Digital Signal Processing

Statistics, Probability and Noise

Part 2

Brief Review from Last Week

- Signal Domain
  - Time, Frequency and Space
- Characterization of signals using statistics
  - Mean, Variance and Standard Deviation
  - Variance is the power of the fluctuations around the mean
- Running Statistics
- Signal to Noise Ratio and Coefficient of Variation

Today's Topics

- Random Variables and Typical Error
- Adding Random Signals
- Process Stationarity
- Histograms – Histogram, PMF, PDF
- The Normal Distribution
- Precision and Accuracy
- Digital Noise Generation

Digital Signal Processing

Random Variables
and
Typical Error

Random Variables

- A random variable is a variable whose values depend on outcomes of a random phenomenon
  - We can describe the variable by its probabilities
  - Example: The output voltage from a sensor can be a random variable. It may consist of a DC value and random noise.

Random Variables

- The variable has a *true* mean, a *true* variance and a *true* standard deviation

- When we calculate the average, we are *estimating* the value of the *true* mean

- When we calculate the standard deviation, we are making an *estimate* of the *true* standard deviation

Estimates of the Mean and Standard Deviation

- When we estimate the mean, there may be an error between the estimate and the true mean

The "hat" $\widehat{\ }$ indicates an estimate

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\hat{\sigma} = \sqrt{\frac{1}{N-1}\sum_{i=0}^{N-1}(x_i - \hat{\mu})^2}$$

Estimates of the Variable Example

- A random variable can be described by its probabilities. Example:
  - Given a random variable with a <u>true</u> mean of 6 and a <u>true</u> standard deviation of 1
  - We can *<u>estimate</u>* the mean of the variable using a test statistic consisting of a set of N samples of the variable:

$$\hat{\mu} = \bar{x} = \frac{1}{N}\sum_{i=0}^{N-1} x_i$$

$\hat{\mu}$ is the *<u>estimate</u>* of the *<u>true</u>* mean

Estimates of the Variable Example

- The estimate of the mean $\hat{\mu}$ may be in error from the *<u>true mean</u>* μ
  - How much in error?
    - <u>The “typical” error </u>of the estimates is determined by:

$$\sigma_{estimate} = \frac{\sigma_{var}}{\sqrt{N}}$$

    - Sometimes called the standard deviation of the <u>estimates</u>
      - Different from the SD of the signal
      - Function of the true standard deviation and the number of samples used to estimate the mean

Typical Error

- The "typical error" of the estimate is a function of the true standard deviation of the variable and the number of samples used in making the estimate.

$$Typical\ Error = \sigma_{estimate} = \frac{\sigma}{\sqrt{N}}$$

- The "typical error" of the estimate decreases by the square root of the number of samples

Law of Large Numbers

- The Law of Large Numbers says that as N approaches infinity, the typical error approaches 0

$$Typical\ Error = \frac{\sigma}{\sqrt{N}} \rightarrow 0 \text{ for large N}$$

- Why is this important?
  - We can control the amount of error in the estimate by selecting the number of samples used in the estimate

In Class Problem
Typical Error

- A variable has a standard deviation of $\sigma = .15$
  - How many samples N, do I need to use in the estimate of the mean, to have a typical error of the estimate equal to .01?

Digital Signal Processing

Adding Random Signals

Adding Random Variables

- The mean of the sum of two random variables will be the sum of the means

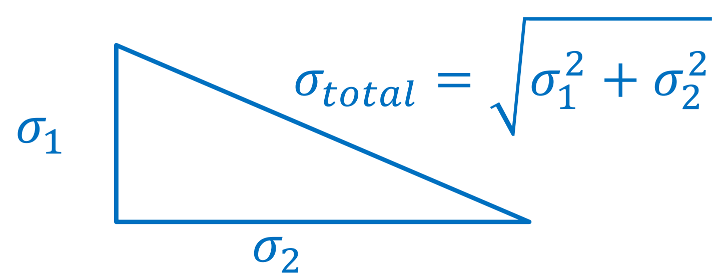
$$\mu_{total} = \mu_1 + \mu_2$$

- If two random variables are added, their variances add:

$$\sigma_{total}^2 = \sigma_1^2 + \sigma_2^2$$

- The standard deviation of the combined signal is:

$$\sigma_{total} = \sqrt{\sigma_1^2 + \sigma_2^2}$$



In Class Problem
Adding Two Signals with Noise

- Given two signals with the following statistics:

Signal 1 $\mu_1 = 2, \sigma_1 = .5$

Signal 2 $\mu_2 = 1, \sigma_2 = .125$

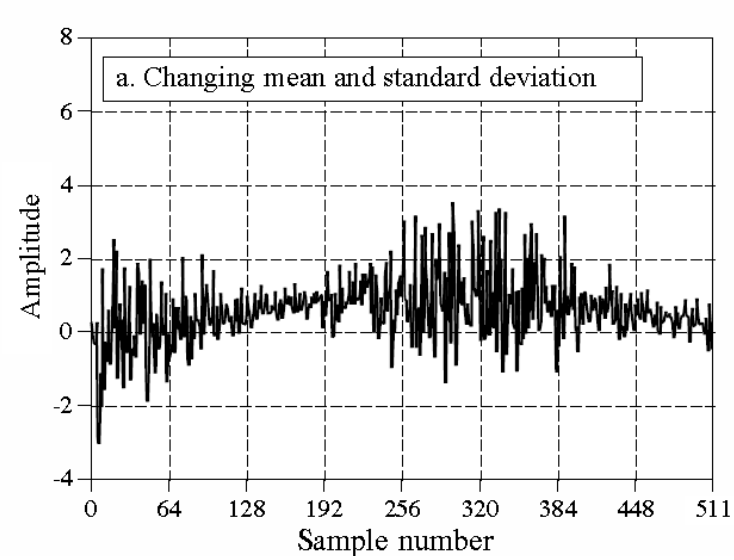
  - Compute their individual SNR’s
  - Compute the <u>SNR of the sum</u> of the 2 signals

