Digital Signal Processing

Dr. Mohannad K. Sabir
Al-Khwarizmi Engineering College
Biomedical Engineering Department

DSP

- Chapter 1 **Statistics, Probability and Noise**
- Chapter 2 **ADC and DAC**
- Chapter 3 **DSP Software**
- Chapter 4 **Linear Systems**
- Chapter 5 **Convolution**
- Chapter 6 **The Discrete Fourier Transform & Fast Fourier Transform**
- Chapter 7 **Digital Filters FIR & IIR**
- Chapter 8 **FFT Convolution**
- Chapter 9 **Data Compression**
- Chapter 10 **Digital Signal Processors**

DSP CHP 1

[Statistics, Probability and Noise]

- ***signal*** is a description of how one parameter is related to another parameter.
- ***continuous signal*** when both parameters can assume a continuous range of values.
- ***discrete signals or digitized signals*** are signals formed from parameters that are quantized.
- ***Vertical axis*** may represent voltage, light intensity, sound pressure, or an infinite number of other parameters. Since we don't know what it represents in this particular case, we will give it the generic label: **amplitude.** This parameter is also called several other names: the **y axis,** the **dependent variable,** the **range,** and the **ordinate.**

DSP CHP 1

[Statistics, Probability and Noise]

- ***Horizontal axis*** represents the other parameter of the signal, going by such names as: the **x-axis, the independent variable, the domain, and the abscissa.** ***Time*** *is the most common parameter to appear on the horizontal axis* of acquired signals.
- we will simply label the horizontal axis: **sample number** for ***discrete signals***, and ***time, distance, x,*** *etc, for* ***continuous signal.***
- The parameter on the y-axis (the dependent variable) is said to be a **function** of the parameter on the x-axis (the independent variable).

DSP CHP 1

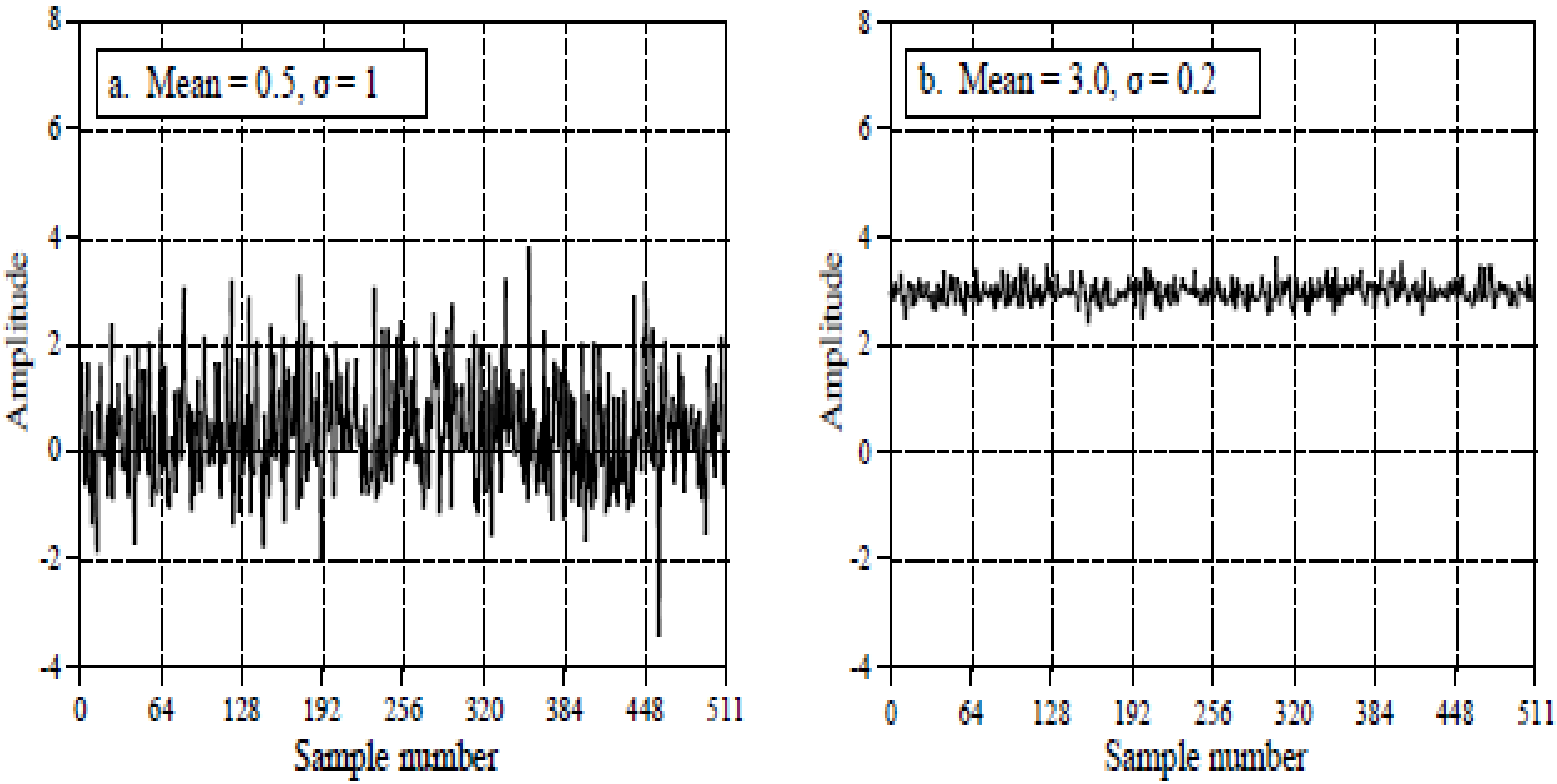


FIGURE 1 Examples of two digitized signals with different means and standard deviations.

