Introduction to Signals

Outline

I. Elementary Signal Concepts

A. Signal Definition and Signal Descriptions

B. Elementary Signal Characteristics

i Signal Support Characteristics

ii. Signal Value Characteristics

ii. Signal Shape Characteristics

II. Elementary Signal Operations

III. Sinusoidal and Complex Exponential Signals

This is discussed in Chapter 2 of the text, not in these notes.

I. Elementary Signal Concepts

Reading Assignment: Chapter 1 and these notes. It is recommended that you review these notes every now and then throughout the term. Some of these elementary concepts will only be needed later in the course, and some will only be well understood after you have had more experience with signals.

A. Signal Definition and Signal Descriptions

Definition:

A "signal" or "waveform" is a time-varying numerical quantity.

More precisely, a signal is a function of time. That is for each value of time t there is number called the *signal value at time* t.

Notation:

We typically use lower case letters like x, y, s or subscripted letters like x_1 to represent signals, i.e. functions of time.

Most frequently, we show time t as the argument of such function, as in $x(t)$.

Beware of the Ever-Present Notational Ambiguity:

When you see "x(t)" written, sometimes the writer intends you to think of the value of the signal at the specific time t, as in $x(3.1)$, and sometimes $x(t)$ means the whole signal -- that is, the writer intends you to think about the whole signal, i.e. the signal values at all times. When it is essential that reader think about the whole signal, writers will sometimes write x or $\{x(t)\}$ instead of $x(t)$.

Continuous- and Discrete-Time Signals:

If the time variable ranges over a continuum of values, we say that the signal is *continuous-time*. If the time variable ranges over a discrete set of values we say the signal is *discrete-time.*

More specifically, we assume unless stated otherwise that every continuous-time signal $x(t)$ has time t ranging over all real numbers from $-\infty$ to $+\infty$. In mathematical terms we say that the *domain* of the function x is the interval $(-\infty,\infty)$.

Similarly, unless stated otherwise, every discrete-time signal is assumed to have time t ranging over the set of all integers: {..., -2, -1, 0, 1, 2, ... }. That is, the domain of the function x is the set of all integers. When dealing with discrete-time signals it is most common to use one of the symbols i, j, k, l, m, or n to denote time rather than t. It is also common to put the time variable within square brackets '[ ]', rather than ordinary parentheses. For instance, the following are examples of the notation used for discrete-time signals $x[n]$, $y[k]$, $z_1[i]$.

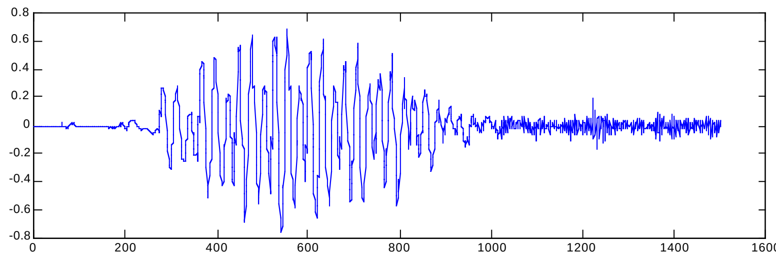
Signal Descriptions

Sometimes signals are described with formulas and sometimes they cannot be so described.

Examples of continuous-time signals described with formulas:

$$x(t) = t^2, \qquad y(t) = 3 \sin(47 t), \qquad z(t) = \begin{cases} 2, & t<0 \\ t^2, & 0 \le t \le 1 \\ 3\sin(4t), & t>1 \end{cases}$$

Example of a continuous-time signal that is not describable with a formula:

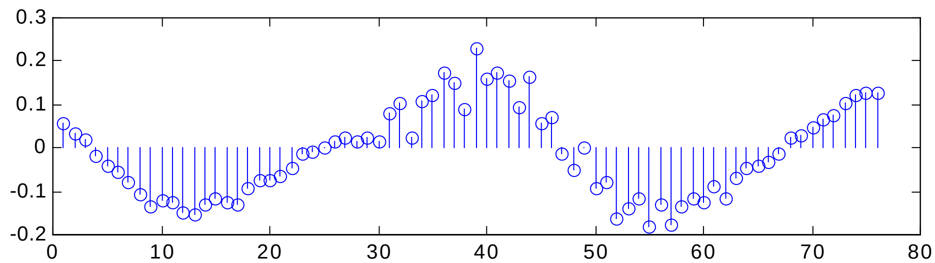


Food for thought: The signal shown above is a recording of me speaking a couple of words. Everything I said and everything you would hear is embodied in the function plotted above.

