Digital Signal Processing

Statistics, Probability and Noise

Part 2

Reminders

- Pre-Lecture Quizzes are due before lecture. See MyCourses for Due Dates
- Homework 01
  - Complete the set of practice problems
  - Compare to solutions as a guide to help your understanding
  - Complete the Homework 01 Quiz to receive homework credit and a grade

Last Class Review

- 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
- Computing Running Variance
- 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

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 some noise.

Random Variables

- The variable has a *true* mean and a *true* variance and 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" $\hat{}$ indicates it is 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
  - 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 using 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

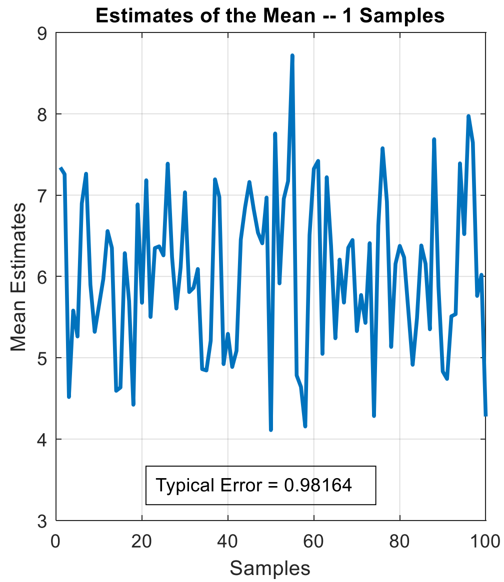
- The estimate of the mean $\hat{\mu}$ may be in error from the *true mean* μ

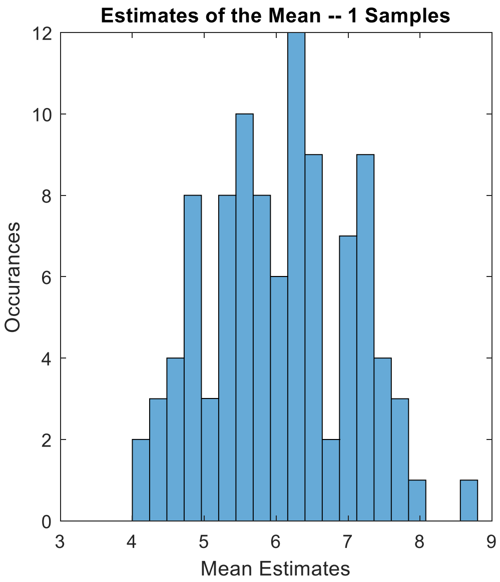
- How much in error?

- We say the "typical" error of the estimates is the standard deviation of the estimates
  - Different from the SD of the signal
  - Function of the true standard deviation and the number of samples in estimating the mean

Typical Error Estimate Example

- Trivial example -- A process with a true mean of 6
- Take one value from the variable and use this as an estimate of the mean
- Repeat this 100 times so that we have 100 estimates of the mean.
- Plot the estimates of the mean of the variable (using 1 sample of the variable)





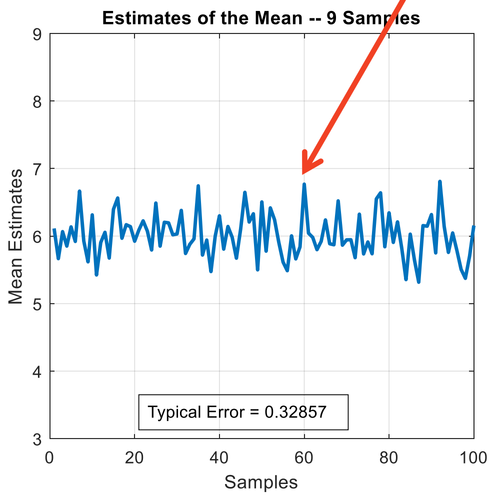
The <u>estimates</u> of the mean are not exactly the <u>true</u> mean value.

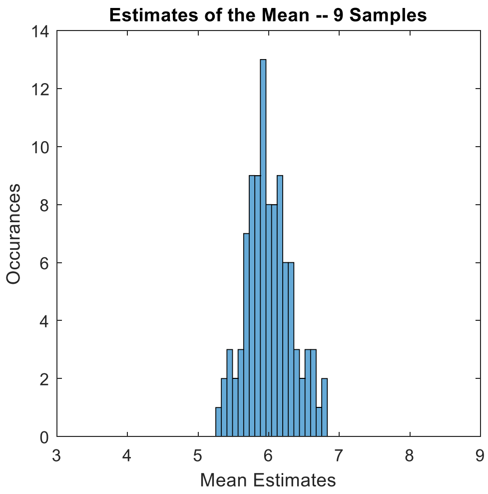
The typical error of the estimates is standard deviation of the <u>estimates</u> or 0.982.

Typical Error Estimate Example

- Increase the number of samples used to make each estimate to 9
- Plot the estimates of the true mean and compare

Each point is the mean of 9 samples





The <u>estimates</u> are now closer to the <u>true</u> mean of 6.

