Chapter 2: Source Coding

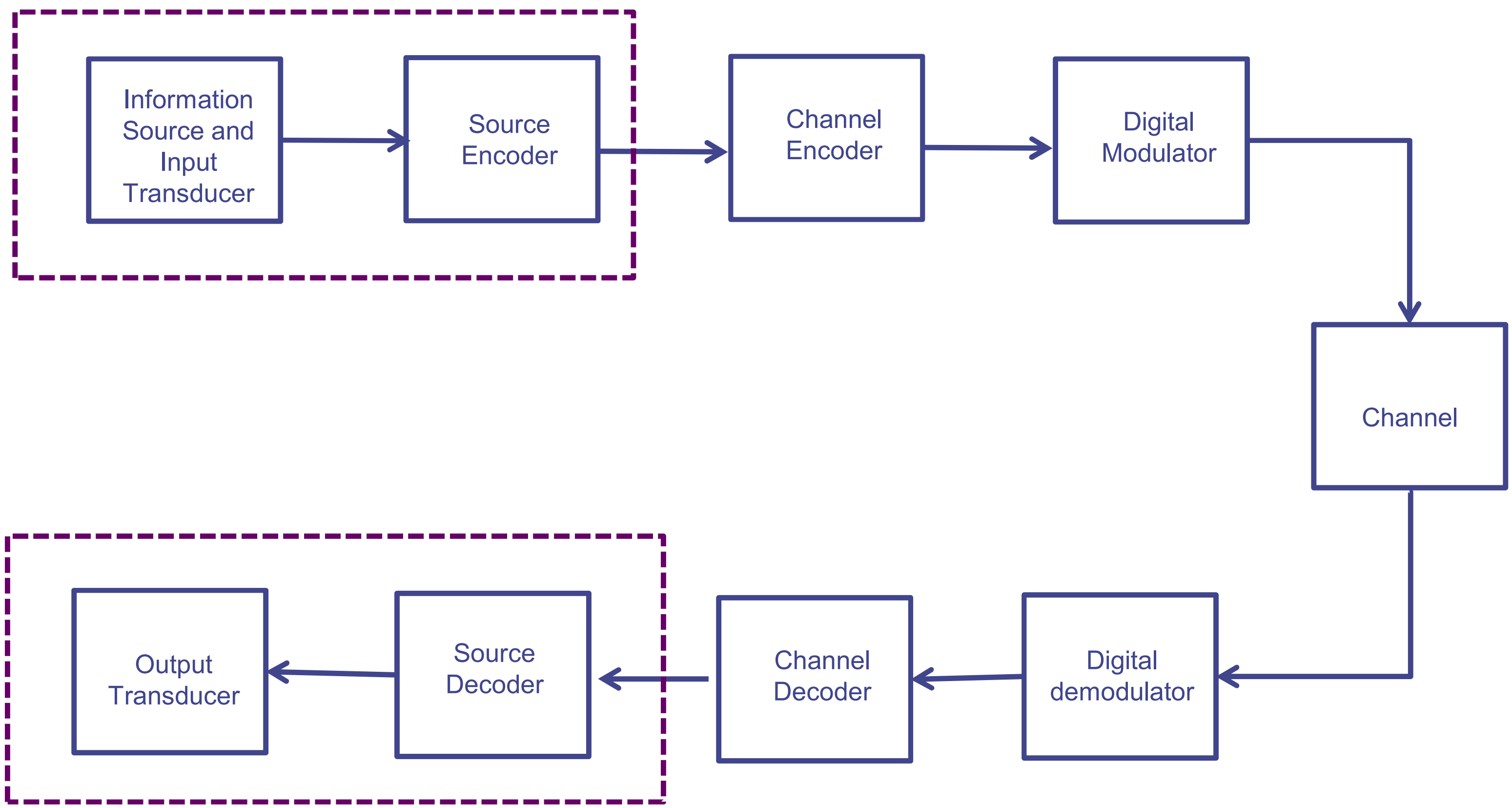
AAiT
Addis Ababa Institute of Technology
አዲስ አበባ ቴክኖሎጂ ኢንስቲትዩት
Addis Ababa University
አዲስ አበባ ዩኒቨርሲቲ

Graduate Program
School of Electrical and Computer Engineering

Overview

- Source coding and mathematical models
- Logarithmic measure of information
- Average mutual information