Examples of discrete-time signals described with formulas:

$$x[n] = n^2, \qquad y[n] = 3 \sin(47 n), \qquad z[n] = \begin{cases} 2, & n<0 \\ n^2, & 0 \le n \le 10 \\ 3\sin(4n), & n>10 \end{cases}$$

Example of a discrete-time signal that is not describable with a formula:



Are signals described by formulas more "real" or "authentic" than signals that are not so describable? What does it mean to "describe a signal with a formula"? Over the centuries, it has been found useful to give names to certain basic mathematical operations, such as '+', '-', '×' '/', x^2, $\ln(x)$, e^x, $|x|$ etc. and certain basic functions, such as $\sin(x)$, $\cos(x)$, $\Gamma(x)$, etc. To "describe a signal with a formula" is simply to say that it can be expressed in terms of previously defined operations and formulas. A signal that is not describable by a formula may simply be a function waiting to be blessed with its own name. Or it may be a function that has not previously occurred and may never occur again. Generally, we do not consider signals described by formulas to be any more real or authentic than those that are not so describable.

Note that a formula describing a signal can be quite complex, as in

$$s(t) = \sum_{i=1}^{N} a_i \cos(b_i t + \phi_i)$$

where N, $a_1,\ldots,a_N$, $b_1,\ldots,b_N$, $\phi_1,\ldots,\phi_N$ are "signal parameters", i.e. constants or variables that one needs to know in order to fully determine the signal. It will be important that you develop the skill of being able to work with complex signal formulas. For example, when you see the summation sign Σ, you should recognize that it is just an abbreviation for a sum of N terms. Indeed, to help you to better understand the signal described by a summation, it is often useful to write it in its unabbreviated form, e.g.

$$s(t) = a_1 \cos(b_1 t+\phi_1) + a_2 \cos(b_2 t+\phi_2) + \ldots + a_N \cos(b_N t+\phi_N)$$

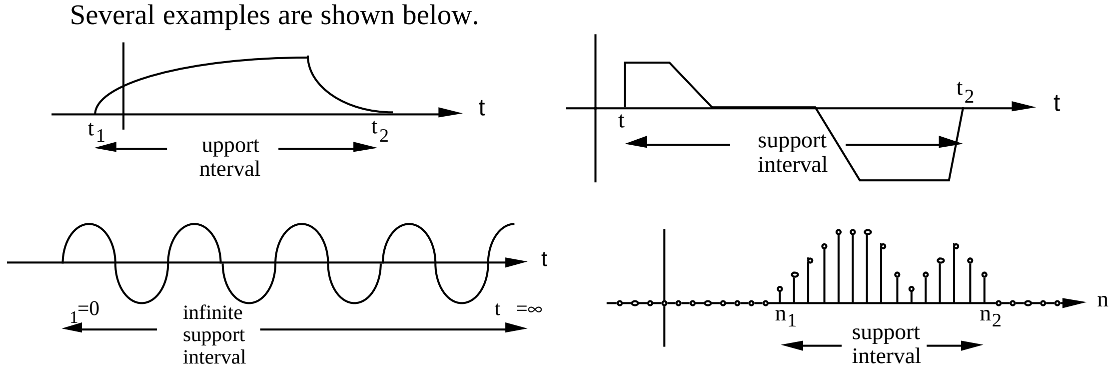
B. Elementary Signal Characteristics[1]

We will primarily present the characteristics of continuous-time signals. There is a discrete-time version of each of these, which will be presented later.

i. Signal Support Characteristics

These are signal characteristics related to the time axis.

Support Interval: Roughly speaking the *support interval* of a signal $x(t)$ is the set of times such that the signal is not zero. More precisely the support interval of a signal $x(t)$ is the smallest interval[2] of times $[t_1,t_2]$ such that the signal is zero outside this interval. We often abbreviate and say simply *support* or *interval* instead of support interval. Several examples are shown below.



Duration: The *duration* or *length* if a signal $x(t)$ is the length of it support interval. Some signals have finite duration and others have infinite duration. For example, the first two signals above have finite duration, and the third signal has infinite duration.

