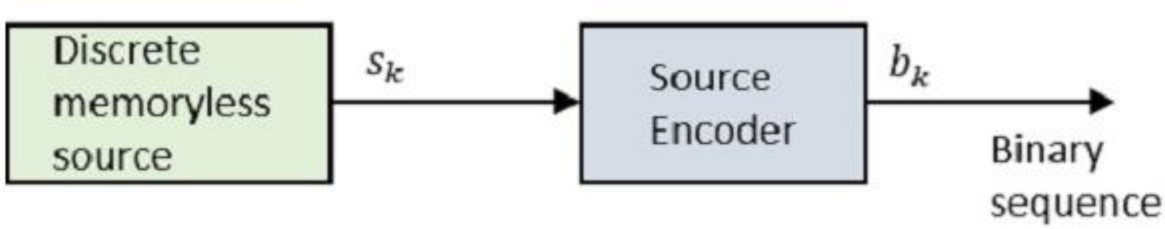
Source Coding Theorem

The Code produced by a discrete memoryless source, has to be efficiently represented, which is an important problem in communications. For this to happen, there are code words, which represent these source codes.

For example, in telegraphy, we use Morse code, in which the alphabets are denoted by **Marks** and **Spaces**. If the letter **E** is considered, which is mostly used, it is denoted by **"."** Whereas the letter **Q** which is rarely used, is denoted by **"--.-"**

Let us take a look at the block diagram.



Where **S_k** is the output of the discrete memoryless source and **b_k** is the output of the source encoder which is represented by **0s** and **1s**.

The encoded sequence is such that it is conveniently decoded at the receiver.

Let us assume that the source has an alphabet with **k** different symbols and that the **k^{th}** symbol **S_k** occurs with the probability **P_k**, where **k = 0, 1…k-1**.

Let the binary code word assigned to symbol **S_k**, by the encoder having length **l_k**, measured in bits.

Hence, we define the average code word length L of the source encoder as

$$\overleftarrow{L}=\sum_{k=0}^{k-1} p_k l_k$$

L represents the average number of bits per source symbol

If L_{min}=minimum possible value $\overleftarrow{L}$

Then **coding efficiency** can be defined as

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

With $\bar{L} \geq L_{min}$ we will have $\eta \leq 1$

However, the source encoder is considered efficient when $\eta=1$

For this, the value L_{min} has to be determined.

Let us refer to the definition, *"Given a discrete memoryless source of entropy* $H(\delta)$*, the average code-word length* ***L*** *for any source encoding is bounded as* $\bar{L} \geq H(\delta)$.

Hence with $L_{min}=H(\delta)$ the efficiency of the source encoder in terms of Entropy $H(\delta)$H(δ) may be written as

$$\eta = \frac{H(\delta)}{\bar{L}}$$