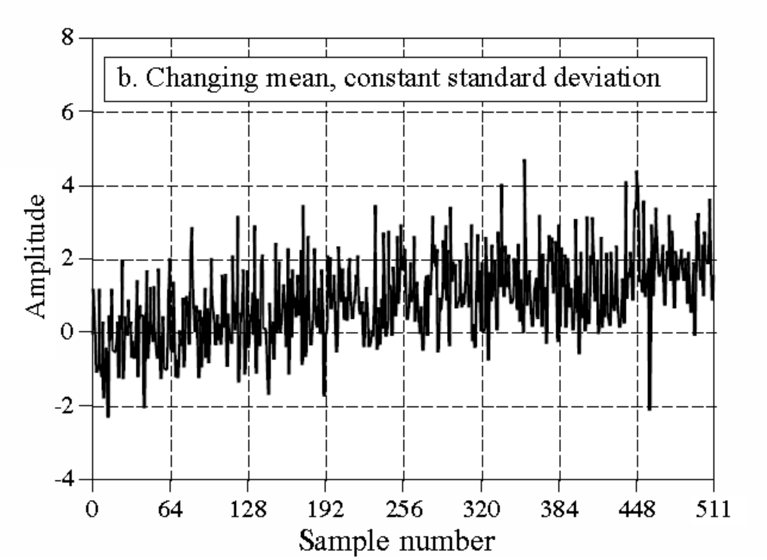
Digital Signal Processing

Process Stationarity

Non-stationary Processes

- If the underlying process changes over time, the process is said to be <u>non-stationary</u>
- Examples:



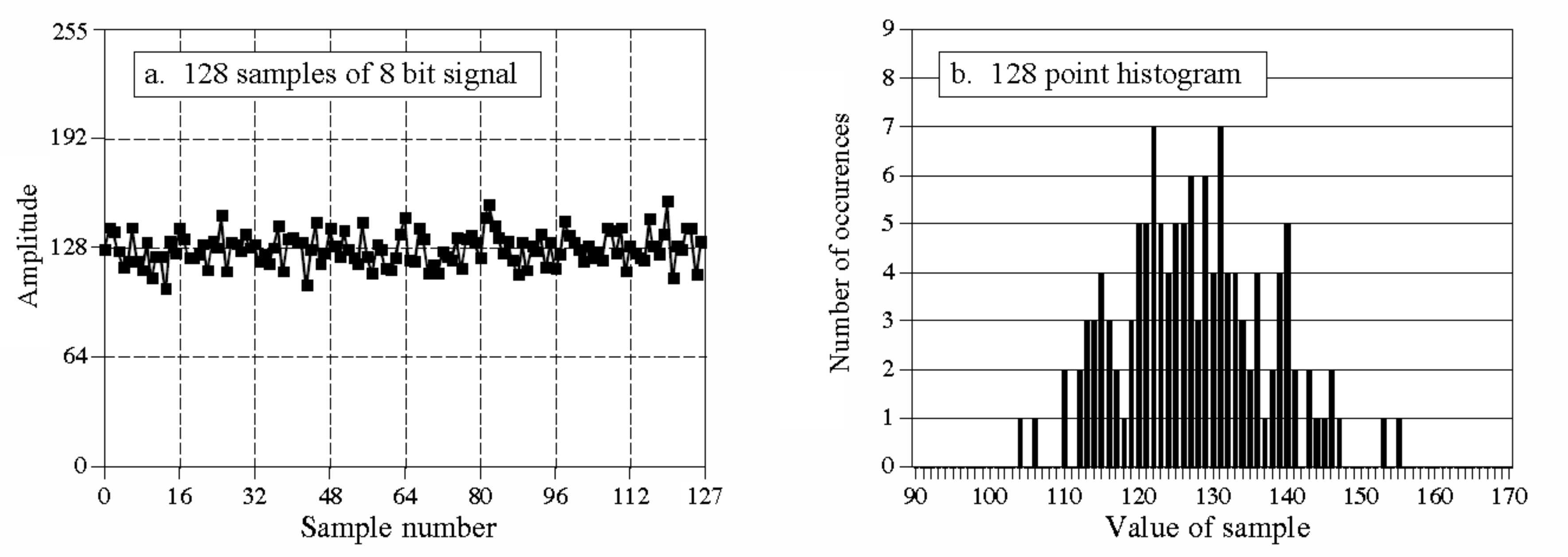


Digital Signal Processing

Histograms

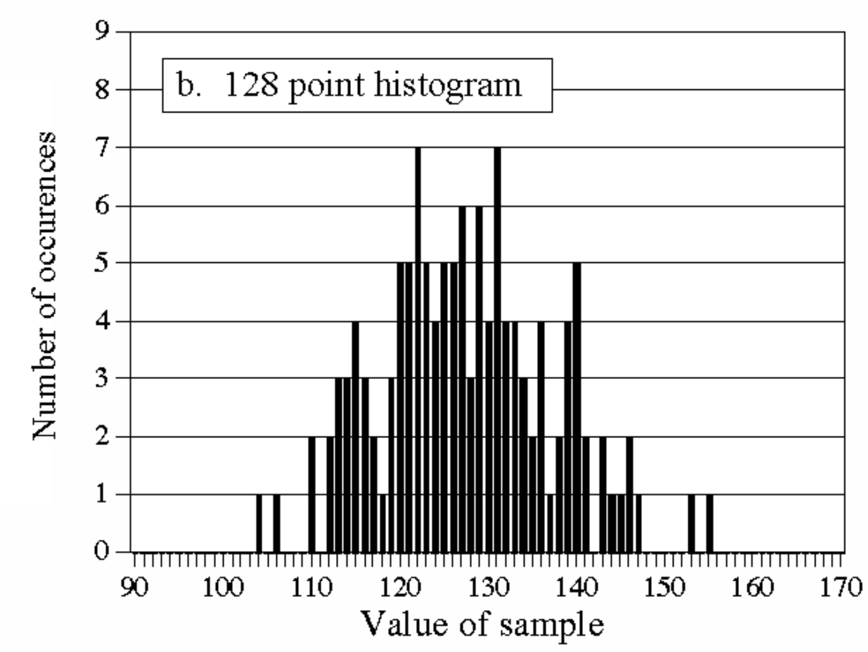
Histograms

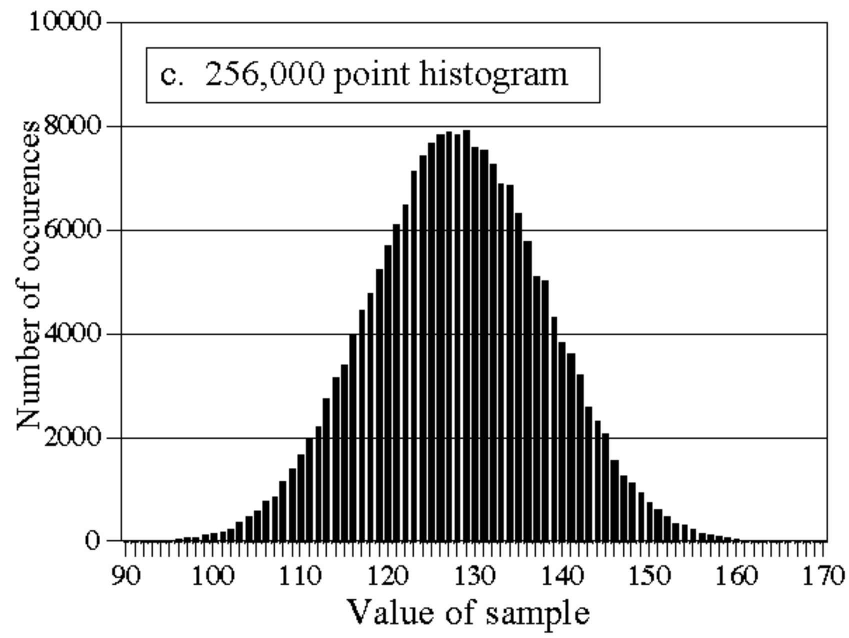
- Describes the number of samples in the data set that have a given value or range of values.
  - Example: Samples from an 8-bit A/D converter



