

第4章 离散无记忆信源无失真编码

4.1 信源编码概论

4.2 码的唯一可译性

4.3 定长编码定理和定长编码方法

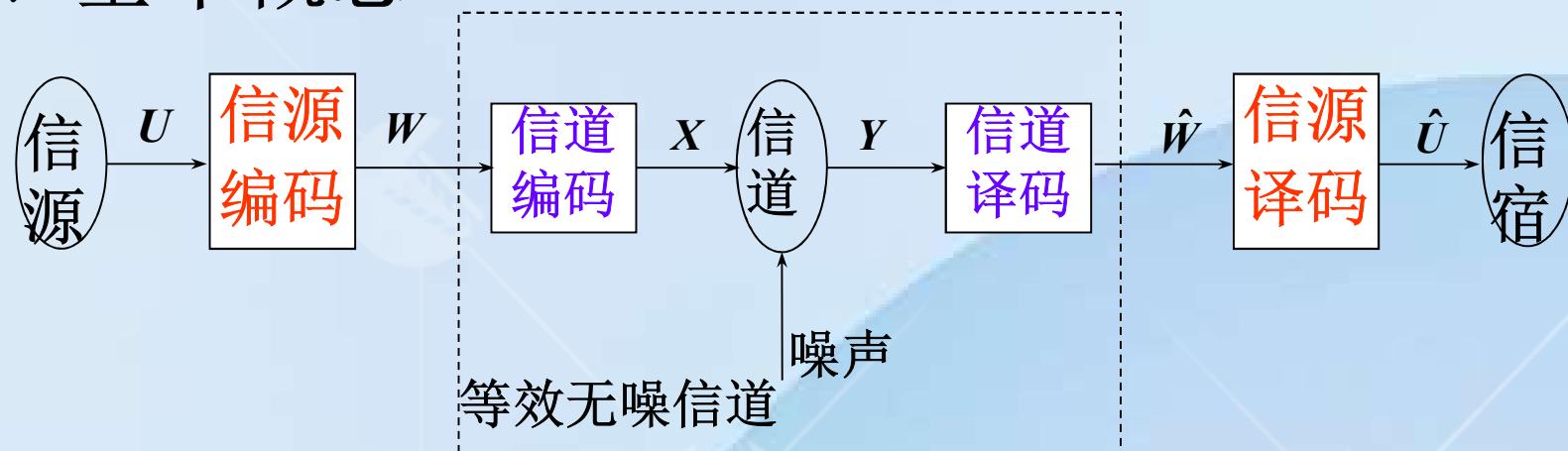
4.4 变长编码定理

4.5 变长编码方法

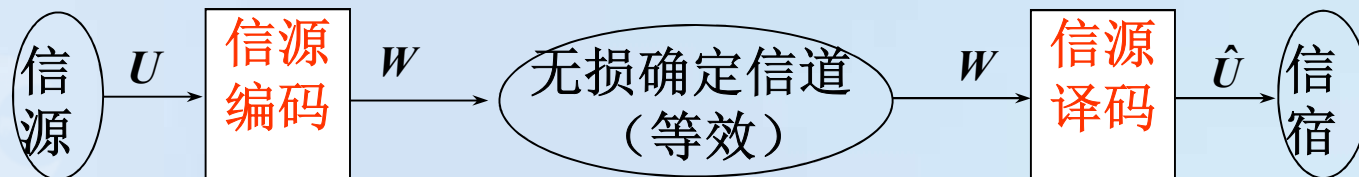
4.6 几种实用的无失真信源编码

4.1 信源编码概论

1、基本概念



- 传输之前的两次变换：信源编码、信道编码。
- 传输之后的两次反变换：信道译码、信源译码。
- 采取适当信道编码和译码措施后，可使信道传送的差错率降到允许的范围之内，因此，图中虚框部分可近似地视为一个等效的无损确定信道，简称为无噪信道，这是我们讨论信源编码