Source Coding



Functional block diagram of a typical digital communication system

Source Coding

- Information sources may be *analog* or *discrete (digital)*
  - Analog sources: Audio or video
  - Discrete sources: Computers, storage devices such as magnetic or optical devices
- Whatever the nature of the information sources, digital communication systems are designed to transmit information in digital form
- Thus the output of sources must be converted into a format that can be transmitted digitally

Source Coding …

- This conversion is generally performed by the source encoder, whose output may generally assumed to be a sequence of binary digits
- Encoding is based on mathematical models of information sources and the quantitative measure of information emitted by a source
- We shall first develop mathematical models for information sources

Mathematical Models for Information Sources

- Any information source produces an output that is random and which can be characterized in statistical terms
- A simple discrete source emits a sequence of letters from a finite alphabet
  - Example: A binary source produces a binary sequence such as 1011000101100; alphabet: {0,1}
  - Generally, a discrete information with an alphabet of L possible letters produces a sequence of letters selected from this alphabet
- Assume that each letter of the alphabet has a probability of occurrence p_k such that

$$p_k = P\{X = x_k\}, \quad 1 \le k \le L; \text{ and}$$

$$\sum p_k = 1$$

Two Source Models

1. Discrete Sources
   - Discrete Memoryless source (DMS): Output sequences from the source are statistically independent
     - Current output does not depend on any of the past outputs
   - Stationary source: The discrete source outputs are statistically dependent but statistically stationary
     - That is, two sequences of length n, $(a_1, a_2, \ldots.. a_n)$ and $(a_{1+m}, a_{2+m}, \ldots.. a_{n+m})$ each have joint probabilities that are identical for all $n \geq 1$ and for all shifts m

2. Analog source has an output waveform x(t) that is a sample function of a stochastic process **X**(t), where we assume **X**(t) is a ***stationary process*** with autocorrelation $R_{xx}(\tau)$ and power spectral density $\boldsymbol{\Phi}_{xx}(f)$

Mathematical Models …

- When **X**(t) is ***bandlimited*** stochastic process such that $\Phi_{xx}(f) = 0$ for $|f| \geq W$; by the sampling theorem

$$X(t) = \sum_{n=-\infty}^{\infty} X\left(\frac{n}{2w}\right) \frac{\sin\left(2\pi W\left(t - \frac{n}{2W}\right)\right)}{2\pi W\left(t - \frac{n}{2W}\right)}$$

  - Where {X(n/2W)} denote the samples of the process at the sampling rate (Nyquist rate) of $f = 2W$ samples/s

- Thus applying the sampling theorem we may convert the output of an analog source into an equivalent discrete-time source

Mathematical Models …

- The output is statistically characterized by the ***joint*** pdf $p(x_1, x_2, \ldots x_m)$ for $m \geq 1$ where $x_n = X(n/2W)$; $1 \leq n \leq m$, are the random variables corresponding to the samples of $X(t)$
- Note that the samples $\{X(n/2W)\}$ are in general continuous and cannot be represented digitally without loss of precision
- Quantization may be used such that each sample is a discrete value, but it introduces distortion (*More on this later*)

Overview

- Source coding and mathematical models
- Logarithmic measure of information
- Average mutual information

Logarithmic Measure of Information

- Consider two discrete random variables **X** and **Y** such that

$$\mathbf{X} = \{x_1, x_2, \ldots\ldots x_n\} \text{ and } \mathbf{Y} = \{y_1, y_2, \ldots\ldots\ldots y_m\}$$

- Suppose we observe **Y** = y_j and wish to determine, quantitatively, the amount of information **Y** = y_j provides about the event

$$\mathbf{X} = x_i, \qquad i = 1,2,\ldots\ldots n$$

- Note that if **X** and **Y** are statistically independent Y = y_j ***provides no information*** about the occurrence of **X** = x_i
- If they are fully dependent **Y** = y_j determines the occurrence of **X** = x_i
  - The information content is simply that provided by **X** = x_i

Logarithmic Measure of Information …

- A suitable measure that satisfies these conditions is the logarithm of the ratio of the conditional probabilities $P\{\mathbf{X} = x_i | \mathbf{Y} = y_j\} = P\{x_i | y_j\}$ and $P\{\mathbf{X} = x_i\} = p(x_i)$
- Information content provided by the occurrence of $\mathbf{Y} = y_j$ about the event $\mathbf{X} = x_i$ is defined as

$$I(x_i, y_j) = \log \frac{P(x_i / y_j)}{P(x_i)}$$

- $I(\mathbf{x}_i, \mathbf{y}_j)$ is called the mutual information between x_i and y_j
- Unit of the information measure is the *nat(s)* if the natural logarithm is used and *bit(s)* if base 2 is used
- Note that $\ln a = \ln 2 . \log_2 a = 0.69315 \log_2 a$