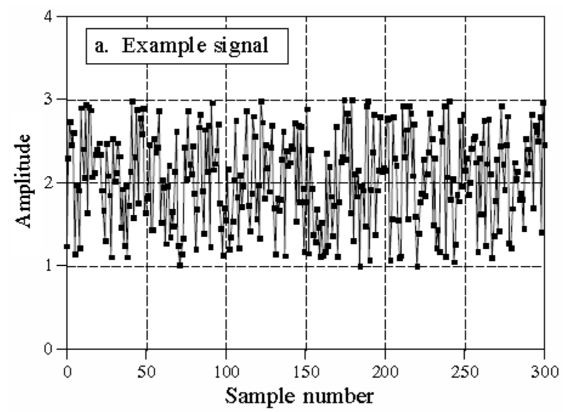
Histogram With More Data Samples

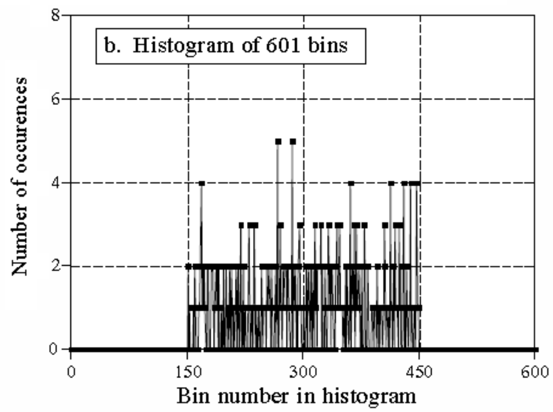
- Increasing the number of samples reveals the underlying distribution:



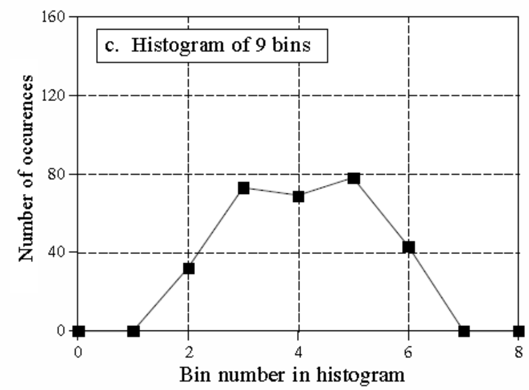


Selection of the Number of Bins





Number of bins is too large
Poor vertical resolution



Number of bins is too small
Poor horizontal resolution

Selection of the Number of Bins

- A rule of thumb is to use between 5 and 20 bins
    - Another approach is to use Sturge's rule*

$$K = 1 + \log_2 N$$

Where:
N is the number of samples
K is the number of bins

https://www.statology.org/sturges-rule/

Estimating the Mean and Variance from the Histogram

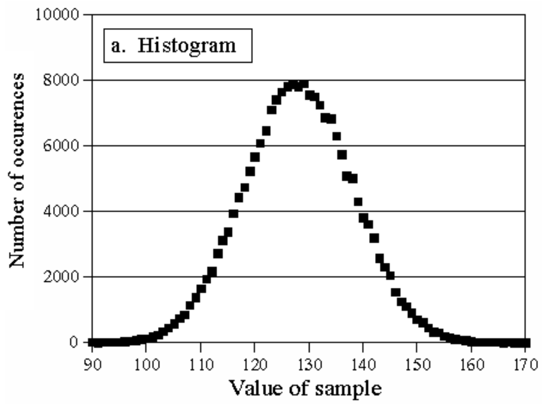
- One can estimate the mean and variance from parameters of the histogram
- H_i is the number samples in the i^{th} bin

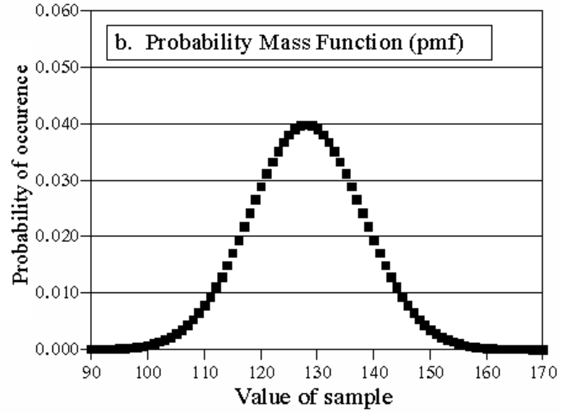
$$N = \sum_{i=0}^{M-1} H_i$$ Total samples N

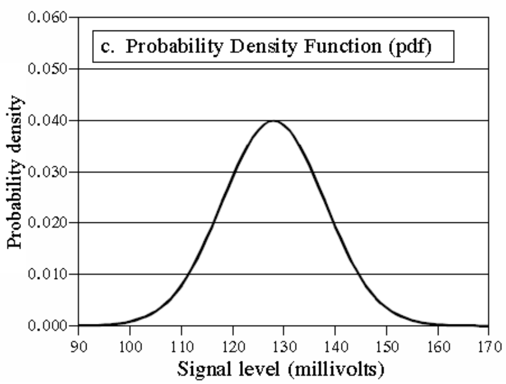
$$\hat{\mu} = \frac{1}{N} \sum_{i=0}^{M-1} i H_i$$

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

Probability Mass Function Vs Histogram







Histogram:

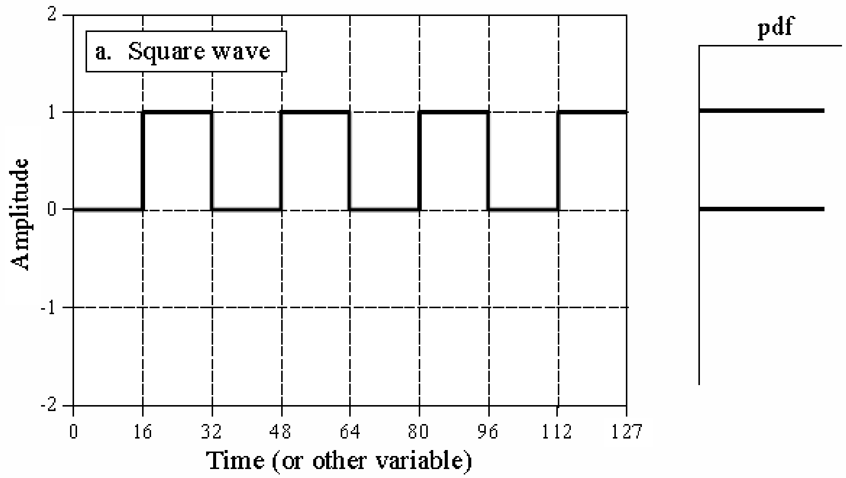
Formed from a finite number of samples of a signal – a statistical estimate of the underlying probability

Probability Mass Function :

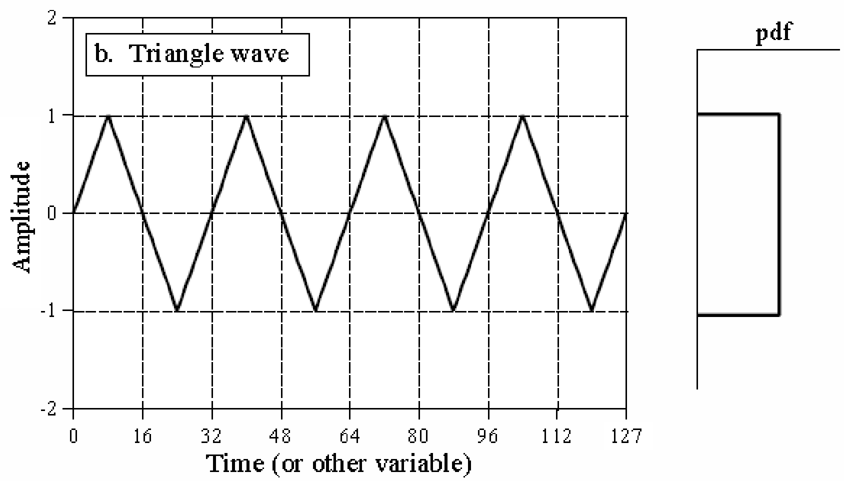
The underlying probability for a signal that takes on discrete values.

Probability Density Function:
The underlying probably for a signal that is a continuous function

PDF For Square and Triangle Waves



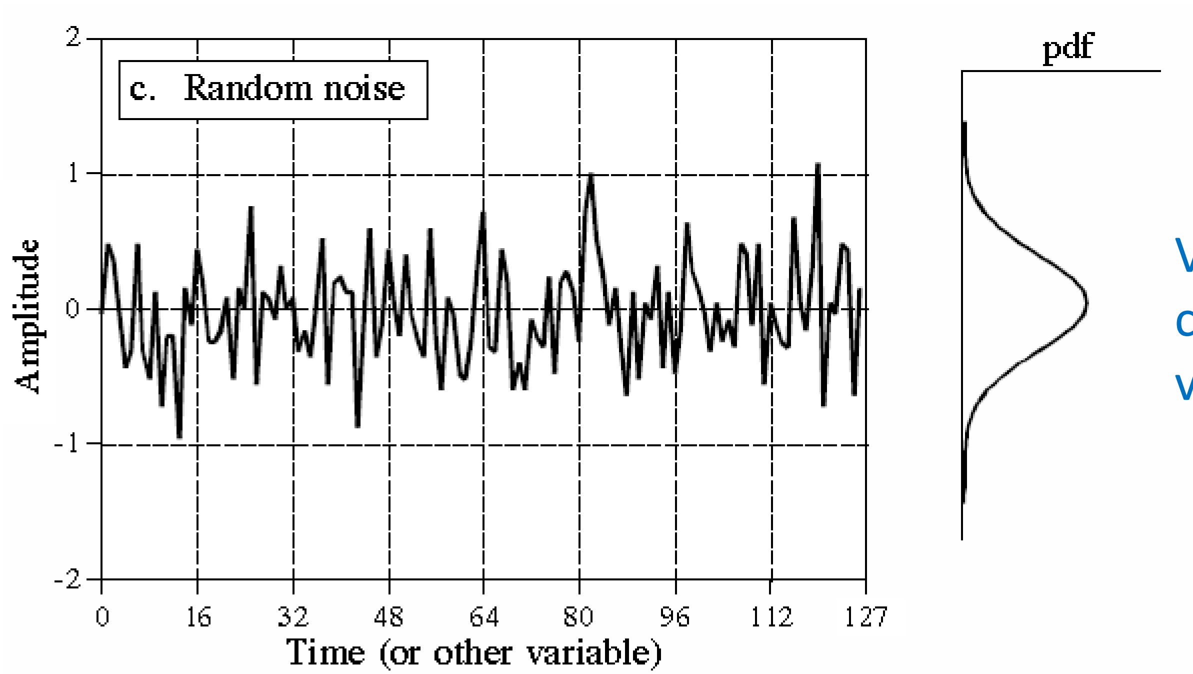
Two discrete values with equal probability



