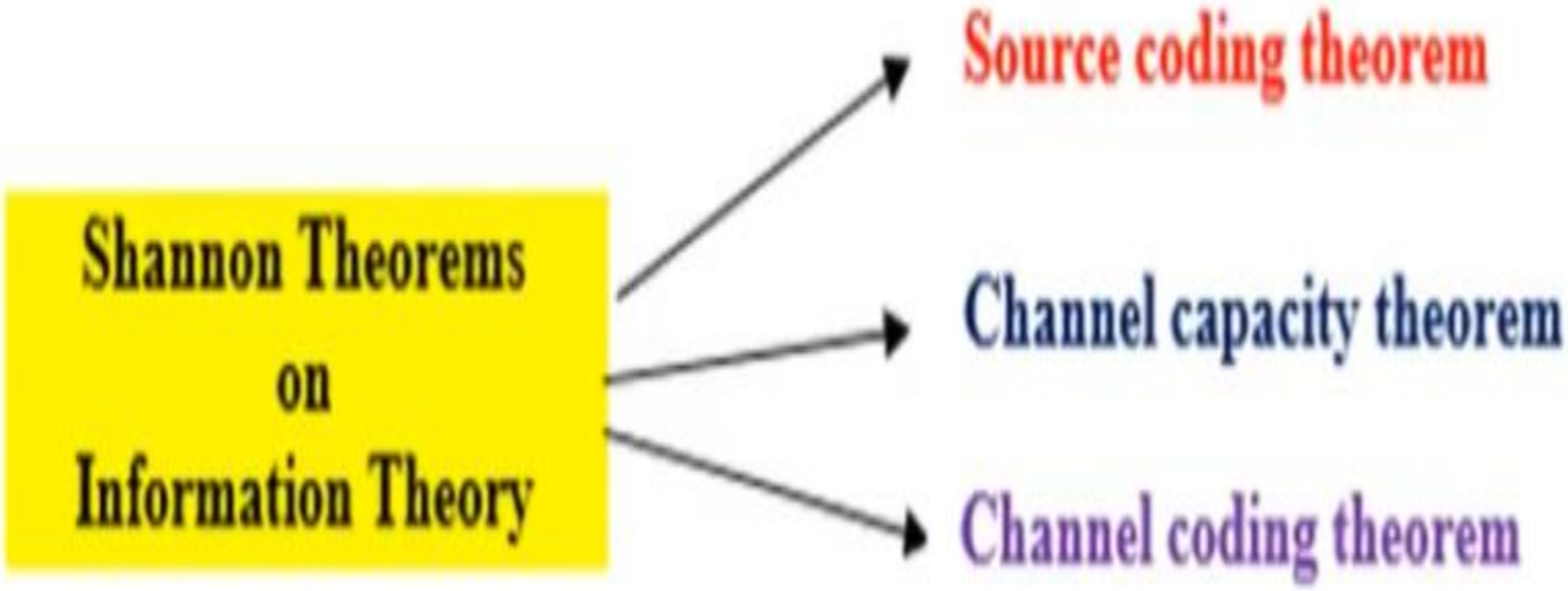
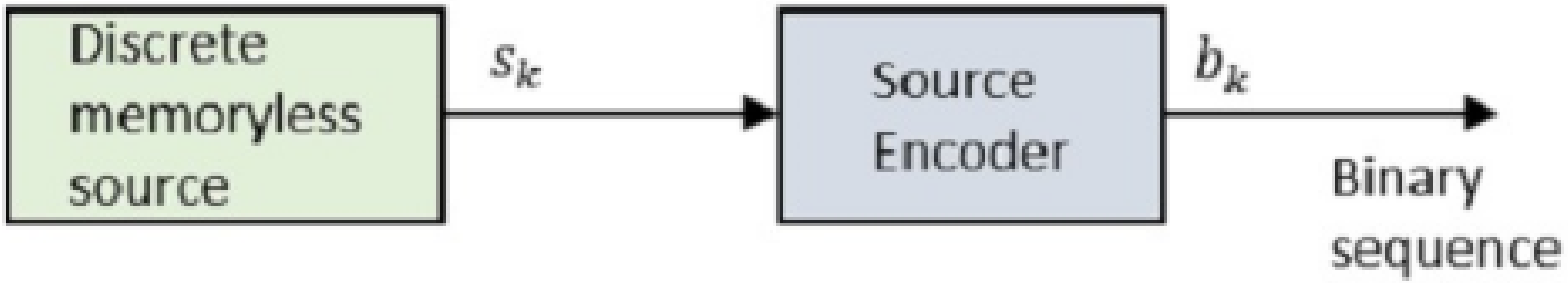
SOURCE CODING THEOREMS I & II


