Fundamentals of Signals

Charan Langton
www.complextoreal.com

When we talk about communications, we are talking about transfer of desired information whether right up close or to far destinations using signals of some sort. Communications via electronic means consists of generation of signals that contain or carry *intelligent* information and its processing at the receiver to extract the intelligent information.

In general a signal is a relationship of a parameter such as amplitude to time. This relationship can be discrete or continuous. The type of signal, analog or discrete, used for communications depends on how far the signal has to travel and the medium it will have to pass through. On the receiving side, this signal is decoded and the intelligent information is recovered. There are various distinct steps a signal goes through from the transmitter (source) to the receiver (sink). These steps although functionally the same, are performed differently for digital and for analog signals. The tools for doing this signal processing whether discrete or analog are governed by communications theory.

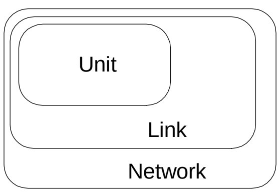


Fig. - 1.1 Study of communications can be conceptualized under unit, link and network level.

In general the field of communications can be conceptualized in three fields of study. There is the physical unit. Here the analysis consists of performance issues of individual units within a communications chain. We may concentrate on the performance of a component such as a filter, an amplifier, an antenna etc.. Then comes a single link between a sender and a receiver. This involves looking at how the signal would go from the handset to the base station and to the intended recipient and vice versa. These are waveform or signal issues. How is the signal distorted as it is going from one unit to the next? As part of a link design, we study modulation, channel effects, power required etc. as shown in Fig. (1.2). The third part is the integration of various communications links to make a network. A cell network has many base stations, many users, how do we design a network so each user is uniquely identified, and does not cause undue interference to other users. Here we study issues of network design, congestion issues and access control.

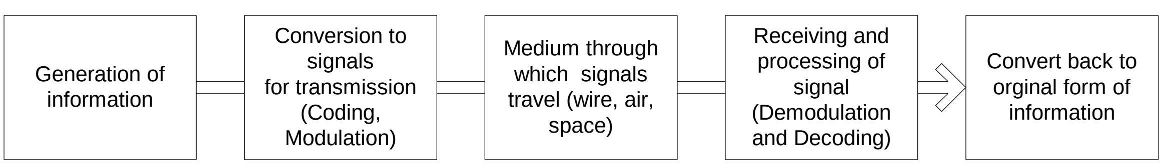


Fig. 1.2 – Generation of signals in a communications link

Signal types

Most signals in nature are analog. Examples of analog signals are sound, noise, light, heat, and electronic communication signals going through air (or space). These signals vary continuously and the processing for these types of analog signals is called Analog Signal Processing (ASP). Examples of naturally occurring discrete signals are Morse code messages, numerical counts such as clock time, the bit streams that constitute digital messages. In communication systems *discrete* signals are often created by *sampling* continuous signals. These discrete signals are sampled versions of the analog signals as shown in (1.3). The amplitude of each sample is same as the original analog signal at the same instant.

Difference between discrete and digital

A *digital signal* is a further refinement of this process. A discrete signal is any signal that has values only at specific time interval. It is not defined between those times. A digital signal on the other hand is one that only takes on a specific set of values. For a two level binary signal, we can take a sample signal which is a discrete signal and then set it to just two levels, -1 or +1. The discrete signal then becomes a digital signal. Digital however does not mean two levels. A digital signal can take on any number of values, usually in power of two. The process of gathering the amplitudes in specific levels is called quantization. A binary signal is a special case of digital signals and digital signal is a special case of a discrete signal.

Discrete signals can be of very small duration so they are essentially like an impulse, or they can hold their value for a certain period of time.

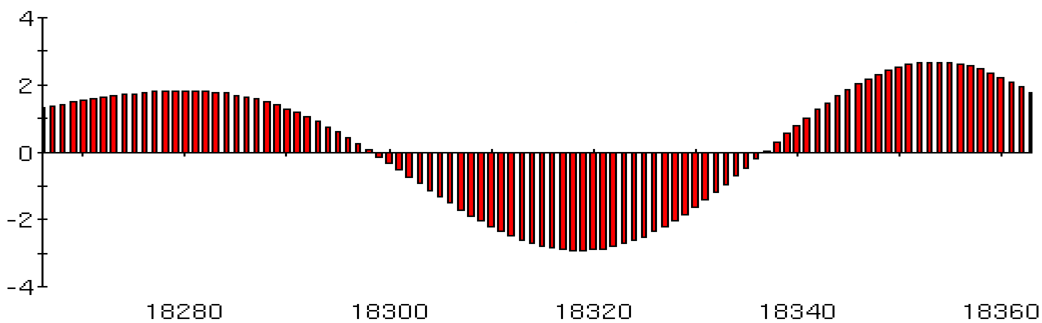


Fig. 1.3 – A discrete signal with varying amplitude at each time instant (sample number).

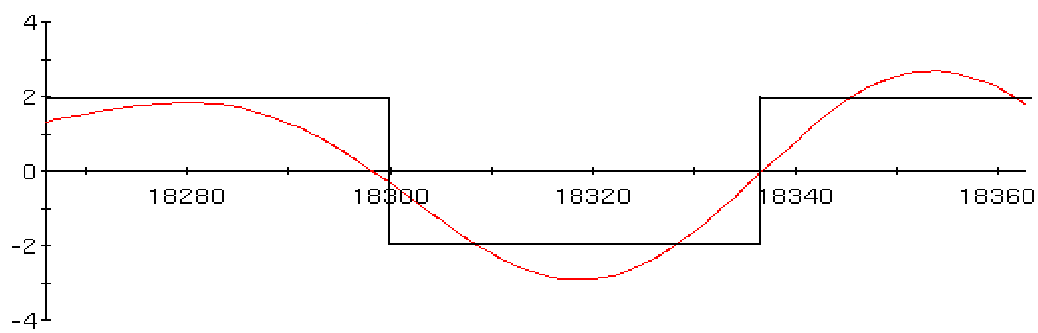


Fig. 1.4 – An analog signal with an equivalent digital (binary) signal (sample number).

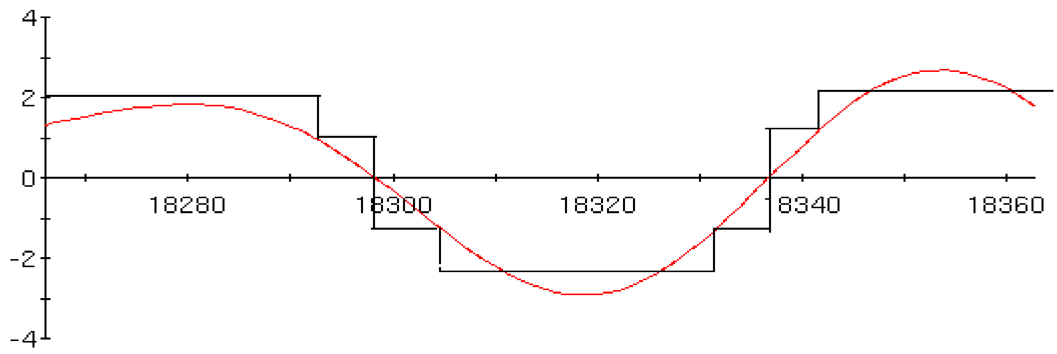


Fig. 1.5 – An analog signal with an equivalent digital (4-level) signal (sample number).

There is another type of signal which is called a pulse train. Here the signal is train of impulses that last a long or a short time. The phase and amplitude of the impulse determines how it is decoded. These pulse-like signals are used in Ultra Wide Band (UWB) systems and for Pulse Position Modulation (PPM). EKG is a good example of this type of signal occurring in nature.

We hear a lot about digital signals but most communication systems would not be possible without analog communications. All wireless signals are analog. A signal may start out analog, converted to digital for modulation, and converted back to analog for radio frequency (rf) transmission and then converted back to digital. Both forms are needed to complete a link. A communications chain is a combination of both of these types of signals. If you work with just the handsets (units) then you will most likely work only in digital domain. If you work on link design or in very high powered equipment such as satellites, then analog issues become important.