DSP CHP 1

[Statistics, Probability and Noise]

Mean and Standard Deviation

- ***Mean***, indicated by **μ** (a lower case Greek *mu),* is the statistician's jargon for the average value of a signal.

$$\mu = \frac{1}{N} \sum_{i=0}^{N-1} x_i \qquad \text{--------- (1)}$$

- The variable, ***N,*** is widely used in DSP to represent the total number of samples in a signal.

DSP CHP 1

[Statistics, Probability and Noise]

- In the first notation, the sample indexes run from 1 to *N (e.g., 1 to 512). In the second* notation, the sample indexes run from 0 to *N& 1 (e.g., 0 to 511).* Mathematicians often use the first method (1 to *N), while those in* ***DSP*** commonly uses the second (0 to *N- 1).*
- In electronics, the ***<u>mean</u>*** is commonly called the ***DC (direct current)*** value.
- **AC (alternating current)** refers to how the signal fluctuates around the mean value.

DSP CHP 1

[Statistics, Probability and Noise]

- ***Sine*** or square wave, its excursions can be described by its peak-to-peak amplitude.
- Most acquired signals do not show a well defined peak-to-peak value, but have a random nature, such as the signals in Fig. 1.
- A more generalized method must be used in these cases, called the **standard deviation,** denoted by **σ (a lower case Greek *sigma*).**

DSP CHP 1

[Statistics, Probability and Noise]

- In most cases, the important parameter is not the ***deviation*** *from the* **mean**, but the ***power*** *represented by the* ***deviation*** *from the* ***mean****.*
- *For example,* when random noise signals combine in an electronic circuit, the resultant noise is equal to the combined ***power*** *of the* ***individual signals****,* *not their combined* ***amplitude****.*
- As a starting point, the expression, $| x_i - \mu |$, *describes how far the sample i^{th} deviates (differs) from the mean.*

DSP CHP 1

[Statistics, Probability and Noise]

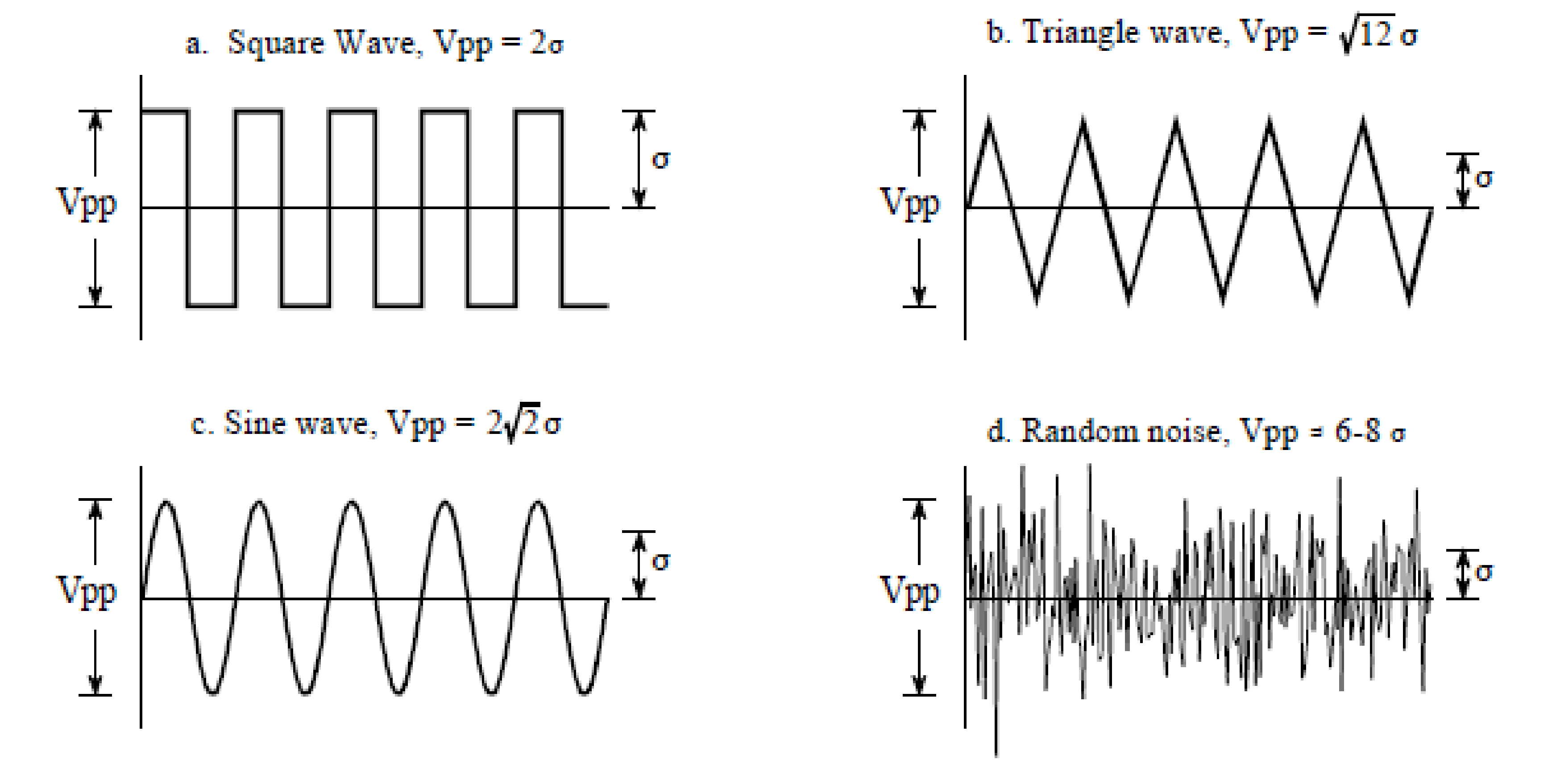
- The *standard deviation is similar to the average deviation, except the* averaging is done with power instead of amplitude.
- To finish, the *square root is taken to compensate for the initial* squaring. In equation form, the standard deviation is calculated:

$$\sigma^2 = \frac{1}{N-1} \sum_{i=0}^{N-1} (x_i - \mu)^2 \quad \text{----------} \; 2$$

DSP CHP 1

[Statistics, Probability and Noise]

- The term, σ^2, occurs frequently in statistics and is given the name **variance.**
- The standard deviation is a measure of how far the signal fluctuates from the mean. The variance represents the power of this fluctuation. Another term you should become familiar with is the **rms (root-mean-square) value,** frequently used in electronics.
- By definition, the standard deviation only measures the **AC** portion of a signal, while the **rms** value measures both the **AC** and **DC** components. If a signal has no **DC** component, its **rms** value is identical to its standard deviation.



- FIGURE 2

Ratio of the peak-to-peak amplitude to the standard deviation for several common waveforms. For the square wave, this ratio is 2; for the triangle wave it is $\sqrt{12} = 3.46$; for the sine wave it is $2\sqrt{2} = 2.83$. While random noise has no *exact* peak-to-peak value, it is *approximately* 6 to 8 times the standard deviation.

- This method of calculating the mean and standard deviation is adequate for many a applications; however, it has two limitations.

1. If the mean is much larger than the standard deviation, Eq. 1 & 2 involves subtracting two numbers that are very close in value. This can result in excessive round-off error in the calculations
2. It is often desirable to recalculate the mean and standard deviation as new samples are acquired and added to the signal. We will call this type of calculation: **running statistics.**

- While the method of Eqs. 1 and 2 can be used for running statistics, it requires that *all of the samples be* involved in each new calculation. This is a very inefficient use of computational power and memory.
- A solution to these problems can be found by manipulating Eqs. 1 and 2 to provide another equation for calculating the standard deviation:

$$\sigma^2 = \frac{1}{N-1}\left[\sum_{i=0}^{N-1} x_i^2 - \frac{1}{N}\left(\sum_{i=0}^{N-1} x_i\right)^2\right]$$

--------- 3

or using a simpler notation,

$$\sigma^2 = \frac{1}{N-1}\left[sum\ of\ squares - \frac{sum^2}{N}\right]$$

- While moving through the signal, a running tally is kept of three parameters:

1. the number of samples already processed,
2. the sum of these samples, and
3. the sum of the squares of the samples (that is, square the value of each sample and add the result to the accumulated value).

- In some situations, the *mean describes what is* being measured, while the *standard deviation represents noise and other* interference. In these cases, the standard deviation is not important in itself, but only in *comparison to the mean.*
- This gives rise to the term: **signal-to-noise ratio (SNR),** *which is equal to the mean divided by the standard deviation.*
- **coefficient of variation (CV).** This is defined as *the standard deviation divided by the mean, multiplied by 100 percent.*

DSP CHP 1

[Statistics, Probability and Noise]

Signal vs. Underlying Process

- **Statistics** is the science of interpreting *numerical data, such as acquired* signals.
- **Probability** is used in DSP to understand the *processes that generate signals.*
- Although they are closely related, the distinction between the **acquired signal** and **the underlying process** is key to many DSP techniques.

DSP CHP 1
[Statistics, Probability and Noise]

- The ***probabilities*** *of the underlying process are **constant**, but the **statistics** of the acquired signal **change** each time the experiment is repeated.*
- This random irregularity found in actual data is called by such names as: **statistical variation, statistical fluctuation, and statistical noise.**
- In particular, for random signals, the typical error between the mean of the *N points, and the mean of the underlying* process, is given by:

DSP CHP 1

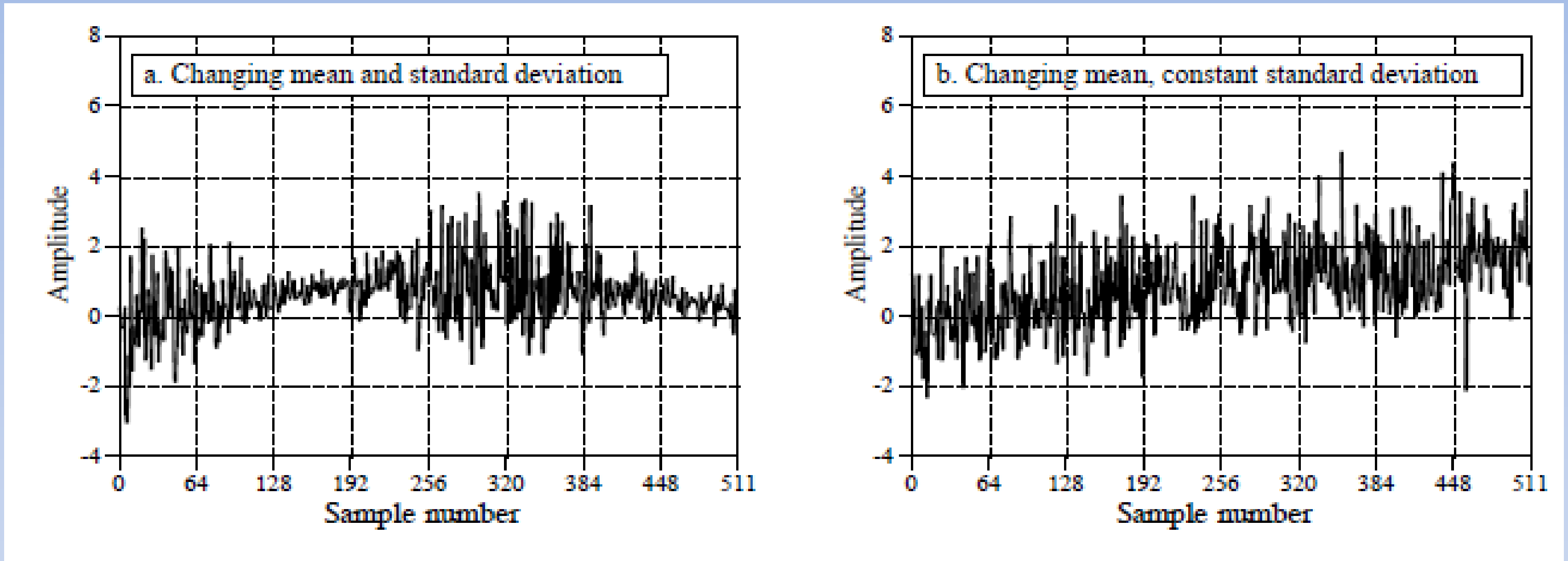
[Statistics, Probability and Noise]

$$Typical\ error = \frac{\sigma}{N^{1/2}} \quad \text{--------- 4}$$

- If ***N*** *is* ***small****, the statistical noise in the calculated mean will be* ***very large****.*
- The **larger** the value of ***N****, the* ***smaller*** *the expected* error will become.
- **The error becomes *zero* as *N approaches* infinity.**

DSP CHP 1

[Statistics, Probability and Noise]



- Figure 3 Examples of signals generated from non stationary processes. In (a), both the mean and standard deviation change. In (b), the standard deviation remains a constant value of one, while the mean changes from a value of zero to two. It is a common analysis technique to break these signals into short segments, and calculate the statistics of each segment individually.

DSP CHP 1
[Statistics, Probability and Noise]

- Processes that change their characteristics in the manner represented in figure 3, are called **nonstationary.**
- In comparison, the signals previously presented in Fig. 1 were generated from a stationary process, and the variations result completely from statistical noise.

DSP CHP 1

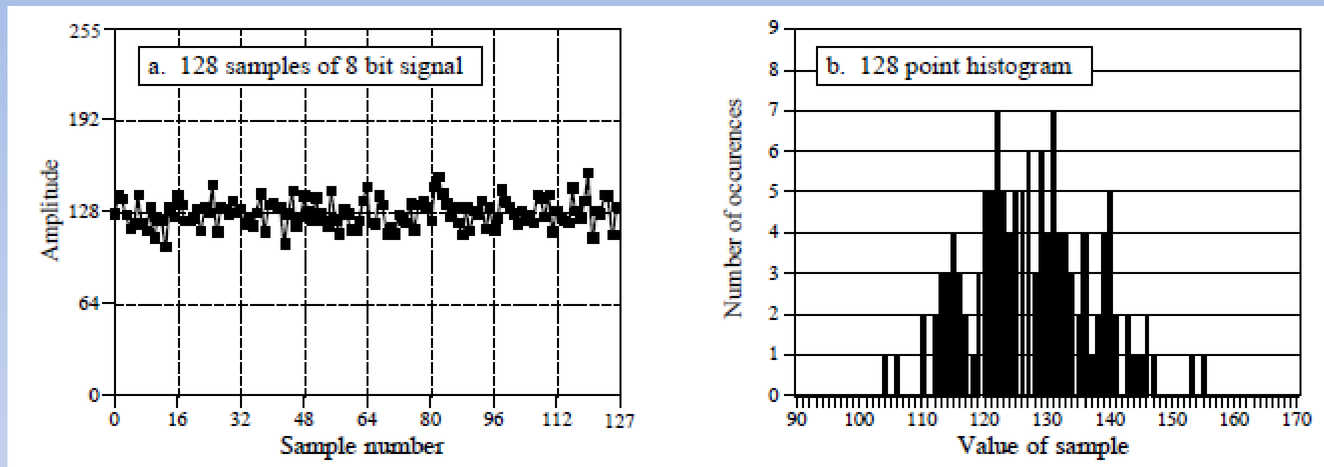
[Statistics, Probability and Noise]

The Histogram, Pmf and Pdf

- The **histogram displays** the *number of samples there are in the signal that have each of these possible values.*
- Suppose we attach an 8 bit analog-to-digital converter to a computer, the value of each sample will be one of 256 possibilities, 0 through 255.

