Modeling of Information Sources

Digital Communication System

\[ x(t) \xrightarrow{\text{Lowpass Filter}} \text{Sampling} \xrightarrow{\text{Quantizer}} \text{Encoding} \xrightarrow{\text{Shannon Coding}} \text{Optimize} \]

Shannon Encoders -- used to encode a signal in a manner that optimizes the use of data (bits).

In other words, it eliminates from the original signal everything that does not represent useful information.

Information:

To define information, we consider a discrete memoryless source (DMS):

\[ \{X_1, X_2, X_3, \ldots \} \]

Alphabet of the source:

\[ \sum_{k=1}^{M} \]

\[ M \]: size of the alphabet

Let

\[ P_i = \Pr [ X = X_i ] = \text{pmf} (X_i), \quad i = 1, \ldots, M \]
**Example:** Consider a computer as an information source

The size of the alphabet is $M = 2$

$X_i \in \{0, 1\}$

$P_r \{X = 0\} = P_0 = p$

$P_r \{X = 1\} = P_1 = 1 - p$

Binary information source

Self information — information contained in each of the transmitted symbols

$I(X_i) = \log_b \left( \frac{1}{P_i} \right), \quad i = 1, \ldots, M$

* Information is a logarithmic function of the PMF

* Different values for $b$ used in practice
  
  If $b = 10 \rightarrow$ Hartley information

  $b = 2 \rightarrow$ Shannon information / bits

  $b = e \rightarrow$ hart (natural unit)

**Example:** Consider a binary source

$P_r \{X = 0\} = p \leq 1$

$P_r \{X = 1\} = 1 - p \leq 1$
Self-information in $X = 0$:

$$I(0) = \log_2 \left( \frac{1}{p} \right) = -\log_2(p)$$

If $p = \frac{1}{2}$ (equal probability of 1 and 0):

$$I(0) = I(1) = \log_2 \left( \frac{1}{\frac{1}{2}} \right) = 1 \text{ bit}$$

Therefore, 1 bit is the information contained in a message about the realization of a random process that has two equally probable outcomes.

**Information at discrete sources:**

* In practical situations, self-information of an individual signal may not be important.
* More important is average information generated by the source.

**Average information generated by the source (Entropy):**

$$H(X) = \sum_{i=1}^{M} P(x_i) \log_2 \left( \frac{1}{P(x_i)} \right)$$

As for: A discrete random variable $X$

$$E[X] = \sum_{i=1}^{M} x_i \cdot P_i$$

$$\sum_{i=1}^{M} P_i = 1$$

$$M$$: Size of source alphabet.

**Example:**
The quantizer possible outputs are $X_1, X_2, X_3, X_4, X_5$

\[
\begin{align*}
\Pr(X_1) &= \frac{1}{2} \\
\Pr(X_2) &= \frac{1}{4} \\
\Pr(X_3) &= \frac{1}{8} \\
\Pr(X_4) &= \frac{1}{16}
\end{align*}
\]

Self Information:

\[
\begin{align*}
I(X_1) &= \log_2\left(\frac{1}{0.5}\right) = 2 \text{ bits} \\
I(X_2) &= \log_2\left(\frac{1}{4}\right) = 2 \text{ bits} \\
I(X_3) &= \log_2\left(\frac{1}{8}\right) = 3 \text{ bits} \\
I(X_4) &= I(X_5) = \log_2(16) = 4 \text{ bits}
\end{align*}
\]

The average Information (entropy) of the source:

\[
H(X) = \frac{1}{2} \cdot x_1 + \frac{1}{4} \cdot x_2 + \frac{1}{8} \cdot x_3 + \frac{1}{16} \cdot x_4 \cdot x_5 = \boxed{1.875 \text{ bits}}
\]

Information bit rate $\Rightarrow R = f_s \cdot H(X)$

\[
= 8000 \times 1.875 = 15 \text{ Kbits/sec}
\]

Maximum entropy of a discrete source:

Consider a source with $M$ symbols

alphabet: $\{X_1, X_2, ..., X_M\}$

\[
P_X(x_i) = p_i, \quad i = 1, ..., M
\]

The entropy of the source:

\[
H(X) = - \sum_{i=1}^{M} p_i \cdot \log_2(p_i)
\]

→ What set of probabilities maximizes the entropy
**Constrained maximization problem:**

Cost function

\[
\text{maximize } H(x) = -\sum_{i=1}^{M} k_i \cdot \log_2 (p_i)
\]

subject to

\[
\sum_{i=1}^{M} k_i = 1
\]

Method of Lagrange multipliers:

1. Form the Lagrange function

\[
\mathcal{L}(p_1, p_2, \ldots, p_M) = -\sum_{i=1}^{M} k_i \cdot \log_2 (p_i) + \lambda \left( 1 - \sum_{i=1}^{M} p_i \right)
\]