Analog signal

An analog signal is one that is defined for all time. You can have a time-limited signal, such as one that lasts only one second but within that second it is defined for all time t. A sine wave as written by this equation is defined for all time, t.

$$f(t) = \sin(2\pi t) \qquad (1.1)$$

A discrete time signal which can also be time-limited, is present only at specific, usually regular intervals. Mathematically such signals are written as function of an integer index n, where n is the n_{th} time tick from some reference time. If we define T as the interval between ticks, the discrete version of the analog sine wave is written as (1.2) and is referred to as the sampling process. The processing of quantization of the discrete signal is called the A/D conversion.

$$f(n) = \sin(2\pi nT), \quad n = 0, \pm 1, \pm 2, ... \qquad (1.2)$$

In definition (1.2), the samples are generated every 1/T seconds. The sampling frequency is the speed at which the samples are taken and is inverse of the sampling time.

$$Sampling\ Frequency = f_s = \frac{1}{T_s} \qquad (1.3)$$

If a signal is sampled at 10 Hz, then the sampling period is T = 1/10 = 0.1 sec. The individual sampled amplitude for a sampling speed of 10 Hz would be given by

$$a_n = f(n) = \sin(2\pi nT) = \sin\frac{\pi n}{5} \qquad (1.4)$$

Information signal

We commonly refer to transmission of voice, music, video and data as **information** signals. These are the intended signals, what we want to communicate and hence are signals containing information. The process of sending these signals through a communications chain tacks on a lot of extra information, such as in a phone call, your phone number, time-date stamps etc. and much more that is invisible to the sender and which facilitates the transfer of this information. These data signals are over and above the information signal and are considered overhead. The information signal can be analog or digital. Voice and music are considered analog signals, video can be either digital or analog. Stock market and financial data is an example of an information signal that is digital. In most case, the analog signals are converted using analog to digital (A/D) converter by sampling it into a discrete signal and quantization them prior to transmission. The discrete signals can have any number of levels but binary or two level signals and then creating are most common in communication. In (1.6-1.8), we see three different digital signals. The levels of 1 and -1 make the first signal a binary signal, it has just two levels. The second signal takes on four discrete values and the third takes on 8.

Information signals are referred to as **baseband signals** because they are at low frequencies, often less than 50 kHz. A baseband signal can be all-information or it can contain redundant bits making it a coded signal but it is still at baseband and still has fairly low frequency contents.

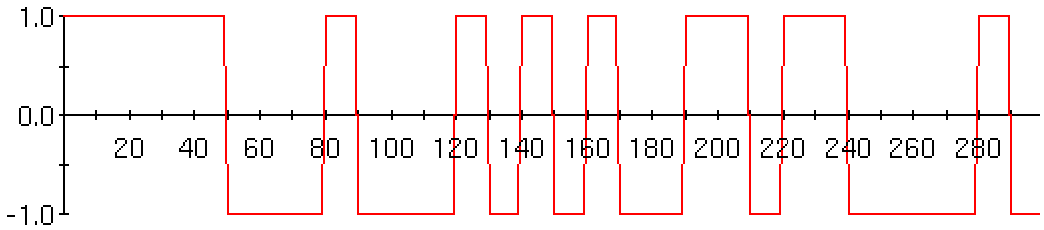


Fig. 1.6 – A binary signal

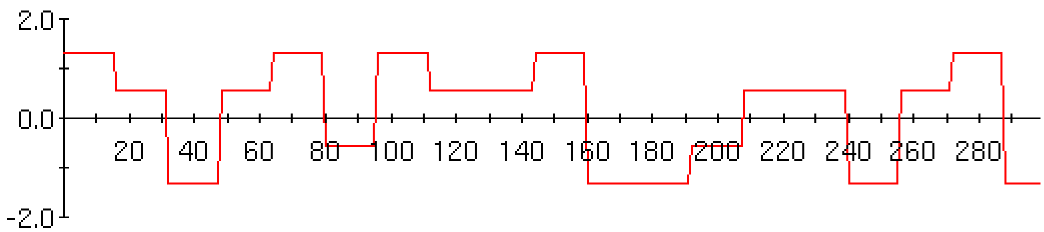


Fig. 1.7 – A 3-level signal

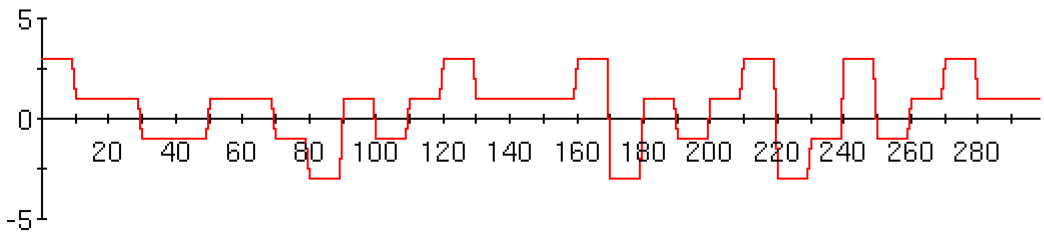


Fig. 1.8 – A 4-level signal

Carrier Signals

Carriers carry the information signal over long distances. But why can't we just transmit the information signals themselves, why use a carrier? Remember that a signal as it travels degrades in power. A line of sight signal attenuates by square of the distance. In order to deliver a discernible signal at the receiver, we need to send a fairly high power signal. This is done by transmission antennas which concentrate the power in a particular direction. The antenna gain is a function of the square of the frequency; the higher the frequency, the higher the gain. The other key parameter is the antenna size; the larger the size, the larger the gain. So to transmit a signal of 1 KHz vs. 1 GHz, would require an antenna that would have to be 10^6 times larger. So even doubling, tripling the area won't make up for the advantage offered by the use of the higher frequency. The use of a high frequency carrier allows us to use smaller antennas. In satellites a Ku-band signal requires only a 0.3 meters diameter dish vs. C-band which requires a dish of app. 2 meters. So that's one reason we use carriers. The other reason is the some media are not friendly to all frequencies. The optical fiber is one obvious example of that.

To transmit baseband information over a carrier requires a process called **modulation**. Modulation is described as the process of mapping the information signal on to the carrier signal. These higher frequency signals that facilitate transfer of information over a variety of media are called **carriers**. The frequency of the carrier is usually much higher than the information signal. The choice of a carrier is function of the medium it must pass through. For wired communications, the carrier may be in KHz range and for wireless and satellites they are in MHz and GHz frequencies. In United States carrier frequencies one can use are prescribed by law for efficient use of the spectrum.

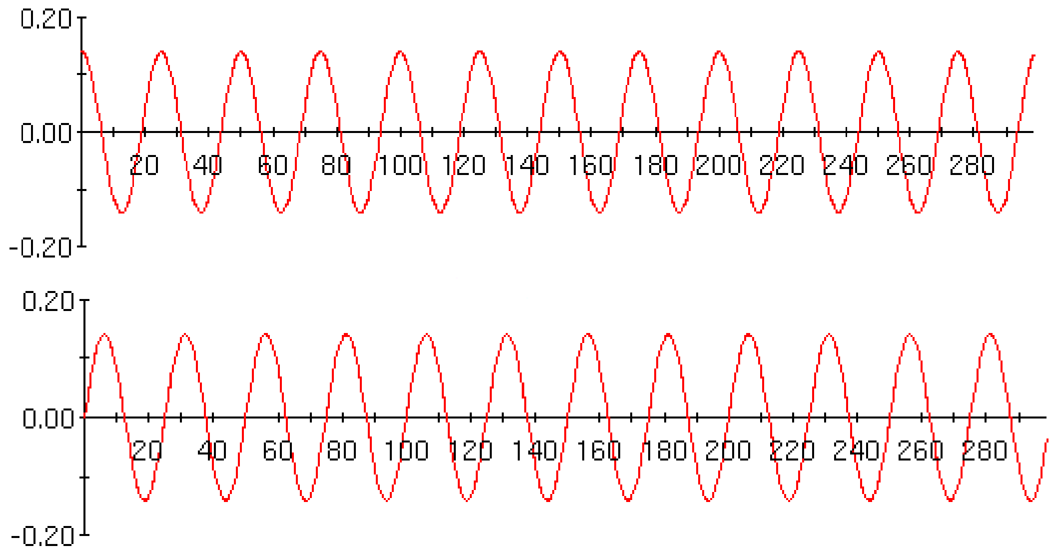


Figure 1.9 – A carrier is a pure sinusoid of a particular frequency.

A carrier is a pure sinusoid of a particular frequency and phase. Carrier signals are produced by voltage controlled oscillators (VCO). In majority of the application, when we talk about carriers, they are analog signals, assumed to be continuous, constant in amplitude and phase and lasting an infinite time. Phase of a carrier and even its frequency can drift or change either slowly or abruptly, and so in reality they are not perfect. The imperfections in the carriers cause problems in removing the information signal at the receiver and so methods have been devised to both track and adjust the carrier signal.

The carriers are carriers in that they *carry* the information signal. This process of "carrying the information" is called modulation. The carrier by itself has no information because it changes in a fixed and predicable manner and information implies change of some sort. So in a modulated signal, we can swap one carrier for an another one (assuming no regulatory constraints) and it would make no difference. However, there is one requirement a carrier must meet: its frequency must be at least two times the highest frequency in the information signal. This is a fundamental requirement that comes from sampling theory.

Modulated Signals

A modulated signal is a carrier that has been loaded with an information signal. To transfer information, it is the modulated signal that travels from place A to B. The information signal in its original shape and form is essentially left behind at the source. A modulated signal can have a well defined envelope as shown here in Fig. 1.10a and 1.10b or it can be wild looking as shown in Fig. 1.10c.