Outside of EECS 206, one will occasionally encounter situations where signals are considered to be *undefined* at times outside their support interval. However, within EECS 206, unless explicitly stated otherwise, we assume the signal value to be 0 outside the support interval. Indeed, we will often define a signal simply by describing its values in some interval, with the presumption that the signal is zero for all times outside this innterval. For example, if we introduce a signal as

$$x(t) = t^2, \quad 1 \le t \le 2 \; ,$$

then it should be understood that $x(t) = 0$ for $t < 0$ and $t > 2$.

---

[1]You do not need to memorize all of these. Rather you need to be aware of the existence of these characteristics, so you can look up and apply the appropriate ones at the appropriate times.

[2]Intervals can be open as in (a,b), closed as in [a,b], or half-open,half-closed as in (a,b] and [a,b). For continuous-time signals, in almost all cases of practical interest, it is not necessary to distinguish the support interval as being of one type or the other.

Pulses: Signals with short duration are often called *pulses*. Note that "short" is a subjective or relative designation.

Negative times and time zero: In some of the examples above the signal interval included negative times. What is the significance of negative time? To answer this, one must first answer the question: What is *time zero*? Basically, time zero is just some convenient reference time. Accordingly, a negative time simply represents a time prior to the reference time. For example, a radar system sends a pulse and waits to record the return times of reflections of this pulse from distant objects. It is usually convenient to let "time zero" be the time at which the original pulse was transmitted. Then $t = -1.8$, means 1.8 units of time before the reference time.

ii. Signal Value Characteristics, aka Signal Statistics

We now consider the values a signal $x(t)$ takes.

Maximum and minimum values: If $x(t)$ denotes some generic signal, then it has a *maximum value* x_{max} and a *minimum value* x_{min}. If these are both finite, i.e. $x_{max} < \infty$ and $x_{min} > -\infty$, then the signal is said to be *bounded*.

What do negative vs. positive signal values represent? The answer depends on the application. As an example, when a microphone responds to a sound, there is usually a diaphragm that moves back and forth, tracking the fluctuations in air pressure that constitute the sound. When the diaphragm is pushed one way, the microphone produces a positive voltage; when pulled the other way, it produces a negative voltage.

Average value: A signal also has an average value. Specifically, the *average* or *mean value* of $x(t)$ over the interval from t_1 to t_2 is

$$M(x) = \frac{1}{t_2-t_1} \int_{t_1}^{t_2} x(t) \, dt \, .$$

Typically a microphone recording has average equal to zero, or very nearly so. In electrical systems, $M(x)$ is often called the *DC value*, where DC stands for *direct current*. If the interval over which the average is sought is infinite, then the average needs to be defined as a limit. For example, the average of the interval $[0,\infty)$ is

$$M(x) = \lim_{T\to\infty} \frac{1}{T} \int_{0}^{T} x(t) \, dt \, ,$$

and the average over the interval $(-\infty,\infty)$ is

$$M(x) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t) \, dt \, .$$

When a signal average is indicated but an interval is not specified, we mean the average over the entire support of the signal.

Absolute value: Quite often, when a signal has values that are both positive and negative, we are interested in a measure of the signal strength apart from its positive or negative sign. With signal strength in mind, one can compute its *magnitude* or *absolute value*, denoted $|x(t)|$.

Squared value, a.k.a. instantaneous power: In most physical situations, the square of $x(t)$, i.e. $x^2(t)$, is a more useful measure of signal strength a time t than magnitude, because it is proportional to the instantaneous power in the signal $x(t)$ at time t, and because power is a quantity of fundamental importance. For such reasons, we will sometimes refer to $x^2(t)$ as the *instantaneous power* of $x(t)$ at time t. However, one must remember that the actual power is a constant times this, where the constant depends on the specific physical situation. For example, if $x(t)$ represents the current in amperes flowing at time t through a resistor with resistance R ohms, then the instantaneous power absorbed by the resistor is $R\,x^2(t)$ watts.

Mean-squared value, a.k.a. average power: Whereas $x^2(t)$ is an excellent measure of signal strength at an individual time t, quite frequently we need an aggregate measure of signal strength that applies to the whole signal, or to the signal over some specified time interval. In such cases, we will typically use the *mean-squared value* (MSV). Specifically, the MSV of a signal $x(t)$ over the interval t_1 to t_2 is

$$MS(x) = \frac{1}{t_2-t_1} \int_{t_1}^{t_2} x^2(t)\; dt \;.$$

This is also called the *average power* in $x(t)$ over the interval t_1 to t_2. As with the definition of average value of x, this definition needs to incorporate a limit when the interval is infinite. And when no interval is specified, the entire support interval is intended.

