第3章 信源压缩编码理论（第一部分）

樊平毅　教授

清华大学电子工程系 WIST LAB.

Email: fpy@tsinghua.edu.cn

2022年9月14日

渐近均分性 ·······

渐近均分性定理 ·······

AEP 的推论:数据压缩 ·

高概率集与典型集 ····

- **例题：** 考虑一个二项分布，

$\mathrm{P}(1)=p, P(0)=q, p+q=1, p>0, q>0.$

如果序列 $X_1X_2\cdots X_{n-1}X_n$ 是独立同分布，服从上述二项分布。

显然，在$p\neq q\neq 1/2$, 任何长度为n的0，1序列不可能等概率出现，

这意味着有的序列出现的概率大，有的序列出现的概率小。

例如，全1序列和全0序列，它们出现的概率， p=0.6, q=0.4,

总有一个是最小的，一个是最大的。 （全0最小，全1最大）

目标：寻找哪些出现概率基本相同的序列（有限长）？

（1）这样的序列的特点是什么？ 基本特征

（2） 这样的序列数目有多少？

（3）对于这样的序列而言，符号间的相关性应很弱，看起来是独立同分布的

（4） 序列的可扩张性很好，保持了其分布的稳定性

考虑大数定律

$\frac{1}{n}logP(X_1X_2\cdots X_n)=\frac{1}{n}\sum_{k=1}^{n}\log P(X_k)\rightarrow -H(X)$,

这意味着有一部分长度为n的序列，其概率值主要集中在均值附近

在($2^{-nH(X)-\varepsilon}, 2^{-nH(X)+\varepsilon}$) 内，$\varepsilon>0$ 的概率很大，

研究发现：

$$Pr\left((X_1X_2\cdots X_n): P(X_1X_2\cdots X_n)\in(2^{-nH(X)-\varepsilon}, 2^{-nH(X)+\varepsilon})\right)\approx 1$$

Shannon 最先看到这一现象，将它应用到1948年的论文中

只保留概率基本相同的序列（长度为n)。

核心思想

“大道至简”
---信息论的基本理念

只需要研究出现大概率的事件集就可以了。删除哪些出现概率小的序列

用于解决信息论两个核心问题：

信源编码问题和信道编码问题。

原因： 我们的系统允许出错，只要是错误量较小，可控或者可忍受就好！

复习：随机收敛性的几个基本定义

定义(随机变量的收敛) 给定一个随机变量序列 $X_1, X_2, \cdots$。序列 $X_1, X_2, \cdots$ 收敛于随机变量 X 有如下三种情形：

1. 如果对任意的 $\varepsilon > 0$, $\Pr\{|X_n - X| > \varepsilon\} \to 0$, 则称为依概率收敛。
2. 如果 $E(X_n - X)^2 \to 0$, 则称为均方收敛。
3. 如果 $\Pr\{\lim_{n\to\infty} X_n = X\} = 1$, 则称为以概率 1(或称几乎处处)收敛。

本课程基本选择依概率收敛的方法加以讨论

定理 3.1.1(AEP) 若 $X_1, X_2, \cdots, X_n$ 为 i.i.d$\sim p(x)$，则

$$-\frac{1}{n}\log p(X_1, X_2, \cdots, X_n) \to H(X) \quad \text{依概率}$$

证明： 独立随机变量的函数依然是独立随机变量。因此，由于 X_i 是i.i.d.，从而 $\log p(X_i)$ 也是 i.i.d.。因而，由弱大数定律，

$$-\frac{1}{n}\log p(X_1, X_2, \cdots, X_n) = -\frac{1}{n}\sum_i \log p(X_i) \tag{3-3}$$

$$\to -E\log p(X) \quad \text{依概率} \tag{3-4}$$

$$= H(X) \tag{3-5}$$

典型集的定义：

定义　关于 $p(x)$ 的典型集 $A_\epsilon^{(n)}$ (typical set) 是序列 $(x_1,x_2,\cdots,x_n)\in\mathcal{X}^n$ 的集合，质：

$$2^{-n(H(X)+\epsilon)}\leqslant p(x_1,x_2,\cdots,x_n)\leqslant 2^{-n(H(X)-\epsilon)}$$