Equal probability for all values over a range

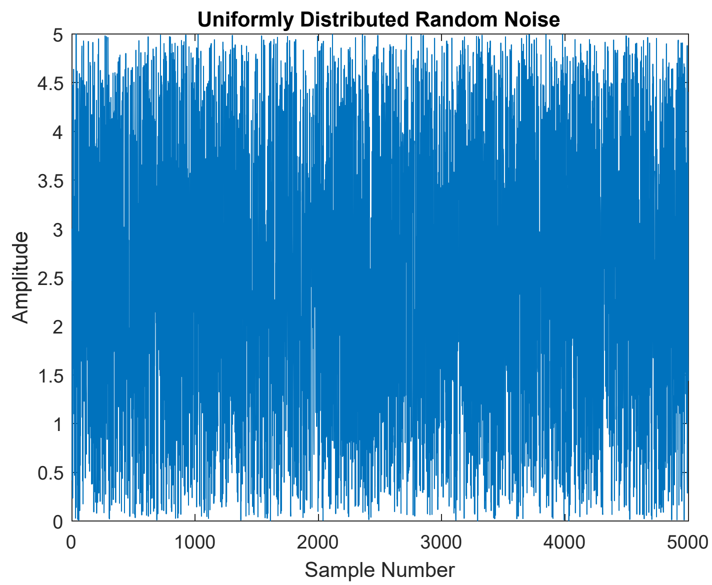
PDF for Random Noise

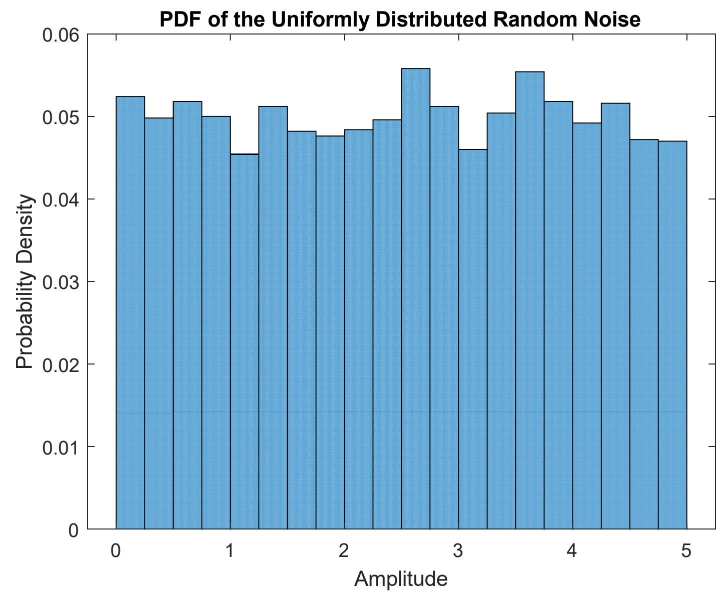
- PDF for this signal is a *normal* distribution



Varying probability density centered on the mean value

PDF for Uniform Random Noise





The PDF is "flat" across the amplitude range

- PDF for this signal is a *uniform* distribution

Digital Signal Processing

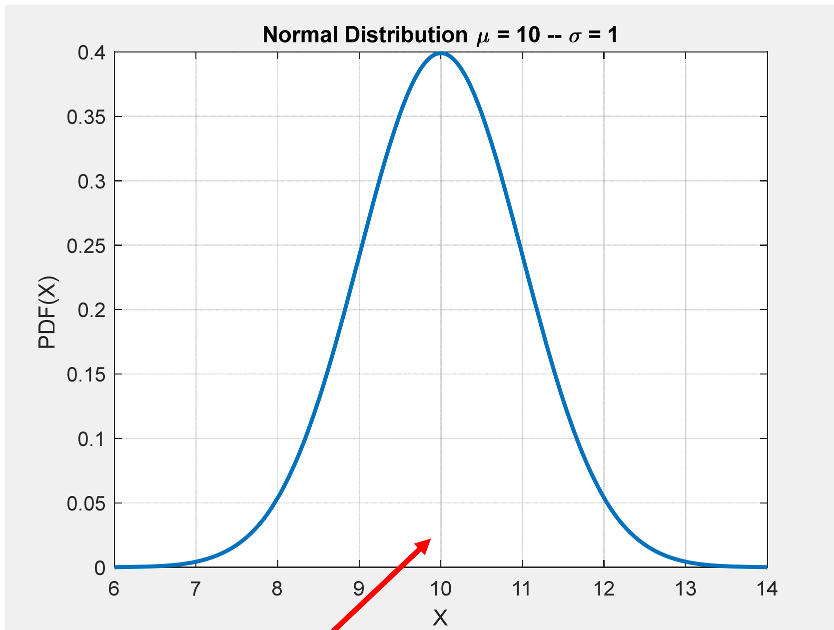
The Normal Distribution

Normal Distribution Example

- Many random signals found in nature have a ***normal distribution***

$$P(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\mu = 10$$
$$\sigma = 1$$



Area under the curve = 1

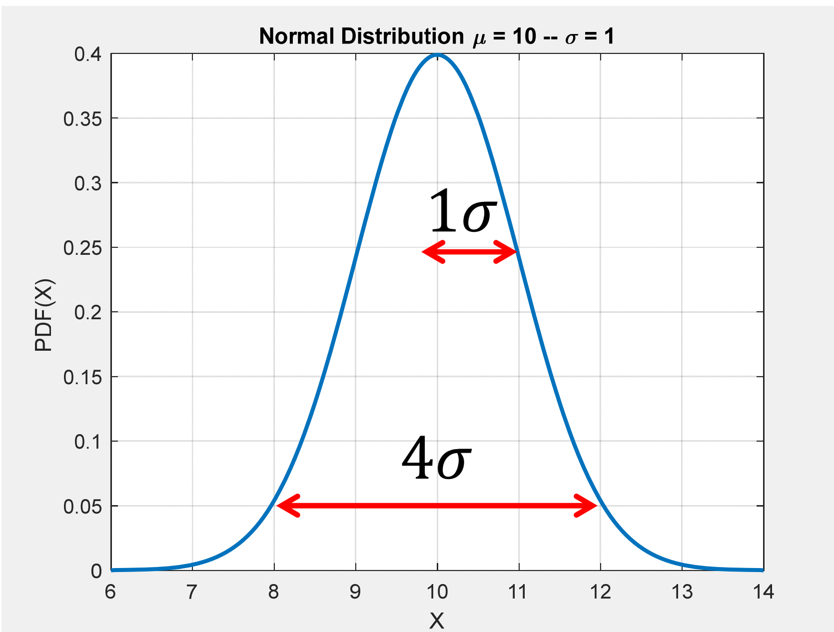
68.3% of values fall within $\pm 1\sigma$.