The process of modulation means taking either an analog or a digital signal and turning it into an analog signal. The difference between a digital modulation and analog modulation is the nature of the signal that is modulating the carrier. The carrier is always analog. In digital modulations, we can see the transitions, Fig. 1.10a and 1.10b, whereas in analog modulated signals the transitions are not obvious, Fig. 1.10c.

Modulation is analogous to another process called D to A, or digital to Analog Conversion (D/A). The D/A conversion is typically done at baseband and does not require any change in the frequency of the signal, whereas modulation necessarily implies a frequency translation.

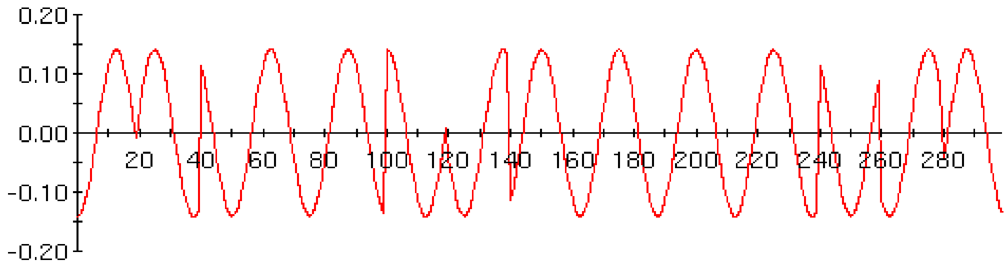


Fig. 1.10a – A modulated carrier signal (digital input)

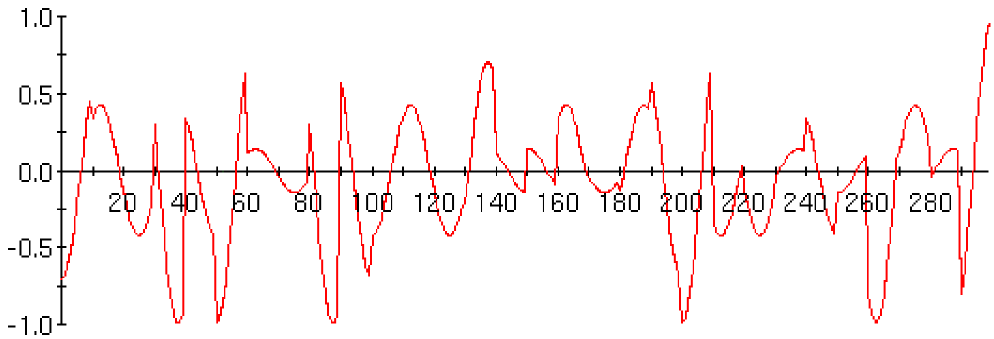


Fig. 1.10b – Another modulated carrier signal (digital input)

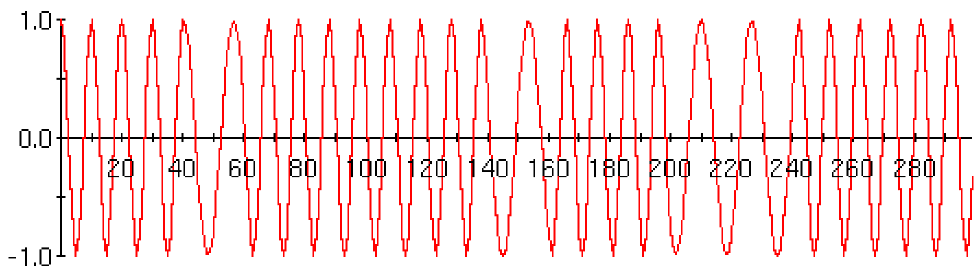


Fig. 1.10c – Yet another modulated carrier signal (analog input)

Bandwidth

Bandwidth can be imagined as a **frequency width**, sort of the fatness of the signal. The bandwidth of a carrier signal is zero. That is because a carrier is composed of a single frequency. A carrier signal is devoid whereas information signals are fat with information. The more information in a signal, the larger the bandwidth of the information signal. To convey information, an information signal needs to contain many different frequencies and it is this span of their frequency content that is called its **bandwidth**. The human voice, for example, spans in frequency from 30 Hz to 10,000 Hz. The range of frequencies in human voice gives it its unique signature. The voice has a bandwidth of approximately 10,000 Hz. Not all of us can produce the same range of the frequencies or amplitudes, so although the spectrum of our voice generally falls within that range, our personal bandwidth will vary within this range. No two people will have the same voice spectrum for the same sounds. It is the same with information signals, they are all unique, although they may occupy the same bandwidth.

If a voice signal is *modulated* on to a carrier, what is the bandwidth of the modulated signal? It is still the same. The modulated signal takes on the bandwidth of the information signal it is carrying like this guy on the motorcycle. He is the modulated signal and his bandwidth just went from near zero, without the load, to the size of the mattress which is his "information" signal.

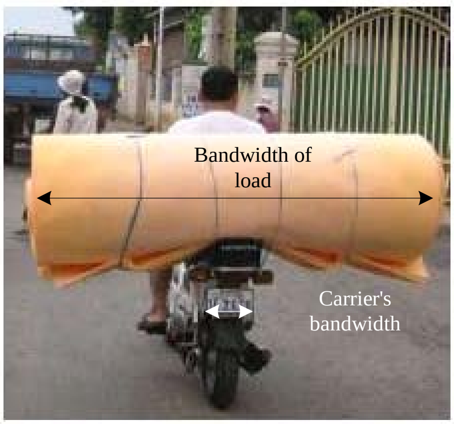


Fig. 1.11 – Bandwidth is a measure of the frequency content of a signal.

Properties of Signals

Periodicity

Carriers have strict periodicity whereas information signals do not have this property. Communications theory and many of the tools used to analyze signals do however rely largely on the concept of periodicity. Conversion of signals from time domain to frequency domain depends on this property and many other analytical assumptions we make about signals also require periodicity. Purely periodic math applies to the carriers whereas math used to describe the information and modulated signals uses stochastic and information theory of random signals.

First we will look at properties of periodic signals and then later at random signals. Mathematically a discrete periodic signal is one that has the following property.

$$f(t) = f(t \pm T) \qquad (1.5)$$

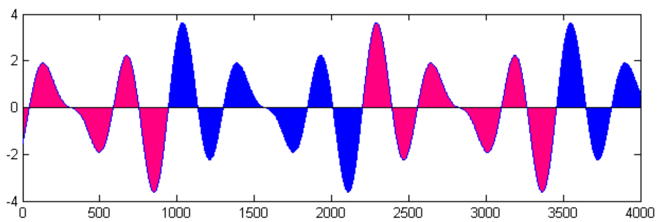


Fig. 1.12 – Carriers are periodic, information signals are not.

This is a sampled discrete signal despite the fact that it looks continuous. (The samples are too close to see.) This is a periodic signal because repeats its pattern with a period T. The pattern can be arbitrary. The value of the signal at any one time is called a sample. The concept of periodicity follows superposition principal. If we add many periodic signals, with different frequencies and phases, the resulting signal is still periodic.

The *mean* of a discrete periodic signal x is defined as the **average value** of its samples:

$$\mu_x = \frac{1}{N}\sum_{n=0}^{N-1} x_n \qquad (1.6)$$

Power and Energy of Signals

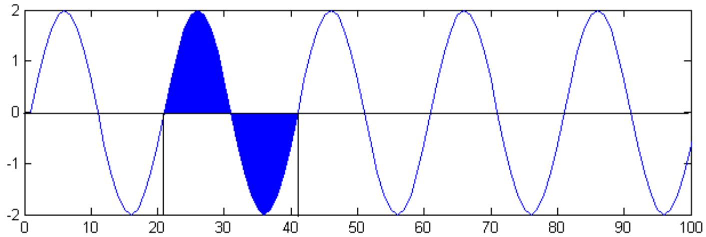


Fig. 1.13 – The area under one period of this signal is zero.

It seems that we ought to be able to say something about the area under a periodic signal. But we see that a zero-mean periodic signal is symmetrical about the x-axis, as is a sine wave; and hence the signal has zero area for an integer number of periods. The negative parts cancel the positive. So the area property does not tell us much. If we square it, we get something meaningful, something we can measure and compare between signals.

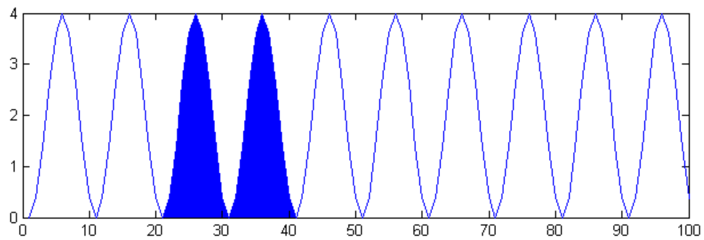


Fig. 1.14 – The area under the squared period is non-zero and indicates the power of the signal.

