Information and Capacity

Model of a Communication System

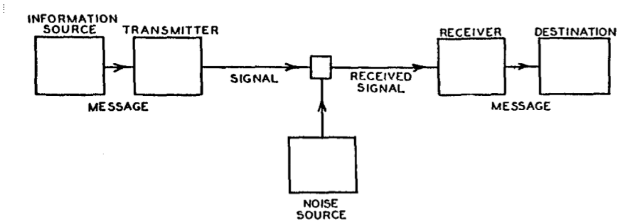

Fig. 1—Schematic diagram of a general communication system.

The diagram above shows a model for a communication system that includes the following[1]:

- information source - generates a sequence of "messages," taken from a limited set of possible values. These values might be a set of voltage levels that taken together convey a perceptible sound or image. The messages might also convey more abstract information called "data" which could represent, for example, the characters in a document or perhaps numbers whose meaning is unknown ("opaque") to the communication system
- transmitter - a device that converts the messages into a time-varying voltage or current (a "signal") that can be carried over the channel
- channel - carries the signal from the transmitter to the receiver, often distorting it and adding random signals called "noise"
- receiver - a device that attempts to recover the messages that were transmitted
- data destination - (sometimes called a "sink") such as a person or computer that makes use of the information

Exercise 1: Give an example of a communication system. If you can, identify the source, transmitter, channel, receiver and destination.

[1]The diagram is from Claude Shannon's fundamental paper, "A Mathematical Theory of Communication," *The Bell System Technical Journal*, Vol. 27, pp. 379–423, 623–656, July, October, 1948.

Review of Random Variables

A *random variable* is one whose value cannot be predicted. Examples in communication systems are the information generated by a source and the noise introduced by the channel.

Although the value of a random variable cannot be predicted, we can define certain properties of these variables called *statistics*.

A statistic called the *expected value* of a random variable X, denoted by $E[X]$ or $\bar{X}$, is the expected average value of X over many "trials" (e.g. many different instances of a noise source or many instance of time).

The *n'th moment* of X is $E[X^n]$ and the *n'th central moment* is $E[(X-\bar{X})^n]$. The first moment is also called the mean and the second central moment the variance (often written σ_X^2).

Random variables can be *discrete* (e.g. bits) or *continuous* (e.g. voltage). The integral of the probability density function between a and b is the probability that the random variable will have a value between these values.

Exercise 2: How would you represent a discrete r.v. in a pdf?

The definite integral of the pdf ($\int_{-\infty}^{\infty} p_X(x)\,dx$) is 1 because the probability that the rv has a finite value is 1.

Stochastic Processes

We are often interested in random variables that are functions of time. These are called *stochastic processes*.

A *stationary* stochastic process is one whose statistics do not vary with time. These are analogous to time-invariant signals and are important for the same reason – we only have to deal with relative delays. There are various types of stationarity depending on which statistics are independent of time (e.g. "'strictly" or "weak-sense" stationarity).

Exercise 3: Is the radio noise generated by the sun a stationary stochastic process? Under what conditions?

Many random processes are *ergodic*. This means the ensemble and time statistics are the same. For example, consider an amplifier. If the statistics of the noise generated by all amplifiers (of the same design) evaluated at one instant are the same as the statistic of the noise from one amplifier evaluated over time then the noise is ergodic.

Exercise 4: Would the amount of data transmitted by cellular subscribers be an ergodic stochastic process?

Multivariate Random Processes

We can define a two-dimensional probability density function (pdf) $p(X, Y)$ which is called the *joint pdf* of the random variables X and Y.

If $p(X, Y) = p(X)p(Y)$ then X and Y are said to be *independent*. This allows us to compute the joint probability using the *marginal* probabilities. We often deal with variables that are independent and identically distributed (i.i.d.).

Exercise 5: Describe the shape of the joint pdf of two zero-mean iid random variables with uniform pdfs. What if they had triangular pdfs extending between ± 1?

The *covariance* of two random variables is defined as:

$$\mathrm{cov}(X, Y) = \mathrm{E}\left[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])\right] .$$

Two random variables are *uncorrelated* if their covariance is zero. This is a weaker condition than independence (two rv can be uncorrelated but not independent).

Functions of Random Variables

The pdf of a sum of two independent random variables is the convolution of the individual pdfs.

Exercise 6: What is the pdf of the sum of two zero-mean iid uniformly-distributed rv's whose pdf has a maximum value of 1?

The *Central Limit Theorem* states (roughly) that the sum of a large number of independent random variables tends to a distribution that has a Gaussian distribution.

The second moment (power) of the sum of two independent random variables is the sum of their powers.

Exercise 7: Prove this.

The *autocorrelation* function of a stationary stochastic process is defined by

$$R_{XX}(\tau) = E[X(t)X(t - \tau)] .$$

The autocovariance is similarly defined (by subtracting the mean).

The autocorrelation function and the power spectrum of a random signal are a Fourier transform pair.

Information Theory

We can model sources as generating one of a limited number of messages. For example, the messages might be letters, words, pixel values, or measurements. Different messages will often have different probabilities. The probability of a particular message is the fraction of messages of that type.