性质（取值范围，概率权重占比，集合大小）

定理 3.1.2

1. 如果 $(x_1,x_2,\cdots,x_n)\in A_\epsilon^{(n)}$，则 $H(X)-\epsilon\leqslant-\frac{1}{n}\log p(x_1,x_2,\cdots,x_n)\leqslant H(X)+\epsilon$
2. 当 n 充分大时，$\Pr\{A_\epsilon^{(n)}\}>1-\epsilon$。
3. $|A_\epsilon^{(n)}|\leqslant 2^{n(H(X)+\epsilon)}$，其中 $|A|$ 表示集合 A 中的元素个数。
4. 当 n 充分大时，$|A_\epsilon^{(n)}|\geqslant(1-\epsilon)2^{n(H(X)-\epsilon)}$。

典型集的例子

考虑 n=5, $P(X=1)=\frac{3}{5}, P(X=2)=\frac{1}{5}, P(X=3)=\frac{1}{5}$ 的离散信源，

- 典型序列

{11123， 21311， 32111，12311，。。。}共计$\binom{5}{3\quad 1\quad 1}$=20个

- 非典型集序列

{11223，33311，22331，。。}

共计 $3^5-20=243-20=223$个

显然，非典型集占据大量的空间数目。

大道至简，可信、易于刻画，可扩展，基本特征不变。

典型集的定义与性质

典型集的定义：

定义 关于 $p(x)$ 的典型集 $A_\varepsilon^{(n)}$ (typical set) 是序列 $(x_1,x_2,\cdots,x_n)\in\mathcal{X}^n$ 的集合，满足如下性质：

$$2^{-n(H(X)+\varepsilon)}\leqslant p(x_1,x_2,\cdots,x_n)\leqslant 2^{-n(H(X)-\varepsilon)}$$

性质（取值范围，概率权重占比，集合大小）

定理 3.1.2

这些序列几乎是等概的

1. 如果 $(x_1,x_2,\cdots,x_n)\in A_\varepsilon^{(n)}$，则 $H(X)-\varepsilon\leqslant -\frac{1}{n}\log p(x_1,x_2,\cdots,x_n)\leqslant H(X)+\varepsilon$
2. 当 n 充分大时，$\Pr\{A_\varepsilon^{(n)}\}>1-\varepsilon$。

序列集概率和的值接近1

扩张稳定性：

单符号的熵是一致的

3. $|A_\varepsilon^{(n)}|\leqslant 2^{n(H(X)+\varepsilon)}$，其中 $|A|$ 表示集合 A 中的元素个数。

序列的数目

4. 当 n 充分大时，$|A_\varepsilon^{(n)}|\geqslant(1-\varepsilon)2^{n(H(X)-\varepsilon)}$。

证明：

证明：性质(1)的证明可直接由 $A_\epsilon^{(n)}$ 的定义得到。第二个性质由定理 3.1.1 直接得到，这是由于当 $n\to\infty$时，事件$(X_1,X_2,\cdots,X_n)\in A_\epsilon^{(n)}$ 的概率趋于1。于是，对任意 $\delta>0$，存在 n_0，使得当 $n\geqslant n_0$ 时，有

$$\Pr\left\{\left|-\frac{1}{n}\log p(X_1,X_2,\cdots,X_n)-H(X)\right|<\epsilon\right\}>1-\delta \tag{3-7}$$

令 $\delta=\epsilon$，即可得到定理的第二个性质。取 $\delta=\epsilon$ 便于以后简化符号。

为证明性质(3)，我们有

$$\begin{aligned} 1 &= \sum_{\mathbf{x}\in\mathcal{X}^n} p(\mathbf{x}) \\ &\geqslant \sum_{\mathbf{x}\in A_\epsilon^{(n)}} p(\mathbf{x}) \\ &\geqslant \sum_{\mathbf{x}\in A_\epsilon^{(n)}} 2^{-n(H(X)+\epsilon)} \\ &= 2^{-n(H(X)+\epsilon)}|A_\epsilon^{(n)}| \end{aligned}$$

