第四章无失真信源编码

第二节渐进等分性

陈兴同

中国矿业大学 数学学院

2021 年 8 月

WPS PDF编辑试用

内容提要

1. 本节预引理

WPS PDF编辑试用

内容提要

1. 本节预引理
2. 弱大数定律

WPS PDF编辑试用

内容提要

WPS PDF编辑试用

内容提要

WPS PDF编辑试用

内容提要

WPS PDF编辑试用

本节引言：

在对信源发出的消息序列进行分组编码时，通常要划分成很多 n 长的消息，然后再进行编码。但每个分组出现的概率可能互不相同，有些会非常小。利用大数定律可以将分组后的 n 长消息进一步分成典型与非典型消息，在编码时只对典型消息进行编码，这种方式进行编码造成的译码误差会充分小。

依概率收敛：

定义4.2.1

设 $\{\xi_n\}_{n=1}^{\infty}$ 是定义在同一个字符集上的随机变量序列，ξ 是一个常数或随机变量。如果对 $\forall \varepsilon > 0$ 总有

$$\lim_{n\to\infty} P\{|\xi_n - \xi| \geq \varepsilon\} = 0,$$

则称序列 $\{\xi_n\}_{n=1}^{\infty}$ 依概率收敛于 ξ，记作 $\xi_n \xrightarrow{P} \xi$。

这是概率论中的一种弱收敛，与它相关的大数定律称为弱大数定律。

WPS PDF编辑试用

弱大数定律：

设 $\{\xi_n\}_{n=1}^{\infty}$ 是随机变量序列，若存在一个数列 $\{a_n\}_{n=1}^{\infty}$ 使得

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - a_n\right| \geq \varepsilon\right\} = 0 \text{亦即} \frac{1}{n}\sum_{i=1}^{n}\xi_i - a_n \xrightarrow{P} 0,$$

则称该随机变量序列满足**弱大数定律**。大数定律有多个，本章只用到辛钦大数定律。

定理4.2.1

（**辛钦大数定律**）*设随机变量序列设* $\{\xi_n\}_{n=1}^{\infty}$ *独立同分布，并且数学期望为* μ*，则*

$$\frac{1}{n}\sum_{i=1}^{n}\xi_i \xrightarrow{P} \mu.$$

更多大数定律参考文献 [1]。

WPS PDF编辑试用

续证明:

消息序列

设离散无记忆信源的字符空间 $\mathcal{X}$ 中有如下分布:

$$X \sim p(x) = \begin{pmatrix} a_1 & a_2 & \cdots & a_N \\ p_1 & p_2 & \cdots & p_N \end{pmatrix}, x \in \mathcal{X}. \tag{3.1}$$

4.2

引入记号（1）n 长随机向量 $X^{(n)} = (X_1, X_2, \cdots, X_n)$ 表示信源发出的 n 长消息，取自全体 n 长消息集合

$$\mathcal{X}^n = \left\{ x^{(n)} = (x_1, x_2, \cdots, x_n) | x_i \in \mathcal{X}, i = 1, 2, \cdots, n \right\},$$

由于是离散无记忆信源，故 $X^{(n)}$ 的分布律为

$$p(x^{(n)}) = p(x_1, x_2, \cdots, x_n) = P\left\{ X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n \right\} = \prod_{i=}^{n}$$

WPS PDF编辑试用

续：

（2）随机变量 X 的函数 $p(X)$ 也是随机变量，它的概率分布是

$$p(X) \sim \begin{pmatrix} p_1 & p_2 & \cdots & p_N \\ p_1 & p_2 & \cdots & p_N \end{pmatrix}.$$