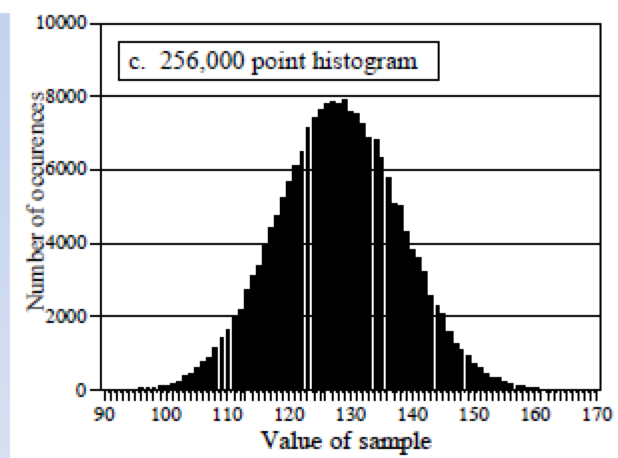
DSP CHP 1

[Statistics, Probability and Noise]



- Figure 4

Examples of histograms. Figure (a) shows 128 samples from a very long signal, with each sample being an integer between 0 and 255. Figures (b) and (c) shows histograms using 128 and 256,000 samples from the signal, respectively. As shown, the histogram is smoother when more samples are used.



DSP CHP 1

[Statistics, Probability and Noise]

- As can be seen, the larger number of samples results in a much smoother appearance. Just as with the mean, the statistical noise (roughness) of the histogram is inversely proportional to the square root of the number of samples used.
- the sum of all of the values in the histogram must be equal to the number of points in the signal:

$$N = \sum_{i=0}^{M-1} H_i \quad \text{--------- } 5$$

DSP CHP 1

[Statistics, Probability and Noise]

- The histogram can be used to efficiently calculate the mean and standard deviation of very large data sets.
- the mean and standard deviation are calculated from the histogram by the equations:

$$\mu = \frac{1}{N} \sum_{i=0}^{M-1} i H_i \quad \text{--------- 6}$$

$$\sigma^2 = \frac{1}{N-1} \sum_{i=0}^{M-1} (i - \mu)^2 H_i \quad \text{--------- 7}$$

DSP CHP 1

[Statistics, Probability and Noise]

- The notion that the acquired signal is a noisy version of the underlying process is very important
- The *histogram is what is formed from an acquired signal.*
- The corresponding curve for the underlying process is called the **probability mass function (pmf).**
- A **histogram** is always calculated using a ***finite number*** of samples, while the **pmf** is what *would be obtained with* ***an infinite*** **number** of samples. The pmf can be estimated (inferred) from the histogram.

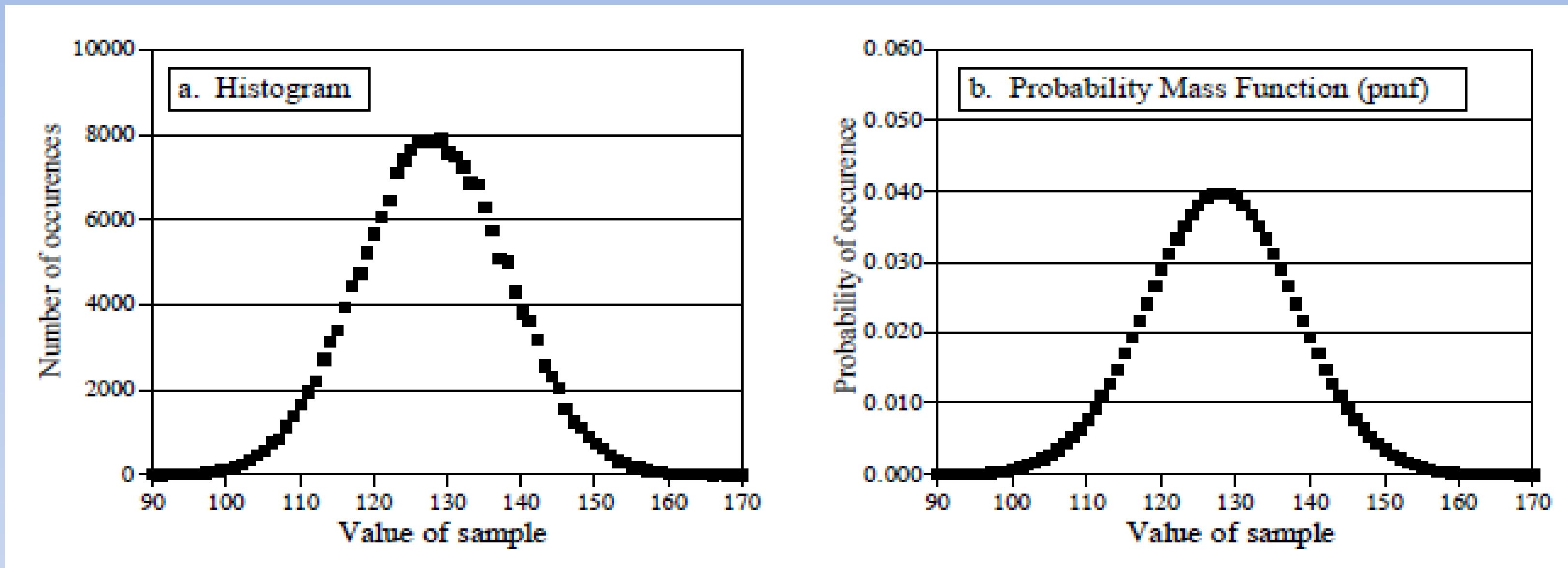
DSP CHP 1

[Statistics, Probability and Noise]

- The vertical axis of the histogram is the *number of times that a particular value occurs in the signal.*
- The vertical axis of the **pmf** contains similar information, except expressed on a *fractional basis.*
- In other words, each value in the histogram is divided by the total number of samples to approximate the pmf.