If we square the sine wave of Fig. 1.13, the area under the squared signal (1.14) for one period is 40 units (2*4*10/2) and the total area under this signal is 40 times the number of periods we wish to count. This is called the **energy of the signal** defined by the area under the signal when squared for a specific time length.

$$E_x = \sum_{n=0}^{N-1} |x_n|^2 \tag{1.7}$$

However, since this sum depends on the time length of the signal, the energy of the signal can become very large and as such it was decided that this is not a useful metric. A better measure is **power of the signal**. The average power of a signal is defined as the area under the squared signal divided by the number of periods over which the area is measured.

$$P_x = \frac{E_x}{N} = \frac{1}{N} \sum_{n=0}^{N-1} |x_n|^2 \tag{1.8}$$

Hence the **average signal power** is defined as the total signal energy divided by the signal time since N is really a measure of time. The power of a signal is a bounded number and is a more useful quantity than the signal energy.

The **root mean square** (RMS) level of a signal is the square root of its average power. Since power is a function of the square of the amplitude, the **RMS value is a measure of the amplitude** (voltage) and not power. We can compare the RMS amplitude with the peak amplitude to get an idea of how much the signal varies. For some applications such as Orthogonal frequency Division signals (OFDM) the peak to average measures are an important metric.

$$x_{RMS} = \sqrt{\frac{x_1^2 + x_2^2 + x_3^2 \ldots}{n}} = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1}|x_n|^2} \qquad (1.9)$$

The **variance** of the signal is defined as the power of the signal with its mean removed. Variance of a signal is defined as (1.10).

$$\sigma_x^2 = \frac{1}{N}\sum_{n=0}^{N-1}|x_n - \mu_x|^2 \qquad (1.10)$$

For zero-mean signals, the variance of the signal is equal to its power. Equation (1.10) becomes equation (1.11) if mean is set to zero. For non-zero mean, variance equals the power minus the mean of the signal. P_x is also often called the DC power.

$$\sigma_x^2 = P_x - \mu_x \qquad (1.11)$$

We can also talk about **instantaneous power**, which is the instantaneous amplitude squared at any moment in time. Since signal power is changing with amplitude, we also have the concept of peak power. Peak power is the peak amplitude squared.

Now we define a quantity called **bit energy** limited over one bit period (this can be a specific number of carrier cycles, usually more than one.)

$$E_b = \frac{Avg\left[x^2(t)\right]}{R_b} \qquad (1.12)$$

For a digital signal, the energy of a bit is equal to the square of the amplitude divided by the bit rate. An another way to write this is C/R_b, with C equal to signal power. If you think about it for a moment, it makes intuitive sense. The numerator is the power of the signal at one instant. We take the power and we divide it by bit rate and what we get is power in a bit, which we call E_b or energy per bit. It is a way or normalizing the energy of the signal. This is a very useful parameter and is used to compare different communication designs.

Random Signals

Information signals are considered random in nature. When you talk, you produce sounds that are essentially random. You don't repeat words in a predictable manner. So signals containing information are not periodic by definition. When compared to the definition of power of a periodic signal, we define the power for random processes slightly differently as

$$P_x = E\left[x^2(t)\right] \qquad (1.13)$$

which is simply the expected value of the mean squared value of the instantaneous amplitude squared. For most communications signals with zero-mean signals, power is not a function of time and is simply the second moment or the variance of the signal x(t).

$$P_x = E\left[x^2(t)\right] = Variance \qquad (1.14)$$

This is intuitively obvious; a signal with a large variance has large power. What is interesting about this relationship is that the variance is also equal to the value of auto-correlation of the signal at zero shift. A zero shift means that we take a signal and multiply it by itself, sample by sample. Of course, that is the same as squaring each instantaneous amplitude! We can write this relationship as

$$P_x = R_x(0) = Variance \qquad (1.15)$$

For these simple relationships to hold for communications signals, certain conditions have to be true about them. One such property is called **Stationarity**. This means that properties of the signal stay put. This is generally true of run of the mill communications signals involving voice data etc. but not for all. If the data is non-random, then most equipment will first randomize the data before transmitting. This averages out the power peaks and is more benign for the transmitters and receivers.

For a signal x(t), if the expected value of the signal amplitude, E{x(t)}, does not change over time, then the signal is called a **stationary** signal. If the mean and the co-variance between signal samples. a certain distance apart are constant, then this type of signal which may not be strictly stationary by the above definition is called a **wide-sense stationary** (WSS) signal.

An **ensemble** is one piece of a particular signal. A collection of ensembles constitutes the whole signal and no parts of one ensemble are shared with another ensemble. (Same as saying, a lot of ensembles make a wardrobe.) If we take the average value of an ensemble, and it turns out to be the same as the average of the whole signal, then this signal would be called **Ergodic**. Not only is the expected value of the signal constant over time but is constant across all ensembles.

An example of a signal not meeting these conditions is the average height of people. Height is not a stationary signal because over time it changes, since humans have been getting taller with the passage of time. It is also not ergodic because the average of one of its ensembles (such as average height in China as compared to average height in England) is not the same. Many signals such as average rain rate are stationary but not ergodic. So ergodic is a much more restrictive condition.

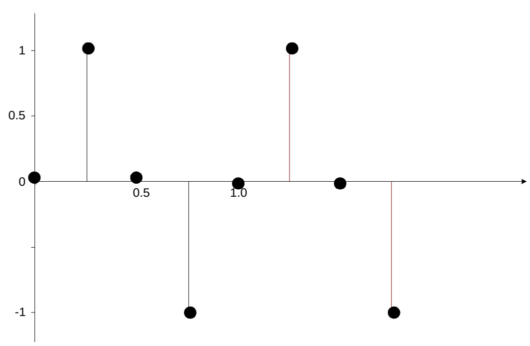
Most signals we deal with in communications are presumed to be stationary and ergodic. They often do not meet these definitions strictly but these assumptions work well enough and give us mathematical tools to analyze and design communication systems. Of course, there are cases that just cannot be assumed as such. Example of non-stationary signals are: Doppler signal (coming from a moving source), variations of temperature, and accelerating, fading and transient signals.

Our Fourier transform-based signal processing techniques are strictly valid only for signals that are both stationery and ergodic but we can use these techniques for signals that don't meet these criterion anyway, as long we understand what errors these assumptions cause.

Sampling

In signal processing, the most challenging part is to receive a signal in an analog form some a source and then figuring out what was actually sent. In signal processing terminology, we want to resolve the received signal. Why should that be a problem? An analog signal is received. The receiver goes through a sampling process and generates some samples. Can't we just then connect these points and know what was sent? Yes, maybe.

Here are sampled values of an unknown signal. This is all we have, just the sampled values. What is the frequency of the signal that was sent? The sampling time is 0.25 seconds, i.e. each sample is 0.25 seconds apart. We connect these and get the waveform on the right of frequency 1 Hz.



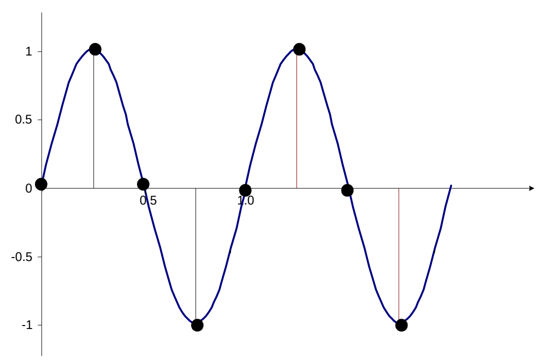


Fig. 1.15 Guessing what was sent based on sampled values.

(a) Received samples, (b) Our assumption of the signal that was sent.

But wait, but what about the following. These two functions also go through the same points. Could not they have been sent? In fact given these sampled points, infinite number of signals can pass through the points of (1.7).

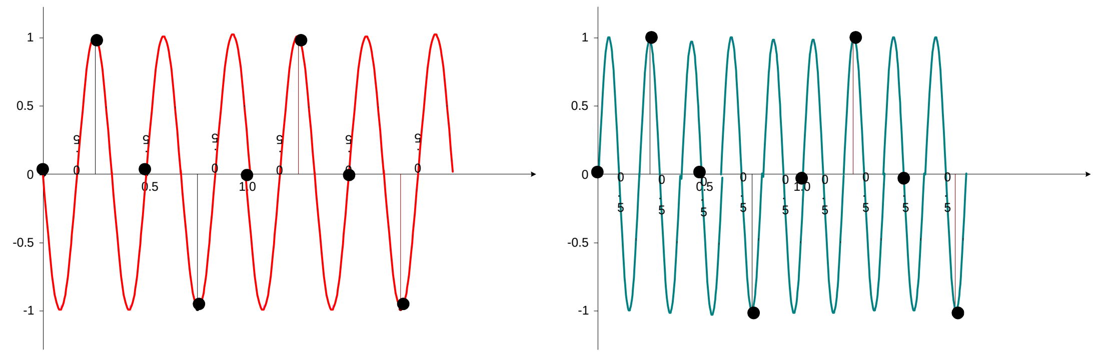


Fig. 1.16 Guessing what was sent based on sampled values.

(a) 2 Hz signal also fits data, (b) 4 Hz signal does too.