As an example, mean-squared value is useful when measuring the strength of the signal received by a radar antenna. If it is large in an interval equal to the length of a radar pulse, then we assume that a reflected pulse has been received during this interval, and determine that this pulse is due to an object whose distance is the elapsed time since the original pulse was transmitted times the speed of light. If it is very small, then we can assume that no reflected pulse has been received during this interval, i.e. there is no object at the corresponding distance.

As another example, mean-squared value is used by electric utility companies to determine how much to charge you for the electricity they have supplied. This is because the amount of fuel required by them to supply your electricity is proportional to the mean-squared value of the current supplied to your home.

As a last example, we mention that mean-squared value is often used as a signal quality measure. For example, suppose $x(t)$ is the signal coming from the leftmost of two microphones that are recording an orchestral concert, and suppose $y(t)$ is the signal fed to the left speaker of your stereo receiver after transmission by an FM radio station. Let $e(t) = x(t) - y(t)$ denote the difference between the two signals, which we consider to be an error signal. Then the MSV of $e(t)$ is a good measure of the quality of the system that records and transmits $x(t)$ to you. It is usually called *mean-squared error.*

RMS Value: A closely related quantity is the root mean-squared value (RMSV), which is simply

$$RMS(x) = \sqrt{MS(x)} = \sqrt{\frac{1}{t_2-t_1} \int_{t_1}^{t_2} x^2(t)\, dts} \;.$$

On the one hand, RMSV is nicer than MSV in that its value is easier to interpret because it is like a typical signal value, whereas the value of the MSV is harder to interpret because it is like the square of a typical signal value. On the other hand, it is usually easier to work with MSV. For example, suppose $x(t)$ has support t_1 to t_2 and suppose $y(t)$ has support t_2 to t_3. Then as derived below, the formula for the MSV of $x(t)+y(t)$ in terms of the MSV's of $x(t)$ and $y(t)$ is simpler than the corresponding formula for the RMSV.

$$MS(x+y) = \frac{1}{t_3-t_1} \int_{t_1}^{t_3} (x(t)+y(t))^2(t)\; dt$$

$$= \frac{1}{t_3-t_1} \int_{t_1}^{t_2} x^2(t)(t)\; dt \;+\; \frac{1}{t_3-t_1} \int_{t_2}^{t_3} y^2(t)(t)\; dt, \;\; \text{since } y(t)=0,\ t<t_2,\ x(t)=0,\ t>t_2$$

$$= \frac{t_2-t_1}{t_3-t_1}\, \frac{1}{t_2-t_1} \int_{t_1}^{t_2} x^2(t)(t)\; dt \;+\; \frac{t_3-t_2}{t_3-t_1}\, \frac{1}{t_3-t_2} \int_{t_2}^{t_3} y^2(t)(t)\; dt$$

$$= \frac{t_2-t_1}{t_3-t_1} MS(x) + \frac{t_2-t_1}{t_3-t_1} MS(y) .$$

$$RMS(x+y) = \sqrt{MS(x+y)} = \sqrt{\frac{t_2-t_1}{t_3-t_1} (RMS(x))^2 + \frac{t_2-t_1}{t_3-t_1} (MS(y))^2} .$$

Signal Energy: Another closely related quantity is the *energy* of the signal $x(t)$ in the interval t_1 to t_2, which is

$$E(x) = \int_{t_1}^{t_2} x^2(t)\ dt .$$

By comparing this, with previous definitions, we see that energy is the integral of instantaneous power. It is also the average power multiplied by the length of the interval. Alternatively, average power is energy divided by the length of the interval over which it is computed. A little thought will convince you that it is energy for which an electric utility company actually charges.

Since signal energy and average power (MSV) are related by a constant, the choice of which to focus on is often a matter of taste. If you focus on one, you can easily compute the other.

However, for signals infinite duration often have infinite energy (over their entire support). For such signals, power is usually a more interesting quantity than energy.