Logarithmic Measure of Information …

- If X and Y are independent $I(x_i , y_j) = 0$
- If the occurrence of **Y** = y_j uniquely determines the occurrence of **X** = x_i,

$$I(x_i , y_j) = \log (1/ p(x_i))$$

- i.e., $I(x_i , y_j) = -\log p(x_i) = I(x_i)$ – ***self information*** of event X
- Note that a high probability event conveys less information than a low probability event
- For a single event x where $p(x) = 1$, $I(x) = 0$ the occurrence of a sure event ***does not convey*** any information

Logarithmic Measure of Information …

Example

1. A binary source emits either 0 or 1 every τ_s seconds with equal probability. Information content of each output is then given by

$$I(x_i) = -log_2\, p = -log_2\, 1/2 = 1 \text{ bit}, \qquad x_i = 0,1$$

2. Suppose successive outputs are statistically independent (DMS) and consider a ***block*** of binary digits from the source in time interval $k\tau_s$

- Possible number of k-bit block = 2^k = M, each of which is *equally probable* with probability 1/M = 2^{-k}

$$I(x_i) = -log_2 2^{-k} = k \text{ bits in } k\tau_s \text{ sec}$$

(Note the additive property)

Logarithmic Measure of Information …

- Now consider the following relationships

$$I(x_i, y_j) = log\frac{p(x_i|y_j)}{p(x_i)} = log\frac{p(x_i|y_j)p(y_j)}{p(x_i)p(y_j)} = log\frac{p(x_i, y_j)}{p(x_i)p(y_j)}$$

- and since $p(y_j|x_i) = \frac{p(x_i, y_j)}{p(x_i)}$

$$I(x_i, y_j) = \frac{p(y_j|x_i)}{p(y_j)} = I(y_j, x_i)$$

- Information provided by y_j about x_i is the same as that provided by x_i about $Y = y_j$

Logarithmic Measure of Information …

- Using the same procedure as before we can also define ***conditional self-information*** as

$$I(x_i|y_j) = \log_2 \frac{1}{p(x_i|y_j)} = -\log_2 p(x_i|y_j)$$

- Now consider the following relationships

$$log\left[\frac{p(x_i|y_j)\,p(x_i)}{p(x_i)}\right] = I(x_i, y_j) - I(x_i)$$

- Which leads to $I(x_i, y_j) = I(x_i) - I(x_i|y_j)$

- Note that $I(x_i|y_j)$ is self information about X = x_i after having observed the event Y = y_j

Logarithmic Measure of Information …

- Further, note that

$$I(x_i) \geq 0 \qquad and \qquad I(x_i|y_j) \geq 0$$ and thus

$$I(x_i, y_j) \geq 0 \quad when \qquad I(x_i) > I(x_i|y_j)$$

- Indicating that the mutual information between two events can either be positive or negative

Overview

- Source coding and mathematical models
- Logarithmic measure of information
- Average mutual information

Average Mutual Information

- Average mutual information between **X** and **Y** is given by

$$I(\mathbf{X};\mathbf{Y}) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) I(x_i, y_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \log_2 \frac{p(x_i, y_j)}{p(x_i)p(y_j)} \geq 0$$

- Similarly, average self-information is given by

$$H(\mathbf{X}) = \sum p(x_i) I(x_i) = -\sum p(x_i) \log_2 p(x_i)$$

- where **X** represents the alphabet of possible output letters from a source
- H**(X)** represents the average self information per source letter and is called the ***entropy*** of the source

Average Mutual Information …

- If $p(x_i) = 1/n$ for all i, then the entropy of the source becomes

$$H(X) = -\sum \frac{1}{n} \log_2 \frac{1}{n} = \log_2 n$$

- In general, **H(X) ≤** log n for any given set of source letter probabilities
  - Thus, the entropy of a discrete source is maximum when the output letters are all equally probable

Average Mutual Information …

- Consider a binary source where the letters {0,1} are independent with probabilities