We can pick any of these at random. But how do we unambiguously decide what was sent? We see that it could have been any number of signals with frequency higher than 1 Hz. Given these samples there is no way we can tell which of these many signals was sent. The only thing we can say definitely is that the signal is of frequency 1 Hz or larger. So with this sampling rate, the lowest frequency we can correctly extract from these samples is 1 Hz.

Working backwards we see that if we have samples separated by time T, then the largest frequency which we can ambiguously resolve is 1/2T.

$$f_{largest} = \frac{1}{2T_s} = \frac{f_s}{2} \quad (1.16)$$

This is the largest frequency in the signal that can be resolved using a certain sampling rate is called the Nyquist frequency and the sampling frequency needs to be twice as large as this frequency or we can not find the signal.

Let's show this by example.

$$f(t) = \sin(2\pi(4)t) + \cos(2\pi(6)t) \qquad (1.17)$$

This simple signal contains two sinusoids, of frequency 4 and 6 Hz. In Fig. 1.8 we show how the signal looks as it is sampled at various speeds. Clearly as the speed decreases, i.e. the number of samples obtained in one second, the reconstructed signal begins to looks bad. On the right side for each sampled signal is its **Fourier Transform**, which is kind of like a frequency detector. For each case we see that the embedded signals are correctly detected until we get to a sampling frequency below 12 Hz, now the larger components (6 Hz) is not detected. If we go below 8 Hz, even the 4 Hz component disappears. There is a component at 2 Hz, but we know this is wrong and is a result of numerical artifacts. In sampling a real signal of course, we would not know that this is an error and if we do not sample a signal fast enough, we are liable to claim that the sampled signal is a signal of frequency 2 Hz.

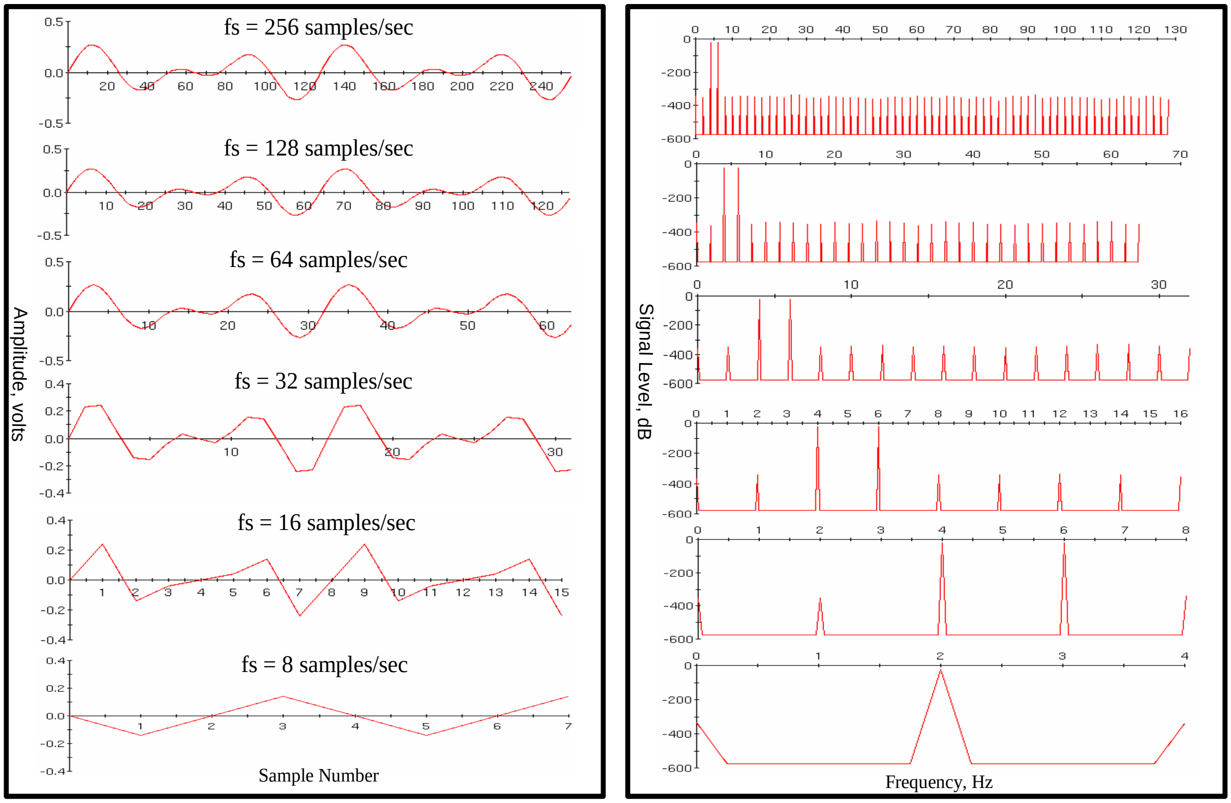


Fig. 1.17 – Sampling speed has to be fast enough (> 2 times the highest frequency embedded) for the processing to detect it and reconstruct the information signal.

Noisy Signals, Random Signals

Being able to decode the original signal successfully depends a lot on the understanding of noise which can enter the link at many points. Noise comes in many flavors and its knowledge is fundamental in study of communications theory. Noise is a non-deterministic random process. (Non-deterministic means, we cannot predict its value ahead of time, just as we can for a carrier signal.) So although we can look at noise signals in time domain, they are often described in frequency domain instead. Information signals similarly are random and can only be analyzed using information theory.

Difference between information theory ,Communications theory and Signal processing

Information theory is a field of science first developed by Clyde Shannon to determine the limits of information transfer. From information theory we learn what is the theoretical capacity of a channel and the envelope of performance that we can achieve. It drives the development of codes and efficient communications but says nothing about how this may be done, similar in idea to the speed of light as a fundamental limit of motion. The important sub-fields of information theory are source coding, channel coding, algorithmic complexity theory, algorithmic information theory, and information-theoretic security.

Communications Theory is all about how to make the information transfer happen from A to B within the constraints of Information theory. It is concerned with choice of media, carriers, maximum number of bits that can be transferred in a given bandwidth, mapping of information to carriers, channel degradation mitigation and link performance. It uses transform theory as its mathematical basis. To be able to fully understand communications, one needs to know Fourier, Hilbert, LaPlace and Z transforms as well as convolution and filtering.

Signal processing is a largely mathematical science of mapping one domain to another, analog to digital, frequency to time. To design digital hardware one needs to understand in detail how operation such as A/D, D/A conversions, and digital math are done. Signal processing provides the tools that implement communications designs.

Signal properties

We need to know the following four properties of random signal distributions; Mean, Variance and Probability Density Function (PDF), Cumulative Density Function (CDF).

Mean

Given a random signal x(t), its mean is given by

$$\overline{x} = E[x] = \int_{-\infty}^{\infty} x \ p(x)\,dx \qquad (1.18)$$

Where $E[x]$ = expected value of x and $p(x)$ = probability density function of x

Mean is fairly easy to understand. For a voltage vs. time signal, to find the mean we take each amplitude and then divide by the number of samples.

After the mean, comes the measure of how much the voltage varies over time. This is called variance, and it is directly related to the power of the signal.

Variance

$$\sigma^2 = E[(x-\overline{x})^2] = \overline{\left(x^2\right)} - \overline{x}^2 \qquad (1.19)$$

Probability Density Function (PDF)

This concept seem to cause confusion because of several reasons; one, the word *density* in it, second it is also known as the Probability Distribution Function (PDF) and third is its similarity to another important idea, the Power Spectral Density (PSD).

PDF is the statistical description of a random signal. It has nothing to do with the power in the signal nor does the word density have any obvious meaning. Let's take the random signal (1.20). It has about 8000 point of data, one voltage value for each. The voltage varies from -18.79 to +17.88v. If we accumulate all the amplitude values and then gather them in bins by quantizing the independent variable, amplitude and plot it, we get a histogram.

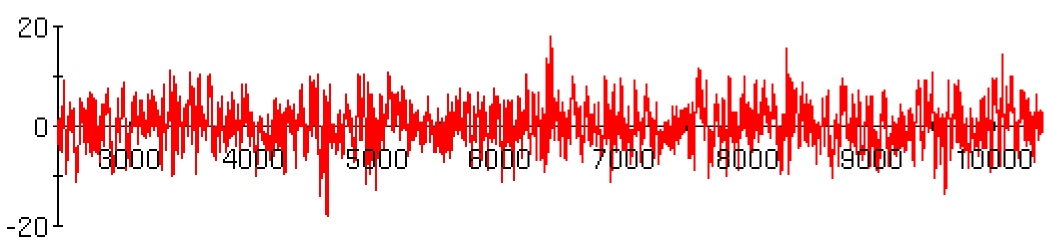


Fig. 1.18 – A data signal that has picked up noise.

