Codes

Recall from last weeks lectures, this table below where a source of size 4 has been encoded in binary codes with symbol 0 and 1.

X	Code 1	Code 2	Code 3	Code 4	Code 5	Code 6
x_1	00	00	0	0	0	1
x_2	01	01	1	10	01	01
x_3	00	10	00	110	011	001
x_4	11	11	11	111	0111	0001

Code Classifications

- **Code 1** This code is fixed length, but not distinct. Two symbols have the same binary representation. Due to this flaw it is no longer considered.
- **Code 2** This code is fixed length and distinct.
- **Code 3** This code is not uniquely decodable. Again due to this flaw, we will no longer consider it.

Prefix-free codes

- A prefix-free code is one in which no codeword is a prefix in another.
- Note that every prefix-free code is decipherable, but the converse is not true.
- In code 4, none of the codewords appear as prefixes for other codewords.
- For code 5, each code word are prefixes for the subsequent codeword.
- Both code 4 and 5 are uniquely decodable.
- Code 6 is prefix free and uniquely decodable.

Word Length

- These codes use code lengths between 1 and 4.
- For code 2; $n_1 = n_2 = n_3 = n_4 = 2$.
- For code 6; $n_1 = 1$, $n_2 = 2$, $n_3 = 3$,$n_4 = 4$.

Word Length

- Suppose that the symbols $\{x_1, x_2, x_3, x_4\}$ appear with the following probabilities $\{0.4, 0.3, 0.2, 0.1\}$
- The average code word length E(L) per source symbol is given by

$$E(L) = \sum_{i=1}^{m} P(x_i) n_i$$

- For each code compute $E(L)$.

Word Length

- **Code 1 and 2** Codes are fixed length $E(L) = 2$
- **Code 3** Recall: Code is flawed

$$E(L) = (0.4 \times 1) + (0.3 \times 1) + (0.2 \times 2) + (0.1 \times 2) = 1.3$$

- **Code 4**

$$E(L) = (0.4 \times 1) + (0.3 \times 2) + (0.2 \times 3) + (0.1 \times 3) = 1.9$$

- **Code 5 and 6**

$$E(L) = (0.4 \times 1) + (0.3 \times 2) + (0.2 \times 3) + (0.1 \times 4) = 2$$

Word Length : Interpretation

- Code 4 would require 190 binary digits to transmit 100 symbols. The transmission would be uniquely decodable.
- Code 3 would require 130 binary digits to transmit 100 symbols. The transmission would be not be uniquely decodable, and the intended message would be unclear.
- For the other codes, each would require 200 digits.
- Code 4 is seemingly the best choice.

Entropy and Efficiency

Given that the entropy of the input source is $H(X) = 1.85$b, compute the efficiency η for each code.

- **Code 1 and 2**
  $\eta = H(X)/E(L) = [1.85/2] \times 100\% = 92.5\%$
- **Code 3**
  Recall that this code is flawed
  $\eta = H(X)/E(L) = [1.84/1.3] \times 100\% = 142\%$
- **Code 4**
  $\eta = H(X)/E(L) = [1.85/1.9] \times 100\% = 97.2\%$
- **Code 5 and 6**
  $\eta = H(X)/E(L) = [1.85/2] \times 100\% = 92.5\%$

Instantaneous Codes

- Recall from previous lecture
- A uniquely decodable code is called an instantaneous code if the end of any code word is recognizable without examining subsequent code symbols.
- The instantaneous codes have the property that no code word is a prefix of another code word.
- Codes 2,4 and 6 are prefix-free codes, hence they are instantaneous codes.

Entropy Encoding

The design of a variable-length code such that its average code word length approaches the entropy of the DMS is often referred to as ***entropy coding***.

In this lecture, we will present two examples of entropy coding.

- Shannon-Fano Coding
- Huffman Coding

A. Shannon-Fano Coding:

An efficient code can be obtained by the following simple procedure, known as Shannon- Fano algorithm:

1. List the source symbols in order of decreasing probability.
2. Partition the set into two sets that are as close to equiprobable as possible, and assign 0 to the upper set and 1 to the lower set.
3. Continue this process, each time partitioning the sets with as nearly equal probabilities as possible until no further partitioning is possible.

A. Shannon-Fano Coding:

- Consider a 6 symbol alphabet: $\{x_1, x_2, x_3, x_4, x_5, x_6\}$ with corresponding probabilities $\{0.30, 0.25, 0.20, 0.12, 0.08, 0.05\}$
- Use the Shannon Fano coding algorithm to compute a variable length code.
- (On Overhead)

A. Shannon-Fano Coding:

- Compute the entropy

$$H(X) = (-0.30 \times \log_2(0.3)) + \ldots + (-0.05 \times \log_2(0.05)) = 2.36\text{b/symbol}$$

- Compute the average codeword length

$$E(L) = (0.30 \times 2) + (0.25 \times 2) + \ldots + (0.05 \times 4) = 2.38\text{b/symbol}$$

- Compute the efficiency of the code.

$$\eta = H(X)/E(L) = 2.36/2.38 = 0.99$$