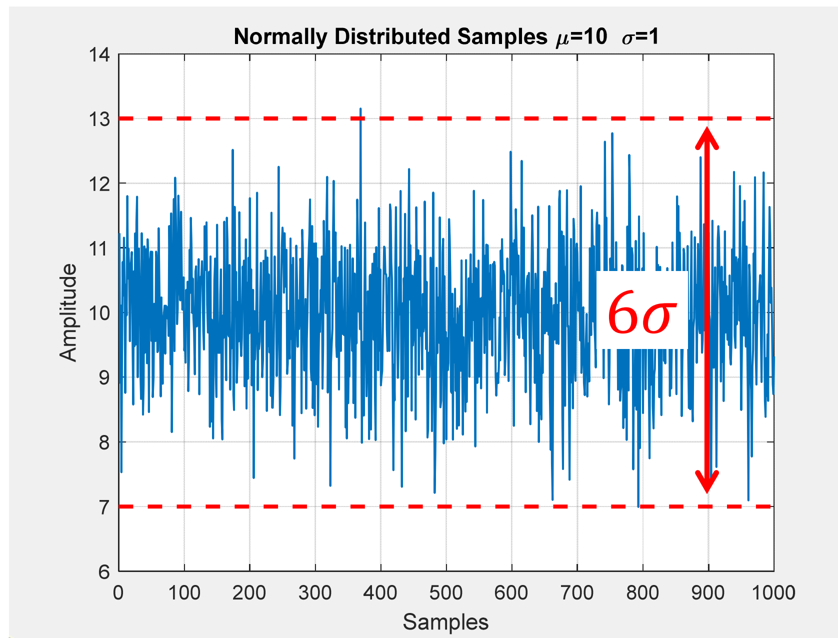
95.5% of values fall within $\pm 2\sigma$

Centered at $\mu = 10$

Characteristics of the Normal (or Gaussian) Distribution

- The likelihood of values far from the mean, e.g. 4 sigma away from the mean, is very low.
- This is why the signal appears to have a bounded peak to peak value of 6-8 times sigma $(\pm 3\sigma\ to\ \pm 4\sigma)$





Digital Signal Processing

The Central Limit Theorem

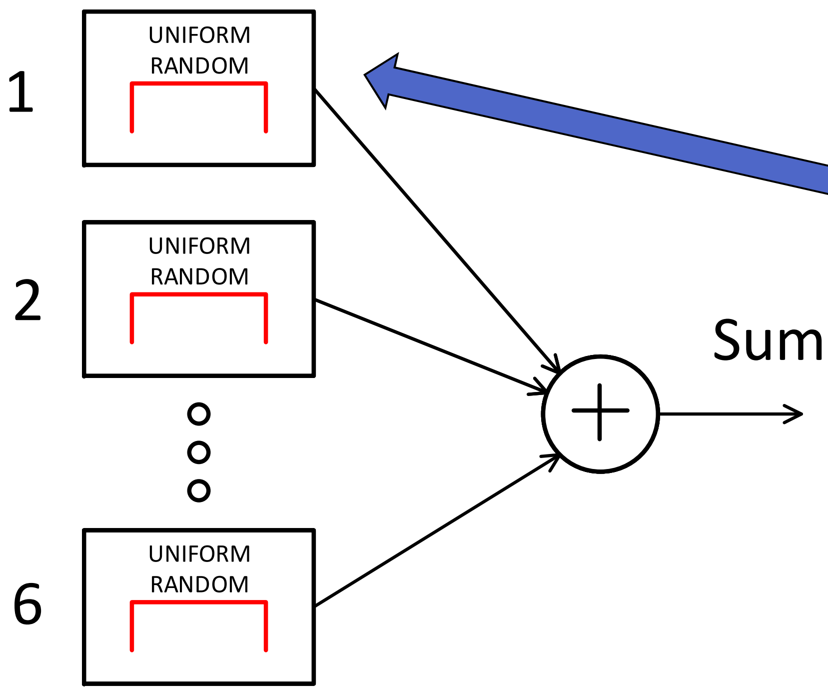
The Central Limit Theorem

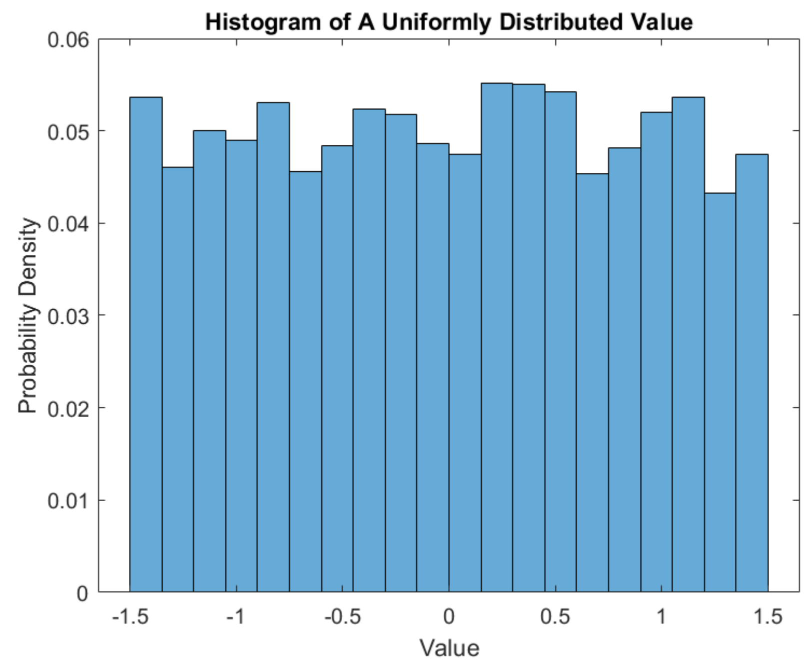
- The sum of random variables becomes normally distributed as more and more random variables are added together.

- True even if the random numbers being added together are from different probability distributions.

Central Limit Theorem MATLAB Example

- Generate 6 uniformly distributed random numbers and add them. What is the distribution of the sum?