In (1.21), All 8000 values of the amplitude have been grouped in 50 bins. Each bin size is equal to the range of the variable, which is 18.79 +17.88 = 36.07 divided by the number of bins or 36.07/50 = 0.7214 v. Each bin contains a certain number of samples that fall in its range. Bin in #23 shown contains 584 samples with amplitudes that fall between voltage levels of -2.1978 v and – 1.4764 v. As amplitudes get large or small, the number of samples gets small. This histogram appears to follow a normal distribution.

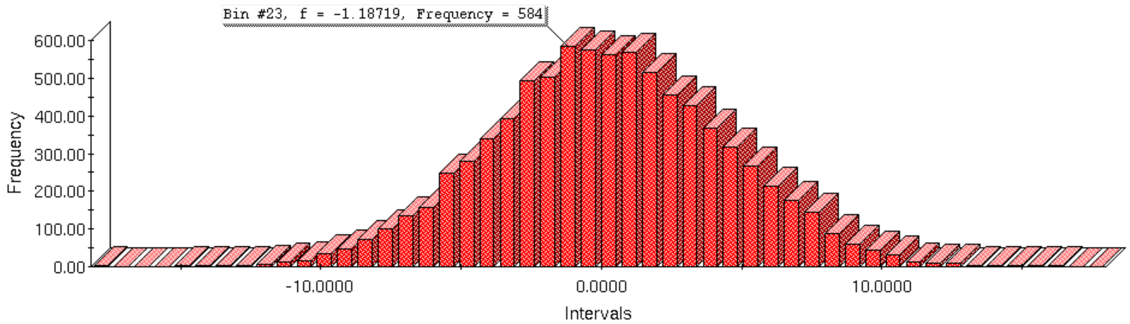


Fig. 1.19 – Histogram developed from the signal amplitude values.

The histogram can be normalized, so the area under it is unity, making it a Probability Density Function. This is done by dividing each grouping of samples by the total number of samples and the quantization interval. For the bin #23, we divide its count by total samples (584/8000) and get 0.07042. This states that the probability that a given amplitude is between of -2.1978 v and – 1.4764 v is 0.07042.

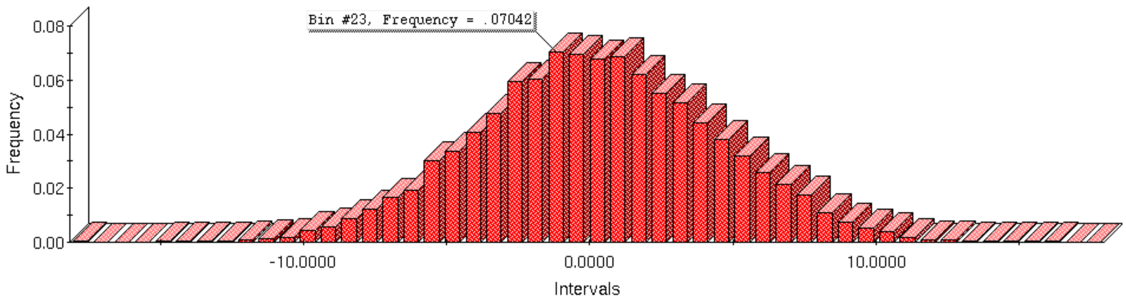


Fig. 1.20 – Normalized histogram approaches a Probability Density Function

In limit, as the width of the bins becomes small, the distribution becomes the continuous Probability Density Distribution or Probability Distribution Function.

From Wikipedia, "A probability density function can be seen as a "smoothed out" version of a histogram: if one empirically samples enough values of a continuous random variable, producing a histogram depicting relative frequencies of output ranges, then this histogram will resemble the random variable's probability density, assuming that the output ranges are sufficiently narrow."

Probability Histogram $p(x) = \Delta x_i \xrightarrow[N \to \infty]{\lim} 0 \frac{N_i}{N \Delta x_i}$ (1.20)

Where

Δx_i = the quantization of the interval of x

N = Total number of samples

N_i = Number of samples that fall in Δx_i

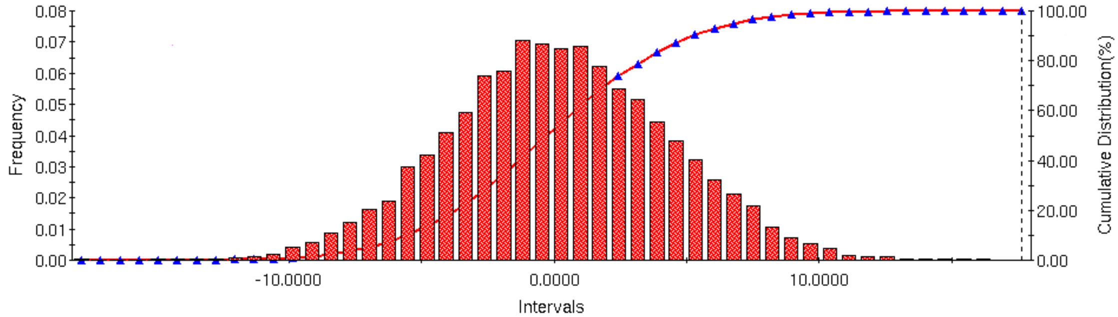


Fig. 1.21 –Probability Density Function and Cumulative Probability Function

To find the probability that a value falls within a specific range x_1 to x_2, we integrate over the range specified. Symbolically:

$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} p(x)\,dx \tag{1.21}$$

Normalizing this, we get

$$P(-\infty < x < \infty) = \int_{-\infty}^{\infty} p(x)\,dx = 1 \tag{1.22}$$

The PDF is always positive. The Cumulative Probability Distribution is the summation of area under the PDF plotted as a function of the independent variable. It is always a positive function with range from 0 to 1.0 as shown in (1.23)

Power Spectral Density (PSD)

A very similar sounding but quite different concept is Power Spectral Density (PSD).

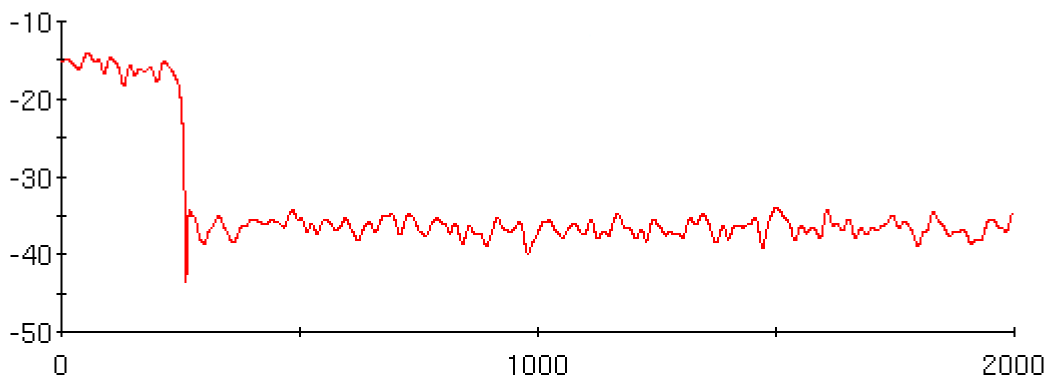


Fig. 1.22 –Power Spectral Density (PSD) of signal in (1.20)

Power Spectral Density (PSD) of a signal is same as its power spectrum. A spectrum is a plot of the power distribution over the frequencies in the signal. It is a frequency domain concept, where the PDF is a time domain concept. You create the PDF by looking at the signal in time domain, its amplitude values and their range. Power Spectral density (PSD) on the other hand is relationship of power with frequency. While variance is the total power in a zero-mean signal, PSD gives the distribution of this power over a range frequencies (power/Hz) contained in the signal. It is a positive real function of frequency. The power is always positive, whereas amplitude can be negative. Power spectral density is commonly expressed in watts per hertz (W/Hz) or dBm/Hz.

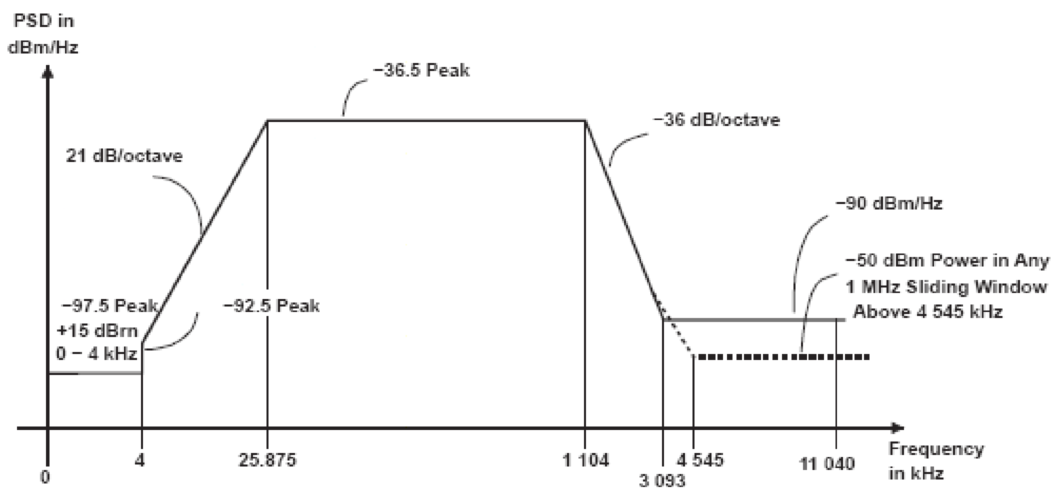


Fig. 1.23 – Specification of Power Spectral Density (PSD) for a specific signal

The two most common random process distributions needed in communications signal analysis and design are Uniform distribution and normal distribution.

