Appendix A

Source Coding Basics

A.1 Introduction

Communication systems are designed to transmit information from sources to destinations. Sources of information can be analog or discrete. An example of the former case can be a phone call, where generally, the source is an audio signal. The output of this source is analog and, hence, they are called ”analog sources”. On the other hand, we have ”discrete sources” which have discrete values as outputs, as can be the daily stock market index, computer files...

Whether the source is analog or discrete, a digital communication system is designed to transmit information, consequently, the output of a source must be conveniently treated to be transmitted digitally. This function is performed by the source encoder at the transmitter side, whose task is not only to quantize the signal, but also to make an efficient representation of the information in digital form (note we haven't established yet what we mean by efficient).

Obviously, the smaller number of bits is used to represent the signal (during quantization), the less capacity of the channel is used, but the worse fidelity is achieved. In general, we will permit a given level of distortion in quantization and we will try to make the most efficient representation of the levels the signal can take.

A.2 Mathematical models for information sources

Any information source has an output that can be described statistically. Indeed, if we knew the output of a source in advance (deterministic output), there would be no need to transmit it.

Therefore, we need now to model each of the sources described before:

- The output of a discrete source is a sequence of letters belonging to a finite alphabet of L possible letters: $x_i \in \{x_1, x_2 ... x_L\}$.
  Each letter has a probability of occurrence:

$$p_k = P(x = x_k) \tag{A.1}$$

$$\sum_{k_1}^{L} p_k = 1 \tag{A.2}$$

  If every letter satisfies statistical independence among all past and future outputs, we say it is a discrete memoryless source (DMS). On the other hand, if the discrete output shows statistical dependence, we should construct a mathematical model which fits this dependence. For instance, a discrete source is said to be stationary if the joint probabilities of two sequences of length n, say $a_1, a_2, ..., a_n$ and $a_{1+m}, a_{2+m}, a_{n+m}$ are identical for $n \geq 1$ and all possible shifts of m.

- An analog source can be modelled as one whose output presents a waveform $x(t)$, which is a sample function of a stochastic process $X(t)$. We assume that $X(t)$ is stationary with correlation $\phi_{XX}(\tau)$ and power spectral density $\Phi_{XX}(f)$ and bandlimited,$\Phi_{XX}(f) = 0 \ \forall \ |f| \geq B$. The sampling theorem helps us to transmit the samples of the analog signal ($x_n = x(n) = x(nT_s) = x(n\frac{1}{2B})$) for further reconstruction at the receiver side as:

$$X(t) = \sum_{n=-\infty}^{\infty} X(\frac{n}{2B}) Sa(2\pi B(t - \frac{n}{2B})) \tag{A.3}$$

  Where $Sa(x) = \frac{sin(x)}{x}$.

  Our previous analog signal is now a discrete-time signal.

> Finally, we should observe that the result of sampling an analog source is usually a discrete-time continuous-amplitude signal, being necessary to perform quantization to obtain a digital signal.

Where $\{X(\frac{n}{2B})\}$ denote the samples of the process $X(t)$ taken at the Nyquist rate ($f_s = 2B$ samples/s).

A.3 A logarithmic measure of information

Now we know the information provided by a source can be measured, we should find an appropriate way to do it.

Suppose we have two random variables which can take values from a finite alphabet each:

$$x_i \in \{x_1, x_2, ..., x_n\}$$
$$y_i \in \{y_1, y_2, ..., y_n\}$$

If both are statistically independent, the information about X provided by an event in Y is zero. On the other hand, if the occurrence of $Y = y_j$ determines completely the occurrence of $X = x_j$, then, the information the event Y provides about X is the same as the information provided by x_i.

Any measure of information we devise must fulfil the previous two conditions; the following function appears to be suitable:

$$I(x_i; y_j) = \log \frac{P(x_i|y_j)}{P(x_i)}\text{, known as } \textit{mutual information} \qquad \text{(A.4)}$$

If we take $\log_2$ the units of the mutual information are called bits.
Let us check if it really satisfies the two conditions previously exposed:

1. If there is independence between both events,

$$I(x_i; y_j) = \log \frac{P(x_i|y_j)}{P(x_i)} = \log \frac{P(x_i)}{P(x_i)} = 0$$

2. If occurrence of Y totally determines the occurrence of X, then

$$I(x_i; y_j) = \log \frac{P(x_i|y_j)}{P(x_i)} = \log \frac{1}{P(x_i)} = I(x_i), \text{ called } \textit{self-information} \tag{A.5}$$

We observe that a high-probability event conveys less information than other with lower probability.

It is also truth that the information about x_i provided by y_j is identical than the information provided by x_i about the occurrence of y_j.

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

We can also define the *conditional self-information* as:

$$I(x_i|y_j) = \log \frac{1}{P(x_i|y_j)} \tag{A.6}$$

Therefore, the following relation holds true:

$$I(x_i; y_j) = I(x_i) - I(x_i|y_j)$$

A.4 Average mutual information and entropy

We can step further and define the average mutual information between X and Y as:

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) I(x_i; y_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log \frac{P(x_i, y_j)}{P(x_i)P(y_j)} \tag{A.7}$$

It holds that $I(X;Y) \geq 0 \; \forall \; \{X, Y\}$.

Identically, we can obtain the *average self-information* as:

$$H(X) = \sum_{i=1}^{n} \log \frac{1}{P(x_i)} \tag{A.8}$$

Where X represents the alphabet of possible output letters from a source, $H(X)$ represents the average self-information per output and it is called the entropy of the source.

$H(X)$ is always less or equal than $log(n)$, where the equality holds when symbols are equally probable.

Finally, we can define the average conditional self-information (or conditional entropy) as:

$$H(X|Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log \frac{1}{P(x_i|y_j)} \tag{A.9}$$

Again it is true: $I(X;Y) = H(X) - H(X|Y) \geq 0$.

The results previously exposed can be generalized to more than two variables as:

$$H(X_1 X_2 ... X_k) = \sum_{i=1}^{k} H(X_i | X_1 X_2 ... X_{i-1}) \tag{A.10}$$

That satisfies:

$$H(X_1 X_2 ... X_k) \leq \sum_{m=1}^{k} H(X_m)$$

This information measures can easily be extended for continuous random variables by simply applying some little changes.