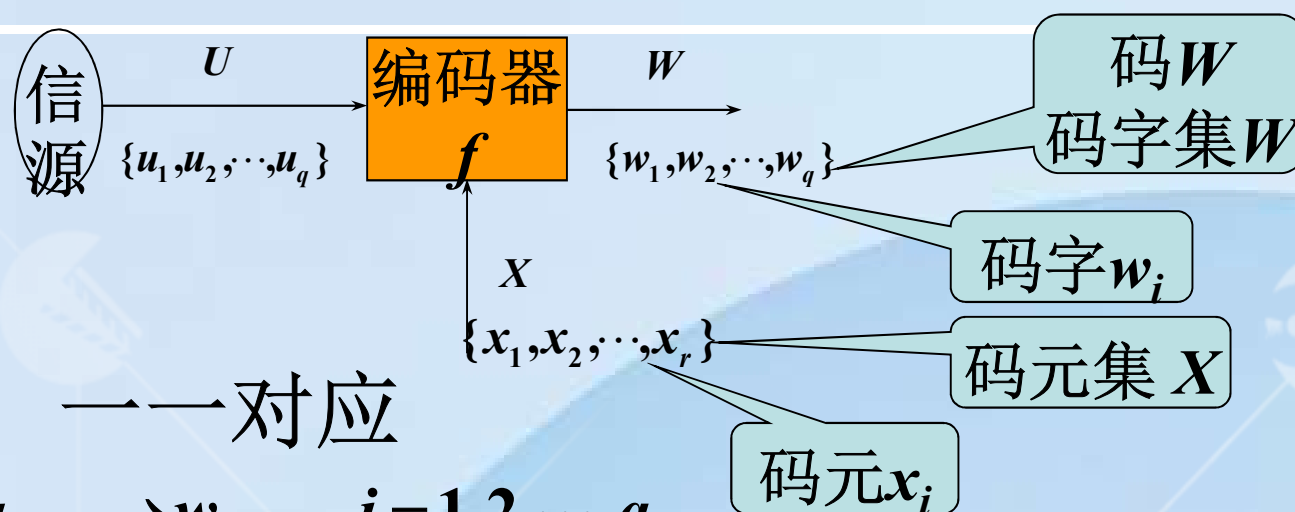


- 信源编码分类：无失真编码、有失真编码。
- 无失真编码：只对信源的冗余度进行压缩，不改变信源的熵，又称冗余度压缩编码，它能保证码元序列经译码后能无失真地恢复成信源符号序列。
- 有失真编码：又称熵压缩编码，将在第6章讨论。

无失真信源编码的作用：

- (1) 符号变换：使信源的输出符号与信道的输入符号相匹配；
- (2) 冗余度压缩：使编码之后的新信源概率分布均匀化，信息含量效率等于或接近于**100%**。

2、编码器模型



信源编码 f ：一一对应的变换。 $f: u_i \longrightarrow w_i$, $i = 1, 2, \dots, q$

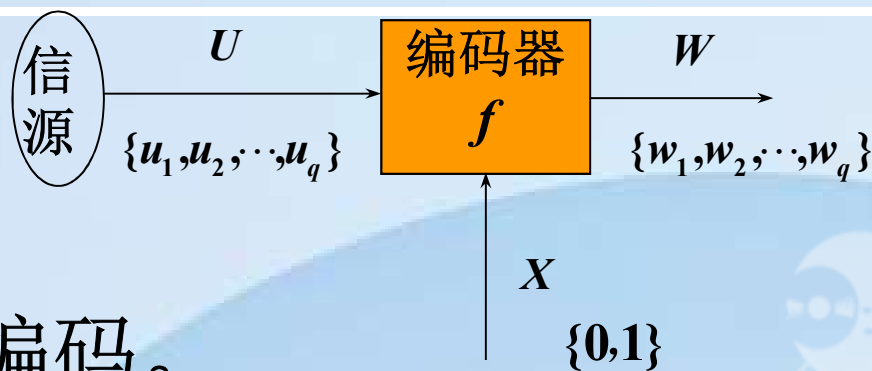
- 码长 l_i ：码字 w_i 所含码元的个数。单位：码元/符号
- 定长码 (FLC, Fixed Length code)：码中所有码字均有相同的码长 l ；否则称为变长码 (VLC, Variable Length code)
- 平均码长 $\bar{l} = \sum_{i=1}^q P(u_i) l_i = \sum_{i=1}^q P(w_i) l_i$ 码元/符号
- 定长码长 $\bar{l} = \sum_{i=1}^q P(u_i) l = l$ 码元/符号

例：编码

设**DMS**的概率空间为：

$$\begin{bmatrix} U \\ P_U \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & u_3 & u_4 \\ 1/2 & 1/4 & 1/8 & 1/8 \end{bmatrix}$$

对其单个符号进行二进制编码。



$$f_1 : \begin{cases} f_1(u_1) = w_1 = 00, l_1 = 2 \\ f_1(u_2) = w_2 = 01, l_2 = 2 \\ f_1(u_3) = w_3 = 10, l_3 = 2 \\ f_1(u_4) = w_4 = 11, l_4 = 2 \end{cases}$$

$$\begin{aligned} \bar{l} &= \sum_{i=1}^4 P(u_i) l_i = \frac{1}{2} \times 2 + \frac{1}{4} \times 2 + \frac{1}{8} \times 2 + \frac{1}{8} \times 2 \\ &= 2 \quad \text{码元/符号} \end{aligned}$$

编码策略：采用等长的码字。

$$f_2 : \begin{cases} f_2(u_1) = w_1 = 0, l_1 = 1 \\ f_2(u_2) = w_2 = 10, l_2 = 2 \\ f_2(u_3) = w_3 = 110, l_3 = 3 \\ f_2(u_4) = w_4 = 111, l_4 = 3 \end{cases}$$

$$\begin{aligned} \bar{l} &= \sum_{i=1}^4 P(u_i) l_i = \frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{8} \times 3 + \frac{1}{8} \times 3 \\ &= 1.75 \quad \text{码元/符号} \end{aligned}$$

编码策略：经常出现(概率大)的符号采用较短的码字，不经常出现(概率小)的符号采用较长的码字。