Uniform distribution

Take a sinusoid signal defined at specific samples as shown below.

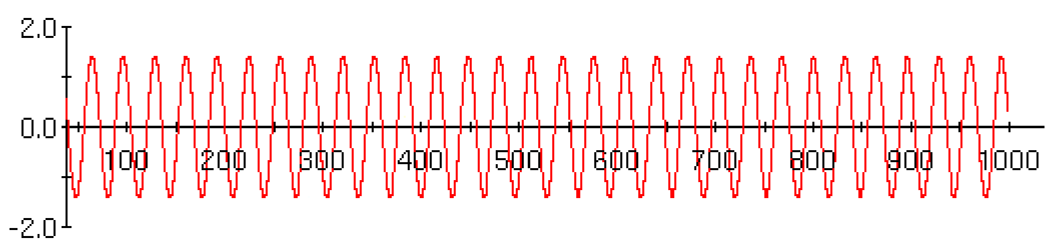


Fig. 1.24 – A sinusoid.

The addition of a -10dBc uniformly distributed noise transforms this signal into (1.27).

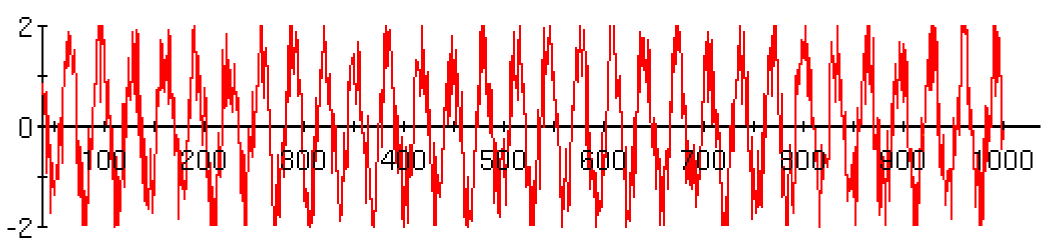
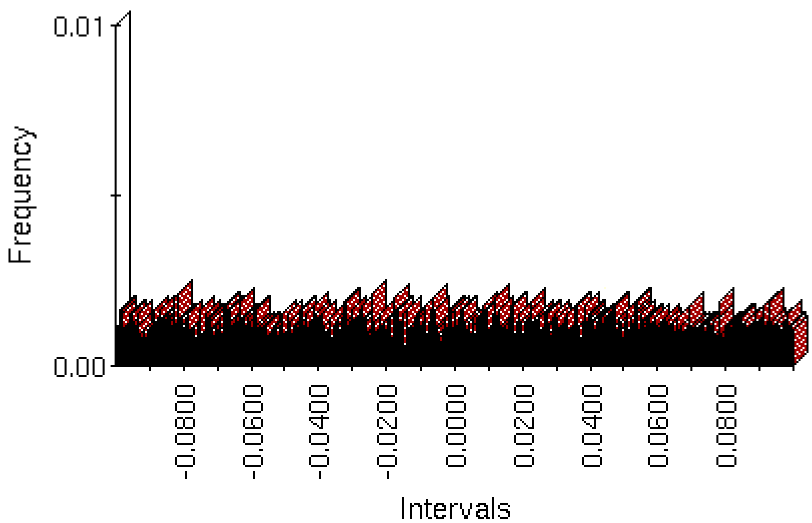


Fig. 1.25 – Transmitted signal which has picked up uniformly distributed noise distributed between -0.1 and +0.1 v.

The noise level varies from +.1 to -.1 in amplitude. If quantize this noise into say 10 levels, at each sample a 11-sided dice is thrown and one of these values is picked [-1 -.8 -.6 -.4 -.2 0 .2 .4 .6 .8 1] and added to the signal as noise. The probability distribution of this noise looks as shown in (1.28)



1.26 – Uniformly distributed noise distributed between -0.1 and +0.1

Uniform distribution noise level is not a function of the frequency. It is uniformly random. This distribution is most commonly assumed for narrowband signals.

Properties of uniformly distributed noise

If a and b are the limits within which a uniformly distributed noise acts, then its properties are given by expressions;

The Mean and variance

$$Mean = \frac{a+b}{2} \qquad (1.23)$$

$$Variance = \frac{(b-a)^2}{12} \qquad (1.24)$$

The PDF and the CDF of the uniform distribution is given by

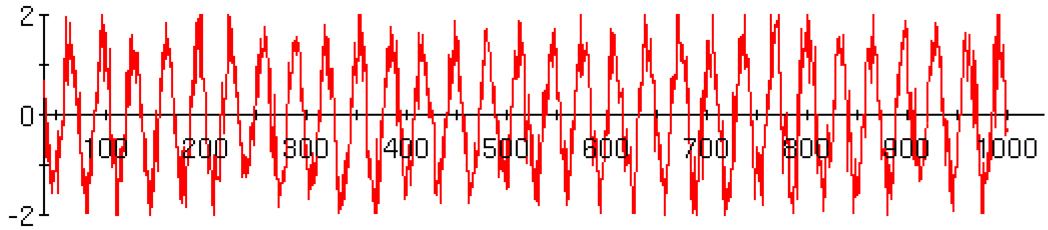
$$f(x) = \begin{matrix} \frac{1}{b-a} & a \le x \le b \\ 0 & x \ge b \end{matrix} \qquad (1.25)$$

CDF

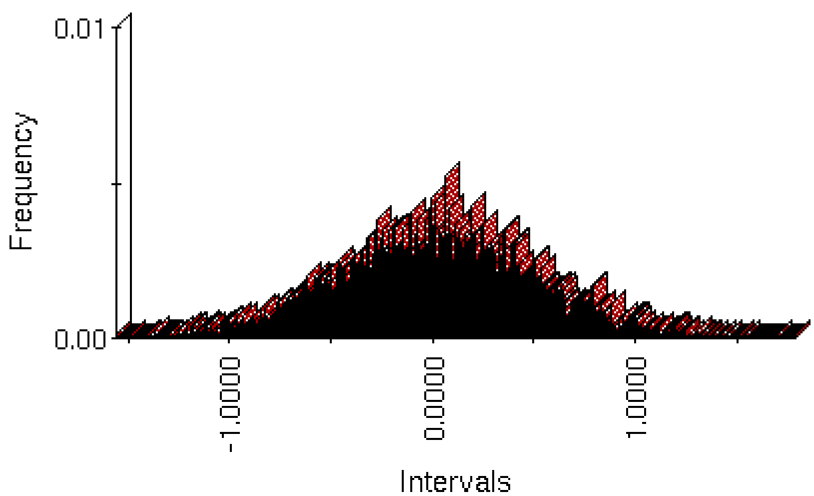
$$F(x) = \begin{cases} 0 & x < a \\ \dfrac{x-a}{b-a} & a \le x \le b \\ 1 & x \ge b \end{cases} \qquad (1.26)$$

Normal Distribution

This is the most important distribution in communications. This is the classic bell shaped distribution also called Gaussian distribution. For a zero-mean distribution, its main parameter is the variance. To the signal (1.26), a normally distributed noise of variance 0.1 has been added. Looking at it, the signal does not look much different than (1.27) but if we plot the distribution we (1.29). We notice not all amplitudes are equally probable. This type of noise describes thermal noise and much of the noise we see in electronic equipment.



1.27 – Transmitted signal which has picked up normally distributed noise of variance 0.1.



1.28 – The bell shaped probability distribution of a normally distributed random variable.

PDF and CDF of a normally distributed random variable for n events

PDF

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad (1.27)$$

CDF

$$\frac{1}{2}\left(1+erf\,\frac{x-\mu}{\sigma\sqrt{2}}\right) \quad (1.28)$$

Mean

$$\mu \quad (1.29)$$

Variance

$$\sigma^2 \quad (1.30)$$

Transforms in signal processing

One word you hear a lot in signal processing is **domain**. The domain refers to the independent parameter of the signal in question. When we are looking at signals with amplitude, phase, current etc vs. time, the time is the independent variable and this signal is in time-domain. Signals in time-domain are easy to understand. The harder one is the frequency domain, which is a transformation of time and where the independent variable is frequency. Here we might be looking at distribution of amplitude over various frequencies in the signal. In communications, signal processing is done in these two domain or dimensions.