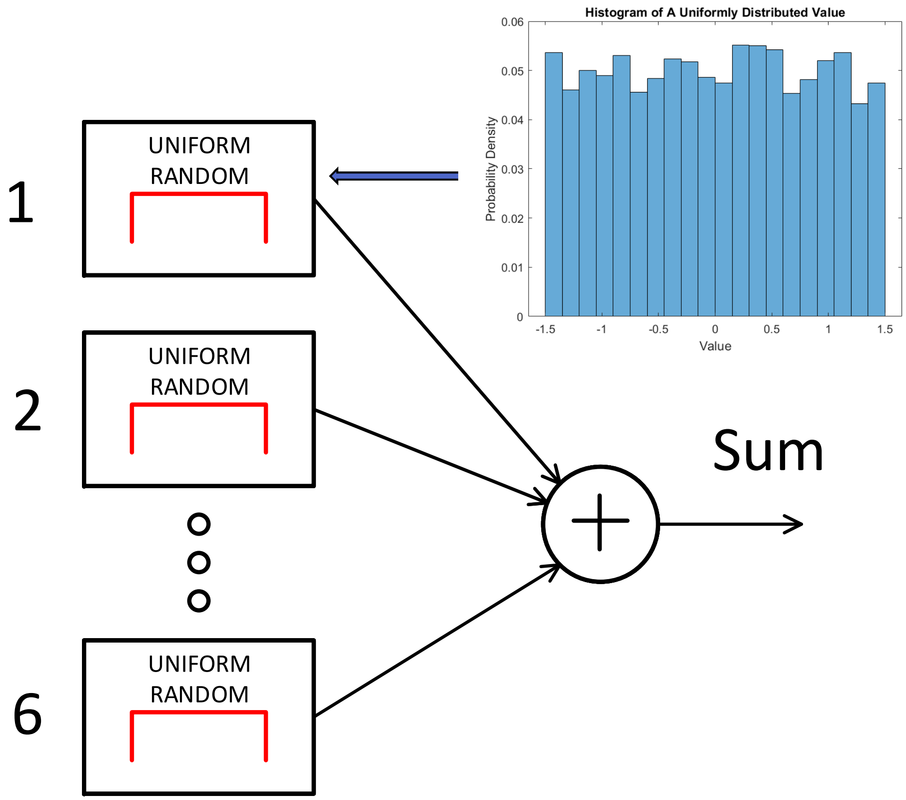


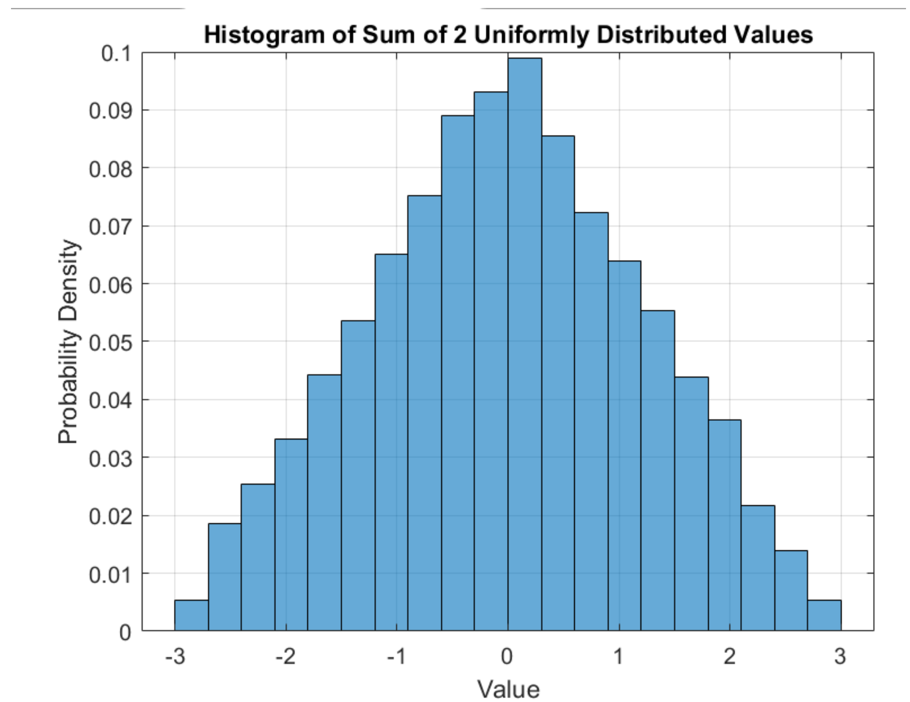


Distribution of 1 Uniform Random Variable

Central Limit Theorem MATLAB Example

- The sum of two uniform random variables starts to look somewhat like a normal distribution

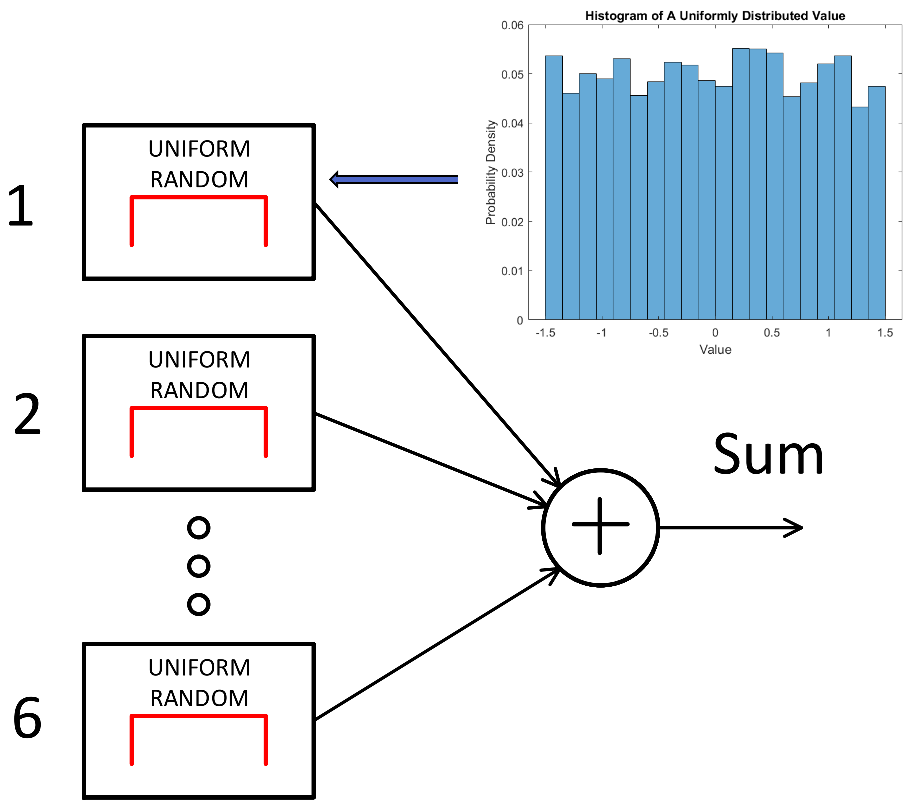


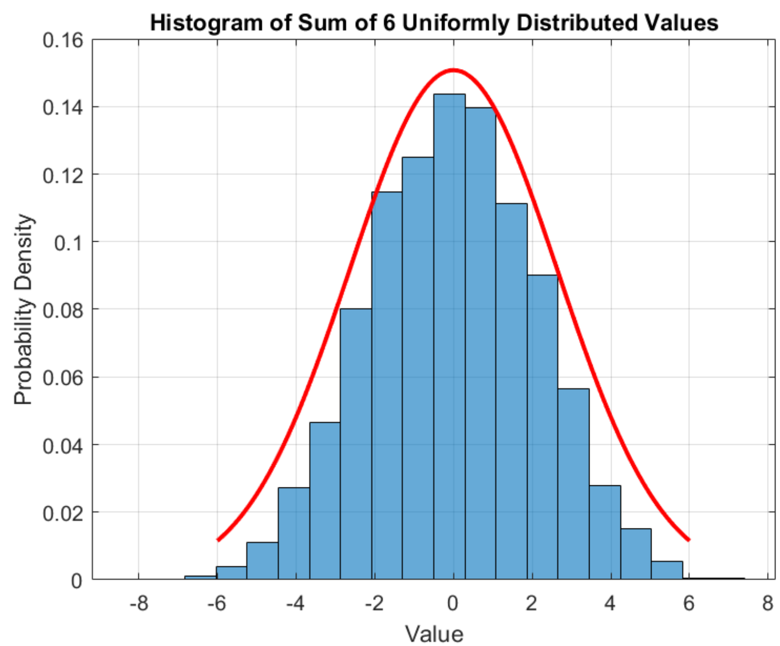


Distribution of the sum of
2 Uniform Random Variables

Central Limit Theorem MATLAB Example

- The distribution of 6 uniform random variables looks very much like a normal distribution