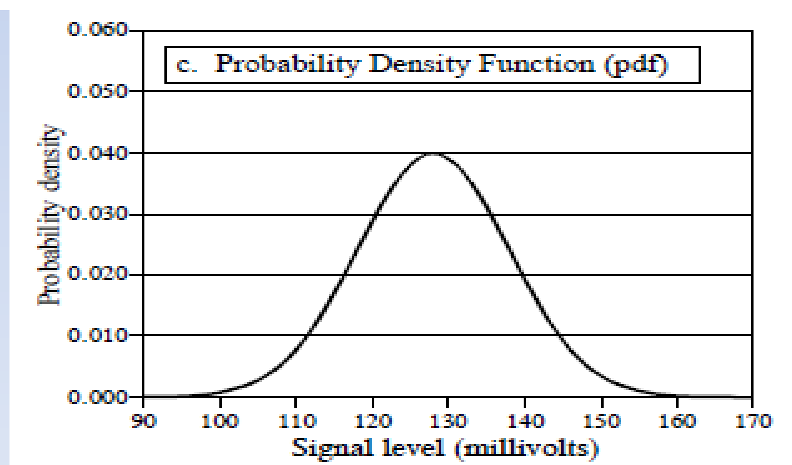
DSP CHP 1

[Statistics, Probability and Noise]



- Figure 5

The relationship between (a) the histogram, (b) the probability mass function (pmf), and (c) the probability density function (pdf). The histogram is calculated from a finite number of samples. The pmf describes the probabilities of the underlying process. The pdf is similar to the pmf, but is used with continuous rather than discrete signals. Even though the vertical axis of (b) and (c) have the same values (0 to 0.06), this is only a coincidence of this example. The amplitude of these three curves is determined by: (a) the sum of the values in the histogram being equal to the number of samples in the signal; (b) the sum of the values in the pmf being equal to one, and (c) the area under the pdf curve being equal to one.



DSP CHP 1

[Statistics, Probability and Noise]

- A problem occurs in calculating the histogram when the number of levels each sample can take on is much larger than the number of samples in the signal.
- This is always true for signals represented in *floating point* notation, where each sample is stored as a fractional value.
- The solution to these problems is a technique called **binning.** This is done by arbitrarily selecting the length of the histogram to be some convenient number, such as 1000 points, often called **bins.**
- The value of each bin represent the total number of samples in the signal that have a value within a *certain range.*

DSP CHP 1

[Statistics, Probability and Noise]

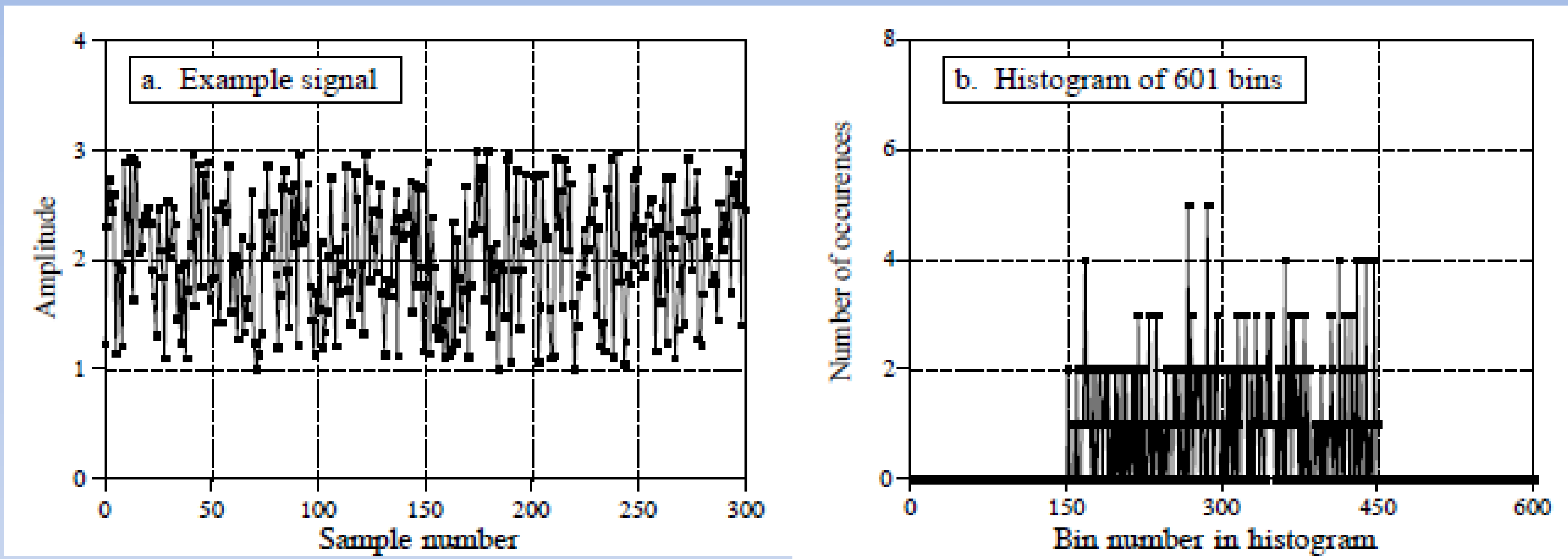
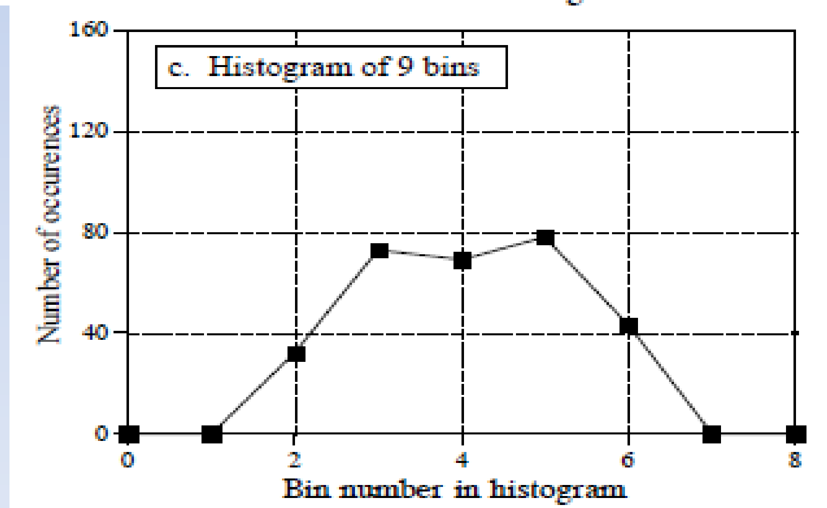