Variance[3] and Standard Deviation[3]: The mean-squared value of $x(t)$ minus its average value is called the *variance* of x. The square root of the variance is called the *standard deviation*. That is, the variance[4] of x over the interval t_1 to t_2 is

$$\sigma^2(x) = MS(x-M(x)) = \frac{1}{t_2-t_1} \int_{t_1}^{t_2} (x(t)-M(x))^2\ dt$$

and the standard deviation is

$$\sigma(x) = RMS(x-M(x)) = \sqrt{\frac{1}{t_2-t_1} \int_{t_1}^{t_2} (x(t)-M(x))^2\, dt}$$

The variance and standard deviations are useful measures of how "variable" is the signal. A signal with small variance or standard deviation stays close to its average value most of the time, whereas a signal with large variance or standard deviation does not. As with MSV vs. RMSV, standard deviation values are usually easier to interpret because their values are commensurate with signal values. On the other hand, variances are usually easier to compute and work with.

Relationship Between Mean-Squared Value, Variance and Average Value: The following is a useful relationship.

$$MS(x) = \sigma^2(x) + M(x)$$

Derivation:

$$\sigma^2(x) = \frac{1}{t_2-t_1} \int_{t_1}^{t_2} (x(t)-M(x))^2\ dt \quad = \frac{1}{t_2-t_1} \int_{t_1}^{t_2} (x^2(t)-2M(x)x(t)+M^2(x))\ dt$$

$$= \frac{1}{t_2-t_1} \int_{t_1}^{t_2} x^2(t)\ dt - \frac{1}{t_2-t_1} \int_{t_1}^{t_2} 2\,M(x)\ x(t)\ dt + \frac{1}{t_2-t_1} \int_{t_1}^{t_2} M^2(x)\ dt$$

---

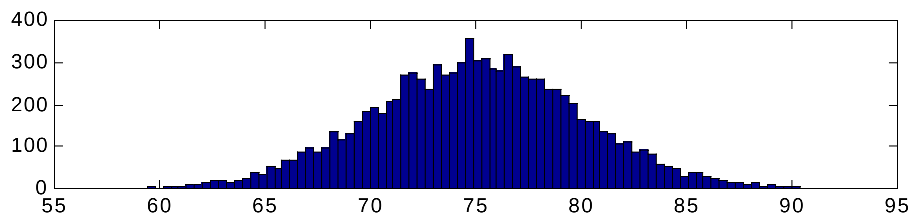
[3]Variance and standard deviation will not be needed early in the course. You can skim them now, and return to them when needed.
[4]The use of the term σ^2 for variance and σ for standard deviation is traditional.

$$= MS(x) - 2\,M(x)\,\frac{1}{t_2-t_1}\int_{t_1}^{t_2} x(t)\,dt + M^2(x)\,\frac{1}{t_2-t_1}\int_{t_1}^{t_2} dt,$$ since M(x) is constant we bring it outside integrals

$$= MS(x) - 2\,M(x)\,M(x) + M^2(x)$$ by definition of M(x) and by doing the integral on the right hand side

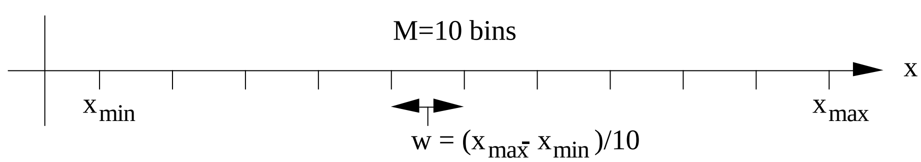
$$= MS(x) - M^2(x)\,,$$ which is the desired relationship.

Signal Value Distribution and Histograms: The minimum, maximum, average, and mean-squared value are numbers that each tell us something about the values that appear in the signal. The *signal value distribution* gives a more complete picture. Before introducing it, let us review the general meaning of the word *distribution*. As one example, consider the collection of grades resulting from an exam. If we speak of the "distribution of these grades", we mean a plot like that shown below. The horizontal axis shows the possible grades, and the height of the plot above a given grade is proportional to the number of exam papers with that grade. As another example, consider the distribution of incomes of residents of Michigan. Again this is a plot like the one shown below. In this case, the horizontal axis shows the possible incomes, and the height of the plot above a given income is proportional to the number of people with that income.