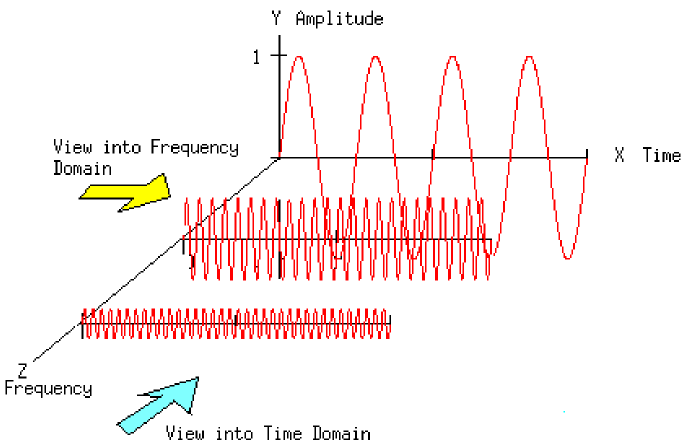


Fig. 1.29 – Frequency and time domain of a signal

The two domains are just two different views of the signal, one from the time perspective and the other from the perspective of the frequency as shown in (1.29). To deal with these dimensions, going back and forth between the time domain and frequency domain, we use **transform theory**. Every field has its own special math and in communication science we use transform theory. A transform is just that, it is a mathematical transformation of data from one type to another with the goal of making handling of data easier and intuitive.

Log function is a form of a transform. It is used extensively in communications to ease multiplication of large numbers. Instead of multiplying 1000 by 10000, we take logs which are 3 and 5 respectively and then add these to determine the result, much easier than multiplying big numbers. Electronic devices like most of us, like adding better than multiplying, so Log transform makes processing easier and faster.

Fourier Transform (FT) and Discrete Fourier Transform (DFT)

Fourier transform is based on the understanding that a periodic signal that is stationery and ergodic can be represented by the sum of various sinusoids. Fig. (1.30b) shows the three sinusoids that form the signal of (1.30a). The Fourier transform is just a plot of the amplitude of each of the component frequencies as in (1.31)

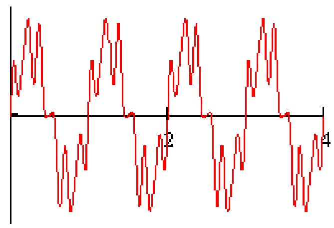


Fig. 1.30a – A periodic signal composed of three sinusoids.

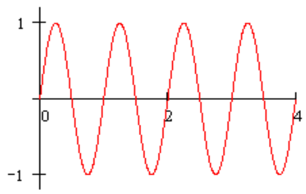


Figure 1.30b - Sine wave 1

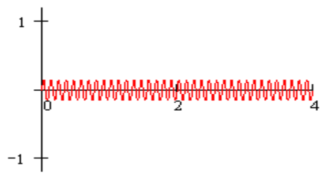


Figure 1.30c - Sine wave 2

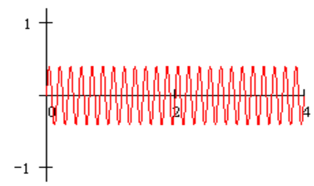


Figure 1.30d - Sine wave 3

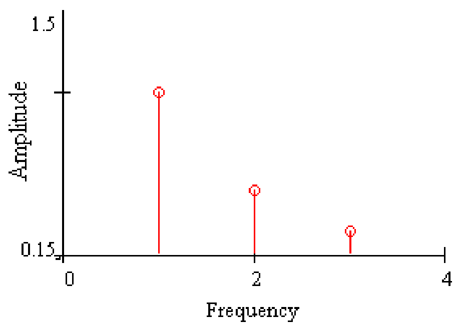


Fig. 1.31 – View into frequency domain a signal consisting of three sine waves of frequencies 1, 2 and 3 Hz.

This fundamental algorithm allows us to look at time-domain signals in frequency-domain and vice-a-versa. Fourier transform makes the signals easier to understand and manipulate. It does that by recognizing that frequency is analogous to time and any signal can be considered a collections of a lot of frequencies of different amplitudes, such as the signal shown in (1.30).

Fourier transform is kind of a de-constructor. It gives us the frequency components which make up the signal. Once we know the frequency components of the signal we can use it to manipulate it.

The Discrete Fourier transform (DFT) operates on discrete samples and is used most often in analysis and simulation of signal behavior.

Laplace transform

Laplace transform is a general transform of which the Fourier transform is a special case. Where the Fourier transform assumes that the target signal is successfully recreated by sum of constant-amplitude sinusoids. Laplace transform basis functions, instead are exponentially enveloped sinusoids as shown in figure (1.32). This allows analysis of transient signals.

An exponential is given by the general expression of

$$x(n) = a^n \tag{1.31}$$

A positive value of n gives a increasing exponential and a negative value a decreasing function. When this function is multiplied by a sinusoids, we get a damped signal. Assorted combinations of these allows us to analyze nearly any signal.

La Place transform is used most often in the analysis of transient signals such as the behavior of phase locked loops. Here a signal may be varying in its amplitude and phase. The error signals computed for these signals that allow correction of the incoming signal frequency have the characteristics of damped signals.

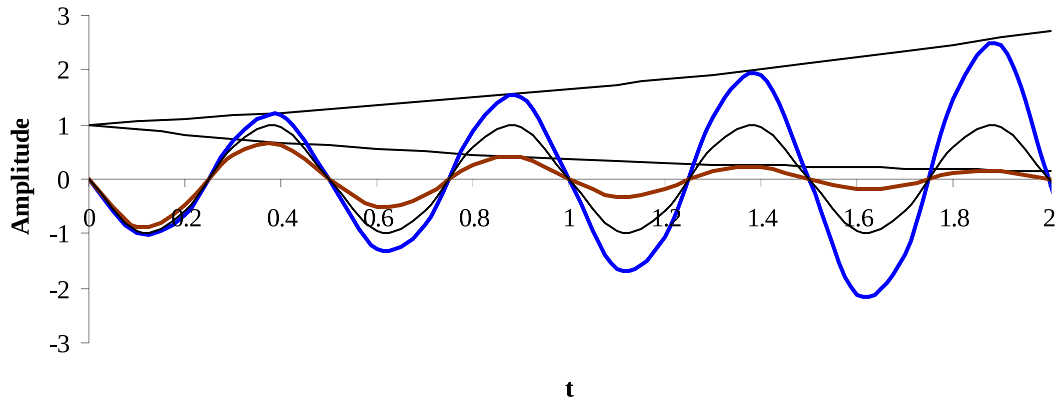


Fig. 1.32 – A signal that is not stationery can be represented by a sum of exponential the amplitude of which is not constant.

Z-Transform

Z-transform is a discrete-time counterpart of the Laplace transform. It is used for transient and time-varying signals (non-stationery) but for discrete signals only. For real-valued signals, $t > 0$, the Laplace transform is a generalization of the Fourier transform, whereas Z-transform is a generalization of the Discrete Fourier transform.

Realness of signals

In communications, we often talk about complex and real signals. Initially this is most confusing. Aren't all signals real? What makes a signal complex? Let's start with what you know. If you scream into the air, this is a real signal. In fact all physical signals are real. They are not *complex* in any sense. What do we mean by complex? Any signal you can create is as real, just as a wire you string between two points. The act of complexification comes when we take these signals into signal processing and do a separation of the signal into components based on some predefined basis functions. Complexness of signals is a mathematical construct that is used in baseband and low frequency processing.

Communications signal processing is mostly a two dimensional process. We take a physical signal and map it into a preset signal space. The best way to explain is to use the Cartesian x and y axis projections of a line. A line is real, no matter at what angle. From the myriad of things we can do to a line, one is to compute its projections into a Cartesian space. The x-axis and the y-axis projections of a line are its complex descriptions. This allows us to compare lines of all sorts on a common basis. Similarly most analog and natural signals are real but can be mapped into real and complex projections which are essentially like the x-y projections of a line.

The x projection of a signal is called its I, inphase or quadrature projection and y projection is called its Q – out of phase or quadrature projection. The word **quadrature** means perpendicular. So I and Q projections are perpendicular in mathematical sense. The real signals are arithmetic sums of the I and the Q signals and not two separate signals that are transmitted orthogonally somehow.

Unlike the projection of a line, which is also a line, we refer to signal projections as **vectors**. Here a vector is not a line but a set of ordered points, such as sampled points of a signal. These ordered sets are called vectors in signal processing and inherit many of the same properties as two dimensional vectors. Just as two plane vectors can be orthogonal to each other, similarly vectors consisting of sets of points can also be orthogonal.

Complex number math is used at baseband and in modulation. In modulation, quadrature representation also includes a frequency shift so the word quadrature representation can also sometimes mean modulation of which more is discussed in the modulation chapter.

For mistakes etc, please contact me at mntcastle@earthlink.net.

Charan Langton

Complextoreal.com

Some references:

http://blackmagic.com/ses/bruceg/COMMBOOK/CH1/ch1commbook.html

http://www.csupomona.edu/~zaliyazici/egr544/egr544-1.pdf

http://www.winlab.rutgers.edu/~crose/545_html/lect1/node5.html