（3）随机向量 $X^{(n)}$ 的函数 $p(X^{(n)})$ 也是随机向量，表达式是

$$p(X^{(n)}) = p(X_1)p(X_2)\cdots p(X_n). \qquad (3.2)$$

4.3

（4）因随机变量 $p(X^{(n)}) > 0$，故

$$Z_n = \log p(X^{(n)}), n = 1, 2, \cdots$$

也是一个随机变量，它的取值为

$$z_n = \log p(x^{(n)}) = \sum_{i=1}^{n} \log p(x_i), n = 1, 2, \cdots ,$$

而取这个值的概率为

$$p(x^{(n)}) = p(x_1)p(x_2)\cdots p(x_n).$$

WPS PDF编辑试用

例题 4.2.1：

设二进离散无记忆信源 $X_1, X_2, \cdots, X_n, \cdots$ 字符空间中分布为

$$X \sim p(x) = \begin{pmatrix} a_1 & a_2 \\ 0.25 & 0.75 \end{pmatrix},$$

考虑消息序列的编码问题。
首先 $H(X) = 0.562$nats$= 0.811$bits。
（1）随机变量 $p(X)$ 的分布
是：$p(X) \sim \begin{pmatrix} p(a_1) & p(a_2) \\ 0.25 & 0.75 \end{pmatrix} = \begin{pmatrix} 0.25 & 0.75 \\ 0.25 & 0.75 \end{pmatrix}$。

WPS PDF编辑试用

续解：

（2）2 长消息或随机变量 $X^{(2)}$ 的取值共有 4 个，分别为 $a_1a_1, a_1a_2, a_2a_1, a_2a_2$，而随机变量 $p(X^{(2)})$ 的取值为

$$p(a_1a_1) = 0.0625, p(a_1a_2) = 0.1875, p(a_2a_1) = 0.1875, p(a_2a_2) = 0.5625$$

其概率分布为

$$p(X^{(2)}) \sim \begin{pmatrix} 0.0625 & 0.1875 & 0.5625 \\ 0.0625 & 0.375 & 0.5625 \end{pmatrix},$$

而随机变量 $\log p(X^{(2)})$ 的分布律为

$$\log p(X^{(2)}) \sim \begin{pmatrix} \log 0.0625 & \log 0.1875 & \log 0.4625 \\ 0.0625 & 0.375 & 0.5625 \end{pmatrix}.$$

WPS PDF编辑试用

续解：

（3）3 长的消息或随机变量 $X^{(3)}$ 的取值为

$$a_1a_1a_1, a_1a_1a_2, a_1a_2a_1, a_2a_1a_1, a_2a_1a_2, a_1a_2a_2, a_2a_2a_1, a_2a_2a_2,$$

它的函数 $p(X^{(3)})$ 取值为

$p(a_1a_1a_1) = 0.015625, p(a_1a_1a_2) = 0.046875 = p(a_1a_2a_1) = p(a_2a_1a_1),$

$p(a_2a_1a_2) = p(a_2a_2a_1) = p(a_1a_2a_2) = 0.140625, p(a_2a_2a_2) = 0.421875,$

其分布律为

$$p(X^{(3)}) \sim \begin{pmatrix} 0.015625 & 0.046875 & 0.140625 & 0.421875 \\ 0.015625 & 0.140625 & 0.421875 & 0.421875 \end{pmatrix},$$

而随机变量 $\log p(X^{(3)})$ 的分布律为

$$\log p(X^{(3)}) \sim \left(\begin{array}{cccc} \log 0.015625 & \log 0.046875 & \log 0.140625 & \log 0.421875 \\ 0.015625 & 0.140625 & 0.421875 & 0.421875 \end{array} \right.$$

WPS PDF编辑试用

续解：

（4）如果要对 1 长的消息进行二进编码，则平均码长为 1，即每个信源字符占用 1bit。如果要对 2 长的消息进行如下二进编码

消息	a_1a_1	a_1a_2	a_2a_1	a_2a_2
码字	111	110	10	0

则平均码长为

$$\bar{L}=\sum_{x^{(2)}}p(x^{(2)})l(x^{(2)})=1.6875bits,$$

平均每个信源字符占用 $R=\bar{L}/2=0.84375$ bits，这说明对 2 长的消息编码时平均每个信源字符需要较少的码字符。

WPS PDF编辑试用

续解：

（5）如果要对 3 长的消息进行二进编码

消息	$a_1a_1a_1$	$a_1a_1a_2$	$a_1a_2a_1$	$a_2a_1a_1$	$a_2a_1a_2$	$a_1a_2a_2$	$a_2a_2a_1$
码字	11111	11110	11101	11100	101	100	110