Exercise 8: We observe a source that outputs letters. Out of 10,000 letters 1200 were 'E'. What would be a reasonable estimate of the probability of the letter 'E'?

We define the information that is transmitted by a message that occurs with a probability P as:

$$I = -\log_2(P) \quad \text{bits}$$

For example, a message with a probability of $\frac{1}{2}$ conveys 1 bit of information. While one with a probability of $\frac{1}{4}$ carries 2 bits of information. Thus, less likely messages carry more information.

Entropy

The information rate (also known as the "entropy") of a source in units of bits per message can be computed as the average information generated by the source:

$$H = \sum_i (-\log_2(P_i) \times P_i) \quad \text{bits/message}$$

where P_i is the probability of the i'th message.

Exercise 9: A source generates four different messages. The first three have probabilities 0.125, 0.125, 0.25. What is the probability of the fourth message? How much information is transmitted by each message? What is the entropy of the source? What is the average information rate if 100 messages are generated every second? What if there were four equally-likely messages?

Mutual Information

The mutual information is defined as:

$$\mathrm{I}(X; Y) = \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} p(x, y) \log_2 \left(\frac{p(x, y)}{p(x)\, p(y)} \right) \frac{\text{bits}}{\text{channel use}}$$

where X and Y are the channel input and output random variables,

Exercise 10: What is the mutual information if X and Y are independent? If they are the same?

Capacity

Shannon defined the capacity of a channel as the maximum mutual information between the input and output of a channel:

$$C = \max_X I(X;Y) \ .$$

where the maximization is over all possible distributions of X.

Shannon showed that it is possible to transmit information with an arbitrarily low error rate if the information rate is less than the capacity of the channel. He also showed that it is not possible to achieve an arbitrarily low error rate if the information rate exceeds the channel capacity.

Shannon's proof does not provide a means to design a system that can achieve capacity. It is therefore an upper bound. Shannon's work also hinted that using error-correcting codes with long codewords (to be discussed later) should allow us to achieve arbitrarily-low error rates as long as we limit the information rate to less than the channel capacity.

In practice, attempting to transmit at information rates above capacity results in high error rates.

Examples

One example of a channel is the Binary Symmetric Channel (BSC). This channel transmits discrete bits (0 or 1) with a bit error probability (BER) of p. The capacity of the BSC in units of information bit per "channel use" (transmitted bit) is :

$$C = 1 - (-p \log_2 p - (1-p)\log_2(1-p))$$

which is 0 for $p = 0.5$ (when each transmitted bit is equally likely to be received right or wrong) and 1 when $p = 0$ (the error-free channel) or when $p = 1$ (a perfectly inverting channel).

Exercise 11: What is capacity of a binary channel with a BER of $\frac{1}{8}$ (assuming the same BER for 0's and 1's)?

For a continuous channel corrupted by Additive White Gaussian Noise (AWGN), the capacity can be shown to be:

$$C = B \log_2 \left(1 + \frac{S}{N}\right)$$

where C is the capacity (b/s), B is the bandwidth (Hz) and $\frac{S}{N}$ is the signal to noise (power) ratio.

Exercise 12: What is the channel capacity of a 4 kHz channel with an SNR of 30dB?

Some systems using modern forward error-correcting (FEC) codes such as Low Density Parity Check (LDPC) codes can communicate at very low error rates over AWGN channels with SNRs only a fraction of a dB more than the minimum required by the capacity theorem.

Exercise 13: What do the Nyquist no-ISI criteria and the Shannon Capacity Theorem limit? What channel parameters determine these limits?

The mutual information of a channel, and thus the capacity, depends on many factors. These include include the statistics of additive noise and interference, fading, feedback channels (allowing the transmitter to change how it encodes information), multiple parallel channels (e.g. diversity), as well as modulation and coding (which define X).

Bit and Frame Error Rates

The bit error rate (BER, P_e) is the average fraction of bits that are received incorrectly.

When these bits are grouped into "frames" we are often interested in the average fraction of the frames that contain one or more errors. This is known as the FER (Frame Error Rate). Sometimes frames include additional bits that allow us to detect most, but not all, errors. We usually want the UEP (Undetected Error Probability) to be very small (e.g. one undetected error per many years).

Exercise 14: You receive 1 million frames, each of which contains 100 bits. By comparing the received frames to the transmitted ones you find that 56 frames had errors. Of these, 40 frames had one bit in error, 15 had two bit errors and one had three errors. What was the FER? The BER?

Compression

When data is not random and we can make use of the redundancy to reduce the amount of data that needs to be transmitted. Both lossless and lossy compression are examples of "source coding."

Lossless. Some types of data contains redundancy such as sequences of bits or bytes that occur more often than others. This type of data can be compressed before transmission and then decompressed at the receiver without loss of information. An example of this "lossless" compression is the 'zip' compression used for computer files.

Another definition of information rate is "the minimum data rate, assuming the best possible lossless compression". Lossless compression does not reduce the information rate but it may reduce the bit rate.

Lossy. Data representing speech and video can often be compressed with little degradation because humans cannot perceive certain details of sounds and images. These details can be removed resulting in lower data rates. Examples of these "lossy" compression techniques include "MP3" for compressing audio and MPEG-4 for video.