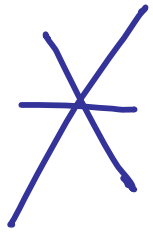


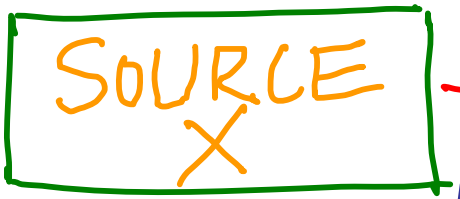
SOURCE CODING

DATA COMPRESSION

Very important in modern communication esp of audio/video/images.

Storage/Archive

Multimedia Content



$x(0), x(1), x(2), \ldots$

symbols in time generated by source.

$x(k) \in S = \{s_0, s_1, \ldots, s_{M-1}\}$

"Source Alphabet"

Discrete Memoryless Source (DMS)

Probabilities of symbols at any time k, do not depend on symbols generated previously.

$$P(x(k) = s_i \mid x(k-1), x(k-2), \ldots x(0))$$

$$= P(x(k) = s_i) = P_i$$

Probabilities of $x(k)$ are fixed and DO NOT depend on past symbols.

Discrete Memoryless Source.

One of the most commonly employed source models.

SOURCE CODING:

convert each symbol to a sequence of binary symbols or codewords.

would like to minimizing the average number of bits used to represent each symbol.

Mapping from symbols to codewords is termed as a code.

Ex: consider $M = 4$

s_0, s_1, s_2, s_3

Fixed Length code:

$s_0 \rightarrow$	1 1
$s_1 \rightarrow$	1 0
$s_2 \rightarrow$	0 1
$s_3 \rightarrow$	0 0

Each symbol is represented by 2 bits

$M = 4$ symbols.

1 0 0 1 0 0 . . .

s_1 s_2 s_3

Uniquely Decodable.

Fixed Length Code.

Low Complexity of Decoding.

Variable Length Code:

$M = 4$

$s_0 \rightarrow 0$ — 1 bit
$s_1 \rightarrow 1$ — 1 bit
$s_2 \rightarrow 00$ — 2 bits
$s_3 \rightarrow 11$ — 2 bits

Variable Length Code. (VLC)

Different symbols are represented by different numbers of bits.

Problem:

0 0 0

s_0 s_2

s_0 s_0 s_0

s_2 s_0.

Same sequence 000 can correspond either to $s_0 s_2$, $s_2 s_0$ or $s_0 s_0 s_0$.

Hence, code is NOT uniquely Decodable.

Uniquely Decodable:

Code is uniquely Decodable if the original source sequence can be reconstructed perfectly from the coded sequence.