则平均码长为

$$\bar{L} = \sum_{x^{(3)}} p(x^{(3)})l(x^{(3)}) = 2.433834 \text{ bits},$$

平均每个信源字符占用 $R = \bar{L}/3 = 0.811278$ bits，这说明对 3 长的消息编码时平均每个信源字符需要更少的码字符。因此在对信源编码时，通常对它的长消息进行编码，这样可能需要更少的码字符。信源字符占用的码符越少说明每个码符能携带的信息越多，对消息压缩的效果越好。

WPS PDF编辑试用

定理 4.2.2:

对离散无记忆信源 $X_1, X_2, \cdots, X_n, \cdots$ *成立*

$$-\frac{1}{n}\log p(X^{(n)}) \xrightarrow{P} H(X).$$

WPS PDF编辑试用

证明：

考虑随机变量序列 $\log p(X_k), k = 1, 2, \cdots$，它具有独立同分布，并且数学期望存在即 $E(\log p(X_k)) = -H(X)$，由辛钦大数定律得

$$-\frac{1}{n}\log p(X^{(n)}) = -\frac{1}{n}\sum_{i=1}^{n}\log p(X_i) \xrightarrow{P} H(X).$$

WPS PDF编辑试用

弱点型序列作用：

4.2.2

定理 ?? 6 告诉我们：对 $\varepsilon > 0$ 存在自然数 N_0，使当 $n > N_0$ 时

$$P\left\{\left|-\frac{1}{n}\log p(X^{(n)}) - H(X)\right| < \varepsilon\right\} > 1 - \varepsilon. \qquad (4.4)$$

如果消息分组长度为 n，令集合

$$W_\varepsilon^{(n)} = \left\{x^{(n)} \in \mathcal{X}^n \middle| \left|-\frac{1}{n}\log p(x^{(n)}) - H(X)\right| < \varepsilon\right\}, \qquad (4.5)$$

于是根据（4.1）可得

$$P\left\{X^{(n)} \in W_\varepsilon^{(n)}\right\} > 1 - \varepsilon \text{或} P\left\{X^{(n)} \notin W_\varepsilon^{(n)}\right\} \le \varepsilon, \qquad (4.6)$$

这说明不在集合 $W_\varepsilon^{(n)}$ 中的那些消息序列 $x^{(n)}$ 出现的概率很小。

WPS PDF编辑试用

消息序列分类：

定义4.2.2

称集合 $W_\varepsilon^{(n)}$ 为离散无记忆信源的 ε ***弱典型序列集****，其中的消息序列称为 n 长 ε* ***弱典型序列****，不在其中的消息称为 ε 非弱典型序列。*

WPS PDF编辑试用

续证明：

这样利用定理 4.2.2 可以将所有 n 长消息进行分类成弱典型序列与非弱典型序列，

$$\mathcal{X}^{(n)} = W_{\varepsilon}^{(n)} \bigcup \overline{W_{\varepsilon}^{(n)}}.$$

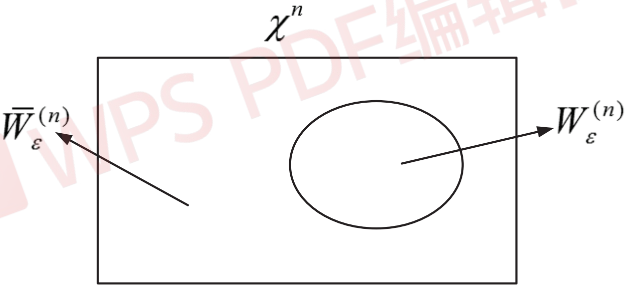

Figure: 图 4-5

WPS PDF编辑试用

定理 4.2.3：典型序列性质

若用以 2 为底的对数，对 $\varepsilon > 0$，当 n 充分大时

（1） *非弱典型序列出现的概率不大于 ε，而典型序列出现的概率不小于 $1-\varepsilon$，即*

$$P\left\{X^{(n)} \notin W_{\varepsilon}^{(n)}\right\} \leq \varepsilon \text{或} P\left\{X^{(n)} \in W_{\varepsilon}^{(n)}\right\} > 1-\varepsilon. \tag{5.1}$$