Distribution of the sum of
6 Uniform Random Variables

Digital Signal Processing

Precision and Accuracy

Precision and Accuracy

- Accuracy
  - A measure of how close the estimated mean $\hat{\mu}$ is compared to the true mean:

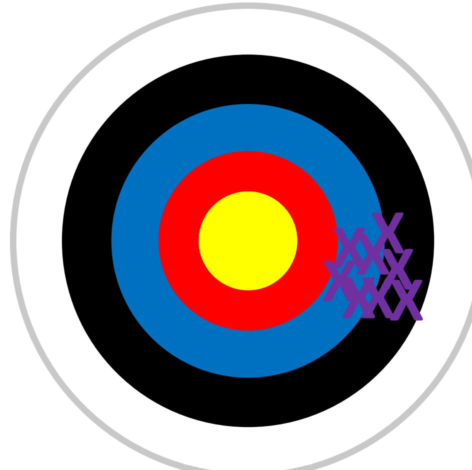
$$Accuracy = \hat{\mu} - \mu$$

- Precision
  - A measure of how well the individual measurements or samples compare with each other:
    - Expressed by the Signal to Noise Ratio (SNR) or by the Coefficient of Variation (CV)

$$SNR = \frac{\hat{\mu}^2}{\hat{\sigma}^2} \qquad CV = \frac{\hat{\sigma}}{\hat{\mu}} \times 100$$

Precision and Accuracy

Neither Accurate
Nor Precise



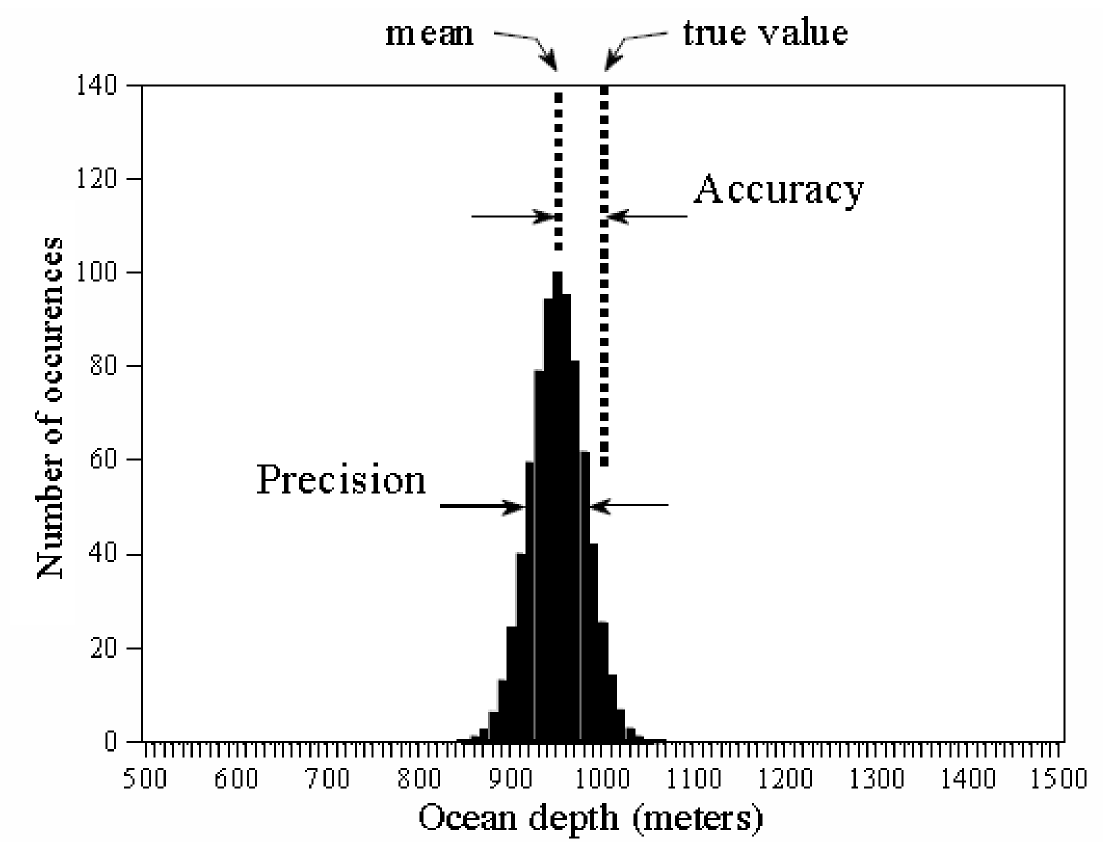
Not Accurate
But Precise

Accurate but
not Precise

Accurate
And Precise

Precision and Accuracy

- Example: For a normally distributed signal



Digital Signal Processing

Digital Noise Generation

Digital Noise Generation

- Generating random noise is helpful for testing how DSP algorithms operate in the presence of noise
- Most programming languages can produce uniformly distributed random numbers
  - MATLAB -> “**rand**” function
- By adding uniformly distributed random numbers, you can create normally (Gaussian) distributed random numbers

Digital Signal Processing

Summary

Summary of Today

- Random Variables and Typical Error
  - The typical error of an estimate of μ is a function of the true σ and the number of samples N
- Adding Random Signals
  - The mean of two signals add algebraically
  - The SD of two signals add in quadrature
- Histograms can help estimate the PMF or PDF
  - PMF is for discrete signals, PDF for continuous

Summary of Today

- The Normal Distribution
  - Many random variables are normally distributed
  - Use the CDF to compute probabilities (text)
- Precision and Accuracy
  - Precision is related to standard deviation and can be express in terms of SNR as well.
- Digital Noise Generation
  - Uniform random variables can be combined according the Central Limit Theorem to produce a normally distributed random variable.