Figure 6

Example of binned histograms. As shown in (a), the signal used in this example is 300 samples long, with each sample a floating point number uniformly distributed between 1 and 3. Figures (b) and (c) show binned histograms of this signal, using 601 and 9 bins, respectively. As shown, a large number of bins results in poor resolution along the *vertical axis*, while a small number of bins provides poor resolution along the *horizontal axis*. Using more samples makes the resolution better in both directions.



DSP CHP 1

[Statistics, Probability and Noise]

The Normal Distribution

- Signals formed from random processes usually have a bell shaped pdf. This is called a **normal distribution, a Gauss distribution, or a Gaussian,** after the great German mathematician, Karl Friedrich Gauss (1777-1855).
- The basic shape of the curve is generated from a *negative squared exponent:*

DSP CHP 1
[Statistics, Probability and Noise]

$$y(x) = e^{-x^2}$$

- This raw curve can be converted into the complete Gaussian by adding an adjustable mean, μ, and standard deviation, σ. In addition, the equation must be normalized so that the total area under the curve is equal to *one, a* requirement of all probability distribution functions.

DSP CHP 1

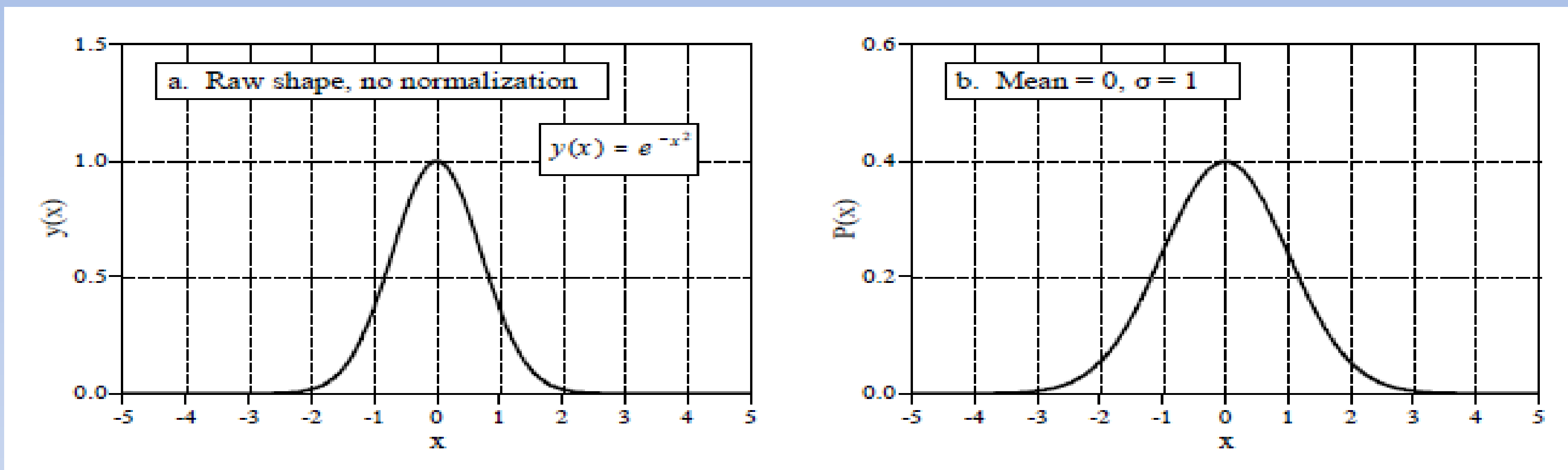
[Statistics, Probability and Noise]

- This results in the general form of the normal distribution, one of the most important relations in statistics and probability:

$$P(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/2\sigma^2} \quad \text{--------- 8}$$

DSP CHP 1

[Statistics, Probability and Noise]



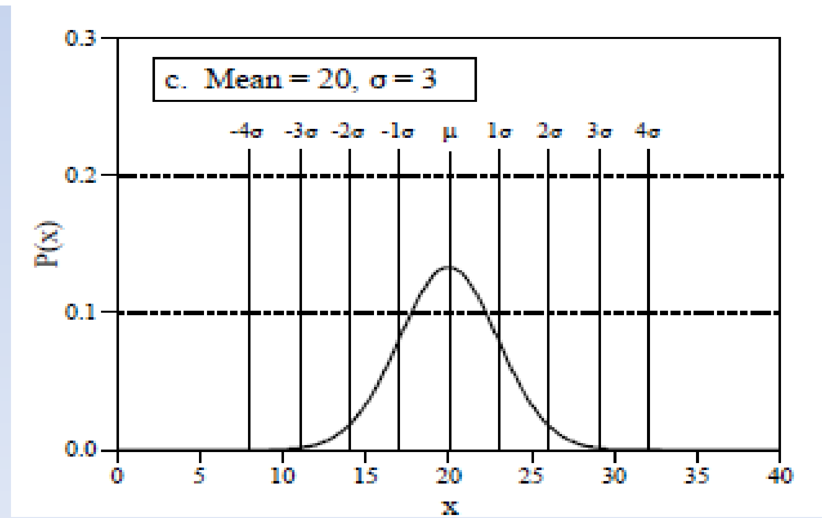


Figure 7

Examples of Gaussian curves. Figure (a) shows the shape of the raw curve without normalization or the addition of adjustable parameters. In (b) and (c), the complete Gaussian curve is shown for various means and standard deviations.