Shannon Theorems
on
Information Theory
Source coding theorem
Channel capacity theorem
Channel coding theorem

Source coding

Definition

- Conversion of the o/p of a discrete memoryless source(DMS) into a sequence of binary symbols (ie binary codeword).
- Device that performs this conversion – source encoder.


Discrete memoryless source
s_k
Source Encoder
b_k
Binary sequence

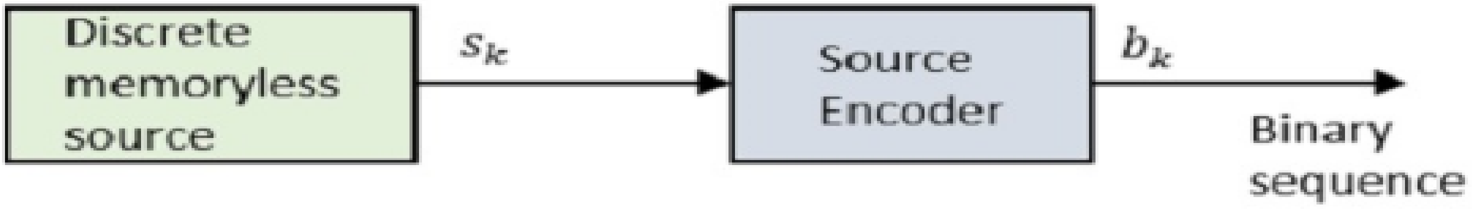
<u>Objective</u>

- To minimize avg bit rate required for representation of the source by reducing the redundancy of the information source.

<u>Terms related to Source coding process</u>

i. Codeword length
ii. Average codeword length ($\bar{L}$)
iii. Code efficiency
iv. Code redundancy

Source Coding theorem I



$\mathbf{S_k}$ is the discrete memory less source output and the b_k is the source encoder output which is represented by 0s and 1s.

- The encoded sequence is easily decoded at the receiver.

- Consider the source has an alphabet with k different symbols and the k^{th} symbol S_k occurs with the probability P_k, where k = 0, 1…k-1.
- S_k is assigned with binary code word b_k, by the encoder having length l_k, is measured in bits.
- The average code word length $\bar{L}$ of the source encoder is defined as

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

- $\bar{L}$ represents the average number of bits per source symbol

- If **L_{min}**=minimum possible value of $\bar{L}$, then coding efficiency is defined as

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

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

- when η approaches unity, the code is said to be efficient (source encoder is efficient when **η=1**).

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

- Given a discrete memory less source of entropy H(S), the average code-word length $\overline{L}$ for any source encoding can bounded as $\overline{L} \geq H(S)$.

- In simple terms, the code word is always greater than or equal to the source code. Code word symbols are greater than or equal to the source code alphabets.

- Therefore with L_{min}=H(S), the source encoder efficiency in terms of Entropy H(S) can be written as

$$\eta = H(S)/\,\bar{L}$$

- This is known as noiseless coding theorem as it provides encoding which is error-free, also known as **Shannon's first** theorem.

- Code redundancy

$$\gamma = 1 - \eta$$

By increasing redundancy of encoding, we can make the prob of error approach 0.

Source Coding theorem II

- Let there be a discrete memory less source with entropy *H(S)*.
- It is always possible to find a *q*-ary prefix code such that the expected value of the length of the code words $\bar{L}$ allocated to the sequences of *L* source symbols satisfies:

$$\frac{H(S)}{\log q} \le \bar{L} < \frac{H(S)}{\log q} + 1/L$$