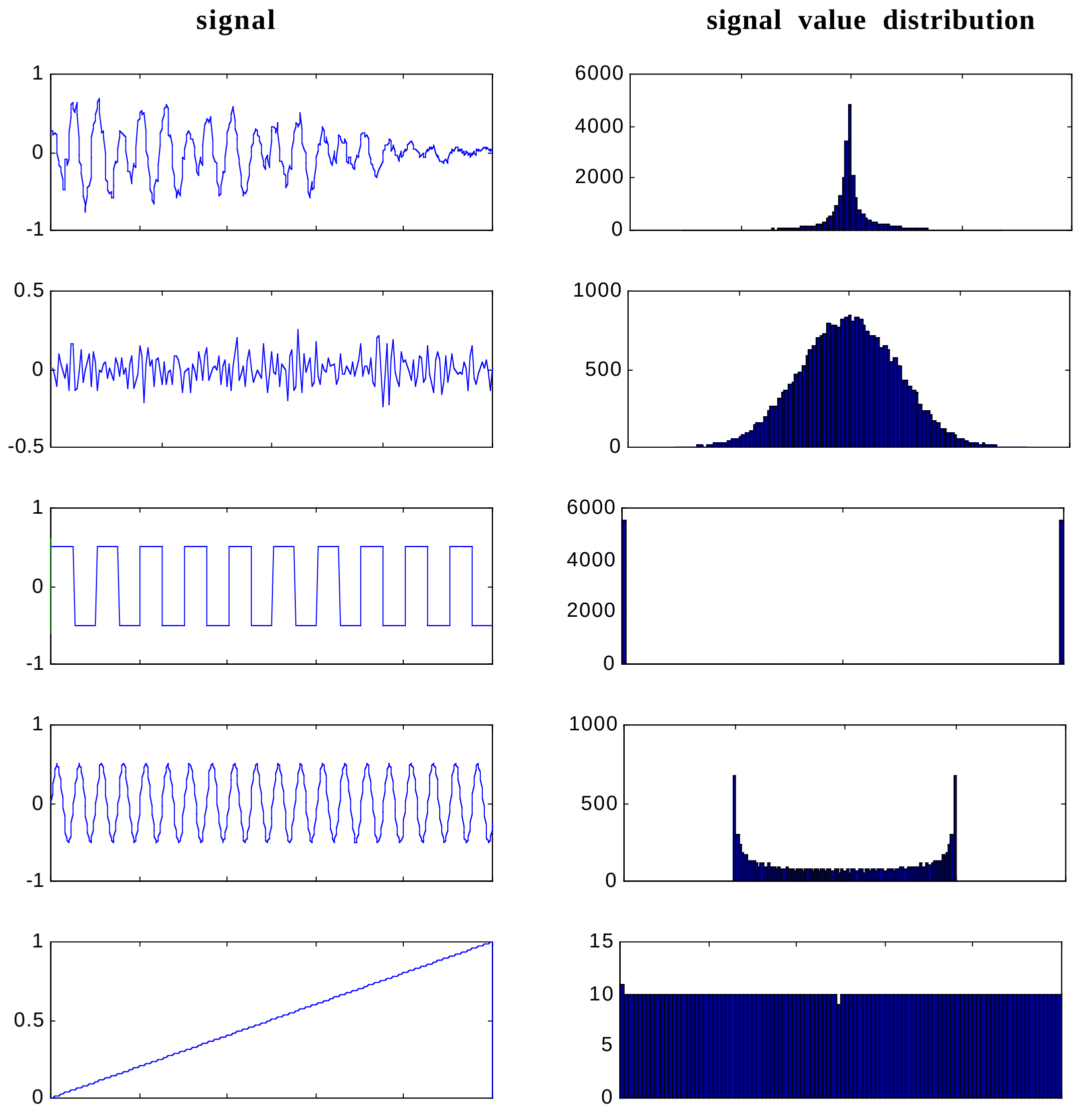
One may similarly consider the distribution of many, many quantities. Not surprisingly, in signals and systems, we are often interested in the distribution of values of a signal x(t), which we call its *signal value distribution*. That is, for a given signal x(t) we want a plot whose horizontal axis shows the signal values and whose height above a given signal value is proportional to the frequency with which that value[5] occurs in the signal.

How do we plot the signal value distribution of a signal x(t)? The most common way is make and plot a *histogram*. Specifically, we divide the range of signal values from x_{min} to x_{max} into M equal width *bins*, as illustrated below, where M is some integer, usually in the range 10 to 1000.



If the signal is discrete-time, we count the number of signal values that lies within each bin. We then plot each count above the bin, as illustrated below. If the signal is continuous-time, then we repeat the same procedure on samples of the image. That is, we repeat the procedure on the set of values x(T), x(2T), x(3T), ... where T is the sample spacing. As examples, several signals and their signal value distributions are shown below.