DSP CHP 1

[Statistics, Probability and Noise]

- In practice, the sharp drop of the Gaussian pdf dictates that these extremes almost never occur.
- This results in the waveform having a relatively bounded appearance with an apparent peak to-peak amplitude of about 6-8σ.
- The integral of the pdf is used to find the probability that a signal will be within a certain range of values.
- This makes the integral of the pdf important enough that it is given its own name, the **cumulative distribution function (cdf).**

DSP CHP 1

[Statistics, Probability and Noise]

- To get around this, the integral of the Gaussian can be calculated by ***numerical integration.***

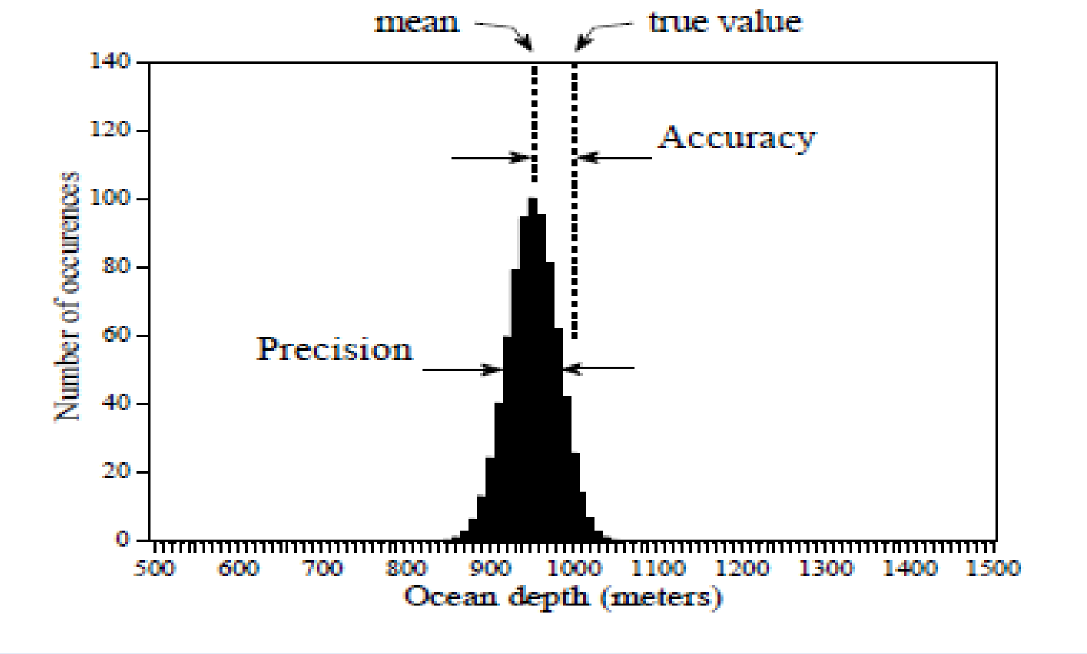
DSP CHP 1

[Statistics, Probability and Noise]

Precision and Accuracy

- **True value** is some parameter you wish to know the value of.
- The method provides a **measured value,** that you want to be as close to the true value as possible.
- ***Precision*** *and* ***accuracy*** *are ways of describing the* error that can exist between these two values.
- The particular measurement could be affected by many factors These measurements are then arranged as the histogram shown below:

DSP CHP 1
[Statistics, Probability and Noise]



- Figure 8

Definitions of accuracy and precision. Accuracy is the difference between the true value and the mean of the under-lying process that generates the data. Precision is the spread of the values, specified by the standard deviation, the signal-to-noise ratio, or the CV.

DSP CHP 1

[Statistics, Probability and Noise]

- The ***mean*** *occurs at the center of the distribution, and* represents the best estimate of the depth based on all of the measured data.
- The ***standard deviation*** *defines the width of the distribution, describing how* much variation occurs between successive measurements.

DSP CHP 1

[Statistics, Probability and Noise]

- This situation results in two general types of error that the system can experience.

1. The mean may be shifted from the true value. The amount of this shift is called the **accuracy** of the measurement.
2. Individual measurements may not agree well with each other, as indicated by the width of the distribution. This is called the **precision** of the measurement.

DSP CHP 1
[Statistics, Probability and Noise]

- Poor precision results from **random errors.**
- ***Precision*** *is a measure of random noise.*
- ***Accuracy*** is usually dependent on how you *calibrate the system.*
- Poor ***accuracy*** results from **systematic errors.**
- ***Accuracy*** *is a measure of calibration.*

DSP CHP 1

[Statistics, Probability and Noise]

- When deciding which name to call the problem, ask yourself two questions.

1. Will averaging successive readings provide a better measurement?

- yes, call the error precision
- no, call it accuracy.

2. Will calibration correct the error?

- yes, call it accuracy
- no, call it precision

DSP CHP 1
[Statistics, Probability and Noise]