其中第二个不等式由式(3-6)得到。因此

$$|A_\epsilon^{(n)}|\leqslant 2^{n(H(X)+\epsilon)}$$

证明部分

最后，当 n 充分大时，$\Pr\{A_\varepsilon^{(n)}\}>1-\varepsilon$，所以

$$
\begin{aligned}
1-\varepsilon &< \Pr\{A_\varepsilon^{(n)}\} \\
&\leqslant \sum_{x\in A_\varepsilon^{(n)}} 2^{-n(H(X)-\varepsilon)} \\
&= 2^{-n(H(X)-\varepsilon)}|A_\varepsilon^{(n)}|
\end{aligned}
$$

其中第二个不等式由式(3-6)得到。因此，

$$|A_\varepsilon^{(n)}|\geqslant(1-\varepsilon)2^{n(H(X)-\varepsilon)}$$

渐近等分性的应用

解决离散信源压缩编码问题

技术特点：

信源压缩编码---无失真压缩技术

两部分压缩技术---无损压缩

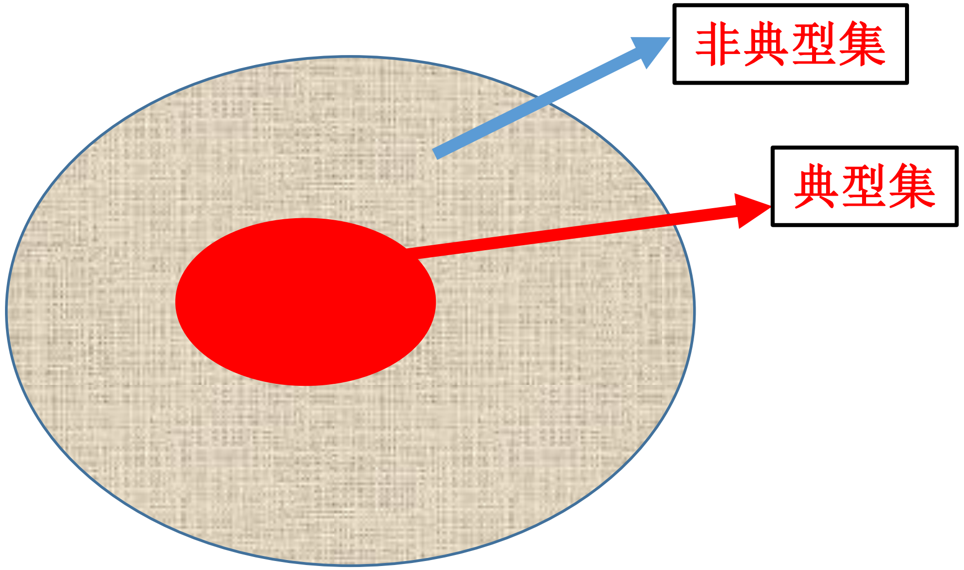
信源压缩：
整个集合分解为：
典型集和非典型集

典型集：

集内每个元素
按等长编码，
长度由信源熵决定
内涵：信源分布信息

非典型集：
采用等长编码，
最大长度 由字母数目决定
（与分布无关）

内涵：等概分布信源
无任何压缩可能；



总序列数: $|X|^n$
典型集中序列数:$2^{n(H+\epsilon)}$

平均编码长度分析

Tsinghua University

下面用记号 x^n 表示序列 $x_1, x_2, \cdots, x_n$。设 $l(x^n)$ 表示相应于 x^n 的码字长度。若 n 充分大，使得 $\Pr\{A_\varepsilon^{(n)}\} \geqslant 1-\varepsilon$，于是，码字长度的数学期望为

