Source Coding

- A conversion of the output of a DMS into a sequence of binary symbols (binary code word) is called ***Source Coding***.
- The device that performs this conversion is called the source encoder.
- An objective of source coding is to minimize the average bit rate required for representation of the source by reducing the redundancy of the information source.

Source Coding : Code Length and Code Efficiency

- Let X be a DMS with finite entropy $H(X)$ and an alphabet $\{x_1, \ldots, x_m\}$, each with corresponding probabilities of occurrence $P(x_i)$.
- Let the binary code word assigned to symbol x_i by the encoder have length n_i b.
- The length of a code word is the number of binary digits in the code word. The average code word length L, per source symbol is given by

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

Source Coding : Code efficiency and Code redundancy

- The parameter L (estimated by $E(L)$) represents the average number of bits per source symbol used in the source coding process.
- The code efficiency is defined as

$$\eta = \frac{L_{min}}{L}$$

  where L_{min} is the minimum possible value of L. When η approaches unity, the codes is said to be efficient.
- The code redundancy γ is defined as $\gamma = 1 - \eta$.

Source Coding Theorem

- The source coding theorem states that for a DMS X with entropy $H(X)$, the average code word length L per symbol is bounded as $L \geq H(X)$
- Furthermore L can be made as close to H(X) as required for some suitably chosen code.
- Thus, with $L_{min} \geq H(X)$, the code efficiency can be rewritten as

$$\eta = \frac{H(X)}{L}$$

- We will use this definition for efficiency. (Remark L is estimable by $E(L)$.)