4.7

（2） *对于弱典型序列 $x^{(n)} \in W_{\varepsilon}^{(n)}$ 有*

$$2^{-n[H(X)+\varepsilon]} \leq p\left(x^{(n)}\right) \leq 2^{-n[H(X)-\varepsilon]}. \tag{5.2}$$

4.8

（3） *弱典型序列总数 M 满足*

$$(1-\varepsilon)2^{n[H(X)-\varepsilon]} \leq M \leq 2^{n[H(X)+\varepsilon]}. \tag{5.3}$$

4.9

WPS PDF编辑试用

证明：

第（1）已经证明。

第（2）证明。对 $x^{(n)} \in W_\varepsilon^{(n)}$ 则有

$$\left| -\frac{1}{n} \log p(x^{(n)}) - H(X) \right| < \varepsilon,$$

$$H(X) - \varepsilon \leq -\frac{1}{n} \log p(x^{(n)}) \leq H(X) + \varepsilon,$$

$$2^{-n(H(X)+\varepsilon)} \leq p(x^{(n)}) \leq 2^{-n(H(X)-\varepsilon)},$$

这就是每个 n 长消息序列出现的概率估计。

WPS PDF编辑试用

续证明：

第（3）条证明。对所有典型序列的概率求和得

$$\sum_{x^{(n)}} 2^{-n(H(X)+\varepsilon)} \leq \sum_{x^{(n)}} p(x^{(n)}) \leq \sum_{x^{(n)}} 2^{-n(H(X)-\varepsilon)},$$

则由上述不等式的左半边得

$$M2^{-n(H(X)+\varepsilon)} \leq \sum_{x^{(n)}} p(x^{(n)}) \leq 1,$$

$$M \leq 2^{n(H(X)+\varepsilon)},$$

由上述双联不等式的右半边得

$$1-\varepsilon \leq \sum_{x^{(n)}} p(x^{(n)}) \leq M2^{-n(H(X)-\varepsilon)},$$

$$M \geq (1-\varepsilon)2^{n(H(X)-\varepsilon)}.$$

WPS PDF编辑试用

典型序列的占比：

渐近等分性告诉我们：当 n 充分大时，n 长的弱典型序列 $x^{(n)}$ 出现的概率 $p(x^{(n)}) \approx 2^{-nH(X)}$；典型序列集 $W_\varepsilon^{(n)}$ 中元素的个数 $M \approx 2^{nH(X)}$，但所有 n 长消息序列的总数为 N^n，因此弱典型序列所占比例为

$$\frac{M}{N^n} \approx \frac{2^{nH(X)}}{2^{n\log N}} = 2^{-n[\log N - H(X)]},$$

由于 n 较大，故这个比例很小，这个事实可以建立离散无记忆信源的无失真压缩编码定理。

WPS PDF编辑账用

例题 4.2.2：求典型序列

弱

已知一个二进离散无记忆信源 $X_1, X_2, \cdots, X_n, \cdots$，字符空间中分布假定为 $p_0 = 0.36, p_1 = 0.64$，求长度不大于 55 并且 $\varepsilon = 0.1$ 所有典型序列个数及所占比例。

弱

利用 MATLAB 可以求得结果，如下表。其中标志为 1 表示弱典型序列集合的概率大于 $1-\varepsilon$。

WPS PDF编辑试用

续解：

消息长度	典型序列总数	占比	概率之和	标志
5	10	0.312500	0.339739	0
10	330	0.322266	0.488621	0
15	15808	0.482422	0.716721	0
20	425714	0.405993	0.756539	0
25	16708810	0.497961	0.857317	0
30	456508140	0.425156	0.816197	0
35	12610033866	0.367000	0.842402	0
40	480458963874	0.436975	0.901778	1
45	13471123866340	0.382872	0.913234	1
50	499693863728080	0.443817	0.945839	1
51	877717667768760	0.389785	0.920984	1
52	1526002175525520	0.338841	0.917612	1
53	3529761910333862	0.391882	0.938274	1

WPS PDF编辑试用