$$
\begin{aligned}
E(l(X^n)) &= \sum_{x^n} p(x^n) l(x^n) \\
&= \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n) l(x^n) + \sum_{x^n \in A_\varepsilon^{(n)^c}} p(x^n) l(x^n) \\
&\leqslant \sum_{x^n \in A_\varepsilon^{(n)}} p(x^n)(n(H+\varepsilon)+2) \\
&\quad + \sum_{x^n \in A_\varepsilon^{(n)^c}} p(x^n)(n \log |\mathcal{X}| + 2) \\
&= \Pr\{A_\varepsilon^{(n)}\}(n(H+\varepsilon)+2) + \Pr\{A_\varepsilon^{(n)^c}\}(n \log |\mathcal{X}| + 2) \\
&\leqslant n(H+\varepsilon) + \varepsilon n (\log |\mathcal{X}|) + 2 \\
&= n(H+\varepsilon')
\end{aligned}
$$

注意，上述编码方案有如下特征：

- 编码是 1－1 的，且易于译码。起始位作为标识位，标明紧随码字的长度。
- 对非典型集 $A_\varepsilon^{(n)^c}$ 的元素作了枚举，没有考虑 $A_\varepsilon^{(n)^c}$ 中的元素个数实际上少于 $\mathcal{X}^n$ 中元素个数。而让人惊讶的是，这足以产生一个有效的描述。
- 典型序列具有较短的描述长度 $\approx nH$。

典型集选择的一致性理论

定理 3.2.1 设 X^n 为服从 $p(x)$ 的i.i.d序列，$\varepsilon>0$，则存在一个编码将长度为 n 的序列 x^n 映射为比特串，使得映射是1-1的(因而可逆)，且对于充分大的 n，有

$$E\left[\frac{1}{n}l(X^n)\right]\leqslant H(X)+\varepsilon \tag{3-23}$$

定义： a_n 与 b_n 在一阶指数意义下是等价的,记$a_n \doteq b_n$

如果 $\lim\limits_{n\to\infty}\frac{1}{n}\log\frac{a_n}{b_n}=0$

典型集选择的一致性理论

定理 3.3.1 设 $X_1, X_2, \cdots, X_n$ 为服从 $p(x)$ 的 i.i.d 序列。对 $\delta < \frac{1}{2}$ 及任意的 $\delta' > 0$，如果 $\Pr\{B_\delta^{(n)}\} > 1 - \delta$，则

$$\frac{1}{n}\log|B_\delta^{(n)}| > H - \delta' \quad \text{对于充分大的 } n \tag{3-25}$$

由此可将上述结果重述为：如果 $\delta_n \to 0$ 和 $\varepsilon_n \to 0$，则

$$|B_{\delta_n}^{(n)}| \doteq |A_{\varepsilon_n}^{(n)}| \doteq 2^{nH}$$

一致性

核心思想： 只要采用典型集的处理，随机选取一种编码模式，得到的结果基本一致， 为编码器的设计带来很大的空间

典型集的选择策略举例

例题：

为说明 $A_\epsilon^{(n)}$ 与 $B_\delta^{(n)}$ 之间的区别，考虑一个伯努利序列 $X_1, X_2, \cdots, X_n$，其参数 $p = 0.9$ (Bernoulli(θ))随机变量是一个二值随机变量，其取1值的概率为 θ)。此时，典型序列中1所占的比例近似等于0.9。然而，这并不包括很可能出现的全部是1的序列。集合 $B_\delta^{(n)}$ 包括所有很可能出现的序列，因而包括全部为1的序列。定理3.3.1表明 $A_\epsilon^{(n)}$ 与 $B_\delta^{(n)}$ 必定包含了所有1所占比例大约为90%的序列，且两者的元素数量几乎相等。

部分总结

- 典型集的定义
- 典型集的基本特征（三条）
- 典型集与无失真信源表示（压缩）
- 典型集选择的一致性规则

探索题2 图示

探索题目：

1. 典型集与统计不等式的关系
2. 利用多少bit 可以完整描述右边的图型。