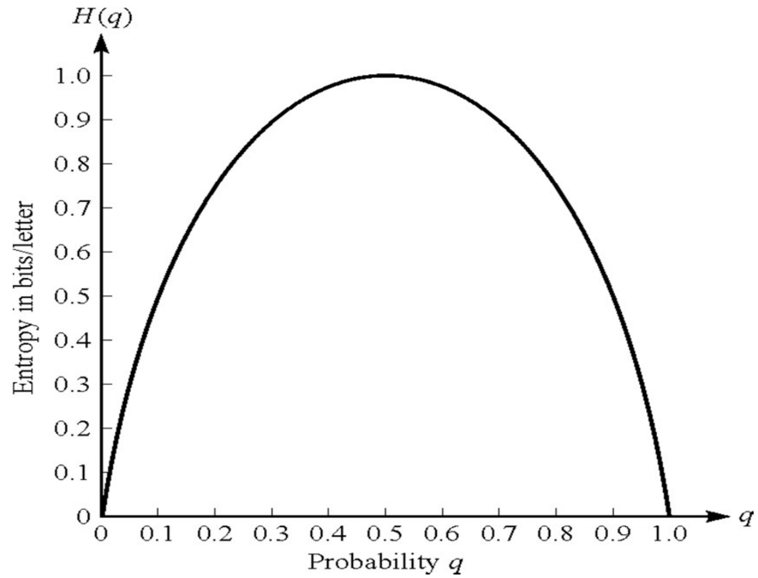
$$p(x_0=0)=q \text{ or } p(x_1=1)=1-q$$

- The entropy of the source is given by

$$H(X)= -q\, log_2 q - (1-q)\, log_2(1-q) = H(q)$$

- whose plot as a function is shown below

Binary Entropy Function



Average Mutual Information …

- In a similar manner as above, we can define ***conditional self-information*** or conditional entropy as follows

$$H(\mathbf{X}/\mathbf{Y}) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log\frac{1}{p(x_i / y_j)}$$

- Conditional entropy is the information or uncertainty in **X** after **Y** is observed
- From the definition of average mutual information one can show that

$$\begin{aligned}\mathbf{I}(\mathbf{X},\ \mathbf{Y}) &= \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\left[\log\ p(x_i / y_j) - \log\ p(x_i)\right] \\ &= -H(\mathbf{X/Y}) + H(\mathbf{X})\end{aligned}$$

Average Mutual Information …

- Thus $I(X,Y) = H(X) – H(X/Y)$ and $H(X) \geq H(X|Y)$, with equality when X and Y are independent
  - H(X/Y) is called the equivocation
  - It is interpreted as the amount of average uncertainty remaining in X after observation of Y
- H(x) – Average uncertainty prior to observation
- I(X;Y) – Average information provided about the set X by the observation of the set Y

Average Mutual Information …

- The above results can be generalized for more than two random variable
- Suppose we have a block of k random variables $x_1,x_2,\ldots..x_k$ with joint probability $P(x_1,x_2,\ldots..x_k)$
- The entropy of the block will then be given by

$$H(X) = -\sum_{j1}^{n1}\sum_{j2}^{n2}\ldots\sum_{jk}^{nk} P(x_{j1},x_{j2}\ldots.x_{jk})\log P(x_{j1},x_{j2}\ldots.x_{jk})$$

Continuous Random Variables - Information Measure

- If X and Y be continuous random variables with joint pdf *f (x,y)* and marginal pdf's *f (x)* and *f (y)*
- The average mutual information between X and Y is defined as

$$I(X,Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x) f(y/x) \log \frac{f(y/x) f(x)}{f(x) f(y)} \, dxdy$$

- The concept of self information does not exactly carry over to continuous random variables since these would require infinite number of binary digits to exactly represent them and thus making their entropies infinite

Continuous Random Variables …

- We can, however, define differential entropy for continuous random variable as

$$H(x) = \int_{-\infty}^{\infty} f(x) \log f(x) dx$$

- And the average conditional entropy as

$$H(X/Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y) \log f(x,y) dxdy$$

- Average mutual information is then given by
  I(X,Y)= H(X) - H(X/Y) or I(X,Y) = H(Y) - H(Y/X)

Continuous Random Variables …

- If X is discrete and Y is continuous, the density of Y is expressed as

$$f(y)=\sum_{i=1}^{n} f(y/x_i)\,p(x_i)$$

- The mutual information about X = x_i provided by the occurrence of the event Y = y is given by

$$I(x_i;\,Y)=log\frac{f(y/x_i)p(x_i)}{f(y)p(x_i)}=log\frac{f(y/x_i)}{f(y)}$$

- The average mutual information between X and Y is

$$I(X;Y)=\sum_{i=1}^{n}\int_{-\infty}^{\infty} f(y/x_i)\,p(x_i)\log\frac{f(y/x_i)}{f(y)}$$