随机矢量) $\mathbf{X} = X_1 X_2 \cdots X_N$, 其中每个随机变量 $X_n (n = 1, 2, \cdots, N)$ 之间存在统计依赖关系。

1. 离散平稳有记忆信源的信源熵

平稳有记忆信源输出符号之间的相互依存关系不仅存在于相邻两个符号之间, 而且存在于多个符号之间。若离散平稳有记忆信源的输出为一个 N 长序列 $\mathbf{X} = X_1 X_2 \cdots X_N$, 则此信源 \mathbf{X} 的平均不确定程度可以利用联合熵表示:

$$H(\mathbf{X}) = H(X_1 X_2 \cdots X_N)$$

$$\begin{aligned}
&= H(X_1) + H(X_2 X_3 \cdots X_N | X_1) = H(X_1) + H(X_2 | X_1) + H(X_3 \cdots X_N | X_1 X_2) \\
&= H(X_1) + H(X_2 | X_1) + H(X_3 | X_2 X_1) + \cdots + H(X_N | X_{N-1} X_{N-2} \cdots X_2 X_1)
\end{aligned} \tag{3.32}$$

式 (3.32) 称为熵函数的链规则, 即 N 维随机变量的联合熵等于随机变量 X_1 的熵与各阶条件熵之和。

2. 离散平稳有记忆信源的信源熵性质

对于离散平稳有记忆信源 $\mathbf{X} = X_1 X_2 \cdots X_N$ 的信源熵有下面的定理。

定理 3.2 信源熵性质

- 条件熵 $H(X_N | X_{N-1} X_{N-2} \cdots X_2 X_1)$ 随 N 的增加是递减的;
- N 给定时平均符号熵大于或等于条件熵, 即

$$H_N(\mathbf{X}) \geq H(X_N | X_{N-1} X_{N-2} \cdots X_2 X_1) \tag{3.33}$$

- 平均符号熵 $H_N(\mathbf{X})$ 随 N 的增加是递减的;
- 若 $H(X_1) < \infty$, 则极限熵 $H_\infty = \lim_{N \rightarrow \infty} H_N(\mathbf{X})$ 存在, 并且

$$H_\infty = \lim_{N \rightarrow \infty} H(X_N | X_{N-1} X_{N-2} \cdots X_2 X_1)$$

证明: (1) 条件熵随 N 的增加而递减:

$$\begin{aligned}
H(X_N | X_{N-1} X_{N-2} \cdots X_2 X_1) &= H(\overbrace{X_N | X_{N-1} X_{N-2} \cdots X_2}^{\text{定义为新的随机变量 } y} | X_1) \\
&\leq H(y) \quad \text{条件熵小于或等于无条件熵: 式 (2.51)} \\
&= H(X_N | X_{N-1} X_{N-2} \cdots X_2) \\
&= H(X_{N-1} | X_{N-2} X_{N-3} \cdots X_1) \quad \text{平稳信源性质: 式 (3.17)}
\end{aligned}$$

所以条件熵 $H(X_N | X_{N-1} X_{N-2} \cdots X_2 X_1)$ 随 N 的增加而递减。

条件熵的递减性说明记忆长度越长, 条件熵越小, 即信源符号之间的依赖关系增加时, 信源的不确定性降低。

(2) 平均符号熵大于或等于条件熵:

$$\begin{aligned}
&N H_N(\mathbf{X}) \\
&= H(X_1 X_2 \cdots X_N) \\
&= H(X_1) + H(X_2 | X_1) + \cdots + H(X_N | X_{N-1} X_{N-2} \cdots X_2 X_1) \\
&= H(X_N) + H(X_N | X_{N-1}) + \cdots + H(X_N | X_{N-1} X_{N-2} \cdots X_2 X_1) \quad \text{平稳信源性质: 式 (3.17)}
\end{aligned}$$