2. \( \frac{\partial \mathcal{L}(\cdot)}{\partial p_i} = 0 \) \( \quad i = 1, \ldots, M \)

\[
-\log_2 (p_i) - k_i \cdot \frac{1}{p_i} - \lambda = 0 \quad \text{(1)}
\]

2. \( \frac{\partial \mathcal{L}(\cdot)}{\partial \lambda} = 0 \) \( \Rightarrow \left( \sum_{i=1}^{M} p_i = 1 \right) \quad \text{(2)}
\]

(1) \( \Rightarrow \quad \log_2 (p_i) + 1 + \lambda = 0 \) \( \quad i = 1, \ldots, M \) \( \text{(3)} \)

(2) \( \Rightarrow \quad \sum_{i=1}^{M} p_i = 1 \) \( \text{(4)} \)

(3) \( \Rightarrow \quad p_i = 2^{-\left(1+\lambda \right)} \quad \text{(5)} \)

(4) \[ \sum_{i=1}^{M} 2^{-\left(1+\lambda \right)} = 1 \Rightarrow M \cdot 2^{-\left(1+\lambda \right)} = 1 \]
\[
\log_2 (p_i) + 1 - 1 + \log_2 (M) = 0
\]
\[
\Rightarrow \log_2 (p_i) = \log_2 \left( \frac{1}{M} \right)
\]
\[
\Rightarrow p_i = \frac{1}{M}, \quad i = 1, \ldots, M
\]

\[
\Rightarrow P_i = \frac{1}{M} \Rightarrow \text{Max entropy is obtained for a uniform PMF}
\]

The maximum entropy is

\[
H_{\text{max}} = - \sum_{i=1}^{M} p_i \log_2 (p_i) = - \sum_{i=1}^{M} \frac{1}{M} \log_2 \left( \frac{1}{M} \right)
\]
\[
= \frac{1}{M} \cdot M \cdot \log_2 (M) = \log_2 (M)
\]

Example: \( M = 5 \) : \( H_{\text{max}} = \log_2 (5) = \log_2 (5) \)
\[
= 2.32 \text{ bits/sample}
\]

If \( M = 3 \) \( H_{\text{max}} = \log_2 (3) = 3 \text{ bits} \)

Redundancy:

Redundancy of a discrete source that has \( M \) symbols in its alphabet is defined as: \( R \)

\[
R = \frac{H_{\text{max}} - H(x)}{H_{\text{max}}} \times 100
\]

\( H_{\text{max}} \): maximum entropy for a source with \( M \) symbols
Maximum entropy: $H_{\text{max}} = \log_2(M)$

$H(x)$: entropy of the source

**Example:**

$$\sum_{i=1}^{5} x_i^2 = 5$$

$$p_1 \sum x_1 = \frac{1}{2}$$
$$p_2 \sum x_2 = \frac{1}{4}$$
$$p_5 \sum x_5 = \frac{1}{8}$$

$$H(x) = 1.875 \text{ bits/sample}$$

$$H_{\text{max}} = \log_2(5) = 2.32 \text{ bits/sample}$$

Worst scenario

$$R = \frac{2.32 - 1.875}{2.32} = 0.1918$$

(Source Coding Theorem) A source with entropy $H(x)$ can be encoded with an arbitrary small error probability at any $R_s$ as long as $R_s > H(x)$. Conversely, if $R_s < H(x)$, then the error probability will be bounded away from zero independent of the employed encoder and decoder.

**Meaning:** Entropy is essentially the minimum number of bits per source symbol that needs to be used for lossless (error-free) encoding of the source.

**Example:** Source with 8 symbols

$p_1 = \frac{1}{2}$
$p_2 = \frac{1}{3}$

If the encoder is binary, how many bits per symbol does it need?
\[ P_1 = P_4 = P_5 = P_6 = P_7 = P_8 = \frac{1}{36} \]

Non-statistical approach: 8 Symbols \( \Rightarrow \) 8 bits

\[(\sigma, 0, 0) \rightarrow 1\]

\[(\sigma, 1, 0) \rightarrow 0\]

\* Statistical approach:

\[ I_1 = \log_2 \left( \frac{1}{0.2} \right) = 1 \times \frac{1}{2} \]

\[ I_2 = \log_2 (3) = 1.5850 \times \frac{1}{3} \]

\[ I_3 = \log_2 (36) = 5.1633 \times \frac{1}{36} \]

\[ I_4 = \ldots = 5.1633 \times \frac{1}{36} \]

\[ I_5 = \ldots = 5.1633 \times \frac{1}{36} \]

\[ I_6 = \ldots = 5.1633 \]

\[ I_7 = \ldots = 5.1633 \]

\[ I_8 = \ldots = 5.1633 \]

\[ R = \frac{3 - 1.89}{3} = 37\% \]

\[ H = 1.89 < 3\]