[5]Strictly speaking it is not the frequency of individual values that matter. Rather for any value x, we want the frequency with which signal values lie in a small neighborhood of x, say from $x-\Delta$ to $x+\Delta$, where Δ is a small constant.



These histograms were computed with Matlab using the command hist(X,M), where X is a vector containing signal samples, and M is the desired number of bins.

We now justify the statement made earlier that the signal value distribution gives a more complete picture of the signal values than its minimum, maximum, average and mean-squared values. We do this by showing that these latter quantities can be determined, at least approximately, from a histogram. First, the minimum and maximum values will be readily apparent from the histogram. For example, the maximum value is approximately equal to the largest bin center for which the histogram is not zero.

Next, let us show how the average value $M(x)$ can be computed from the histogram. Let $x[1], x[2], \ldots, x[N]$ denote the signal samples. If the histogram has M bins, then the width of each bin will be $W = (x_{max}-x_{min})/M$. The first bin is the interval $(x_{min},x_{min}+W)$, the second bin is the interval $(x_{min}+W,x_{min}+2W)$, and so on. Let C_i denote the center of the ith bin. That is, $C_i = x_{min} + iW - W/2$, for $i = 1,\ldots,M$. Let N_i denote the number of signal values that lie in the ith bin. Then the histogram is simply a plot of the points (c_i,N_i), $i = 1,\ldots,M$. The average value of the N signal samples is

$$M(x) = \frac{1}{N} \sum_{n=1}^{N} x[n]$$

Now we observe that we can approximately compute the sum in the above in a different matter. Since there are N_i signal values in the ith bin, we know that there are N_i signal values that approximately equal C_i. The sum of these values is approximately N_iC_i. Making this approximation for each of the bins leads to

$$\sum_{n=1}^{N} x[n] \cong N_1C_1 + N_2C_2 + \ldots + N_M C_M .$$

Therefore,

$$M(x) \cong \frac{1}{N} \sum_{i=1}^{M} N_i C_i = \sum_{i=1}^{N} \frac{N_i}{N} C_i$$

That is, the average signal value $M(x)$ is approximately the weighted average of the C_i's (the bin centers), where the weight multiplying C_i is the fraction of samples that lie in the ith bin.

In an entirely similar fashion one may show that

$$MS(x) \cong \sum_{i=1}^{N} \frac{N_i}{N} (C_i)^2 .$$

Summary of Signal Value Characteristics

The following table shows the definitions of the signal characteristics mentioned previously, with the exception of signal value distribution, which is not easily summarized in table form. It also lists the analogous characteristics for discrete-time signals.

	Continuous-time signal $x(t)$	**Discrete-time signal $x[n]$**
support interval	$[t_1,t_2]$	$\{n_1,n_1+1,...,n_2\}$
duration	t_2-t_1	n_2-n_1+1
maximum value:	$x_{max} = \max_t x(t)$	$x_{min} = \max_n x[n]$
minimum value:	$x_{min} = \min_t x(t)$	$x_{min} = \min_n x[n]$
average value:	$M(x) = \frac{1}{t_2-t_1} \int_{t_1}^{t_2} x(t)\, dt$	$M(x) = \frac{1}{n_2-n_1+1} \sum_{n=n_1}^{n_2} x[n]$
magnitude:	$\lvert x(t)\rvert$	$\lvert x[n]\rvert$
squared value, aka instantaneous power:	$x^2(t)$	$x^2[n]$
mean-squared value, aka average power:	$MS(x) = \frac{1}{t_2-t_1} \int_{t_1}^{t_2} x^2(t)\, dt$	$MS(x) = \frac{1}{n_2-n_1+1} \sum_{n=n_1}^{n_2} x^2[n]$
RMS value:	$RMS(x) = \sqrt{MS(x)}$	$RMS(x) = \sqrt{MS(x)}$
energy:	$E(x) = \int_{t_1}^{t_2} x^2(t)\, dt$	$E(x) = \sum_{n=n_1}^{n_2} x^2[n]$
variance:	$\sigma^2(x) = MS(x-M(x))$	$\sigma^2(x) = MS(x-M(x))$
standard deviation:	$\sigma(x) = \sqrt{MS(x-M(x))}$	$\sigma(x) = \sqrt{MS(x-M(x))}$
relationship:	$MS(x) = \sigma^2(x) + M(x)$	$MS(x) = \sigma^2(x) + M(x)$