The typical error is now 0.329

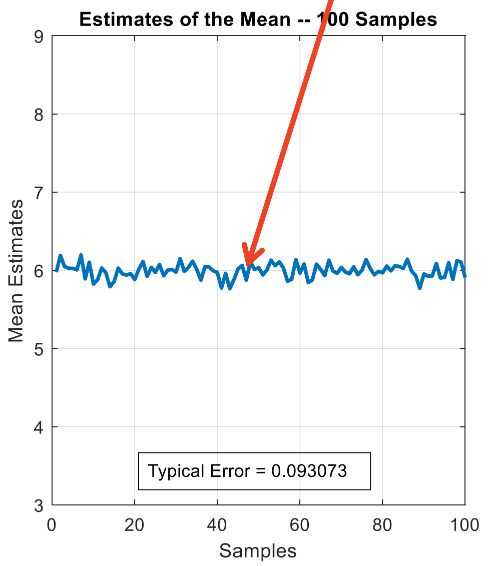
$$\sigma_{estimate} = \frac{\sigma_{var}}{\sqrt{N}}$$

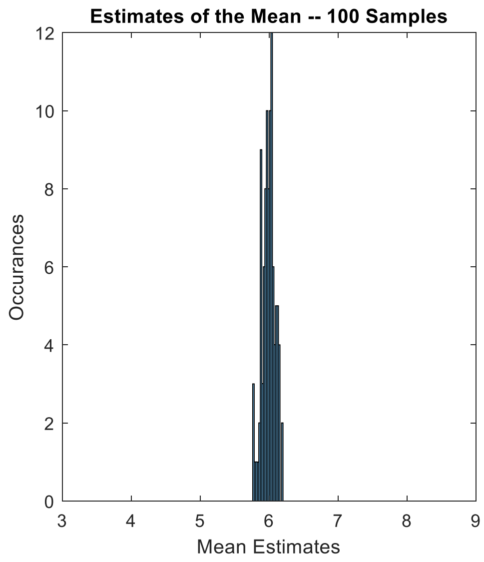
$$\sigma_{estimate} = \frac{1}{\sqrt{9}} = \frac{1}{3}$$

Typical Error Estimate Example

- Increase the number of samples used to make each estimate to 100
- Plot the estimates of the true mean and compare

Each point is the mean of 100 samples





The estimates are now even closer to the true mean of 6.

The typical error is now $\approx \frac{1}{10}$

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

$$\sigma_{estimate} = \frac{1}{\sqrt{100}} = \frac{1}{10}$$

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 .01?

In Class Problem
Typical Error

- A variable has a standard deviation of $\sigma = .15$
- How many samples do I need to average to estimate the mean to within a typical error of .01?

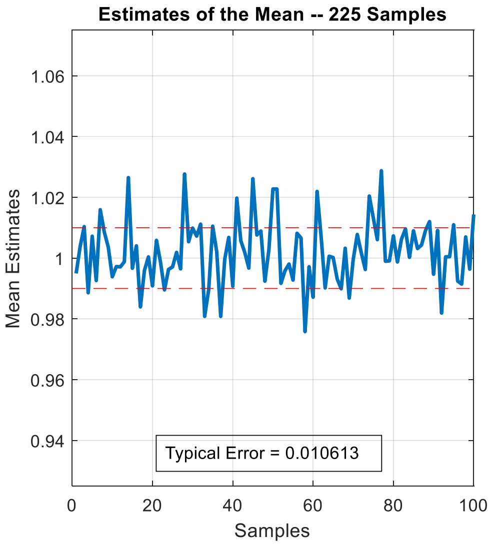
$$Typical\ Error = \sigma_{estimate} = \frac{\sigma_{process}}{\sqrt{N}}$$

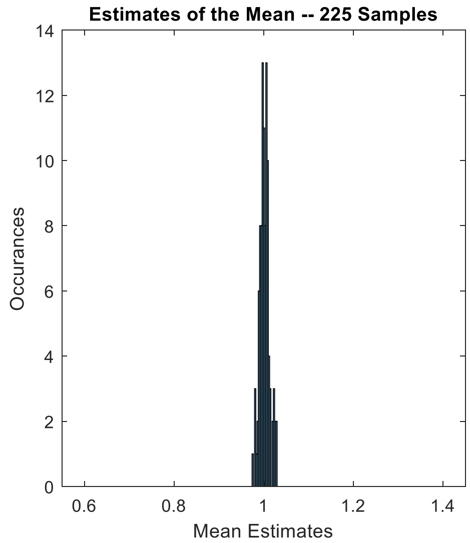
Solving for N

$$N = \left(\frac{\sigma}{.01}\right)^2 = \left(\frac{.15}{.01}\right)^2 = 225 \text{ samples}$$

Example
Typical Error

- Create 225 samples from a variable with $\mu = 1$ and $\sigma = .15$. Estimate μ using the average.
- Repeat 100 times and plot the estimates





Some estimates are outside of typical error window.

Most ($\approx$ 67%) are within the error window

$$Typical\ Error = \frac{\sigma_{var}}{\sqrt{N}} = \frac{0.15}{\sqrt{225}} = .01$$

NOPRINT

Adding Random Signals

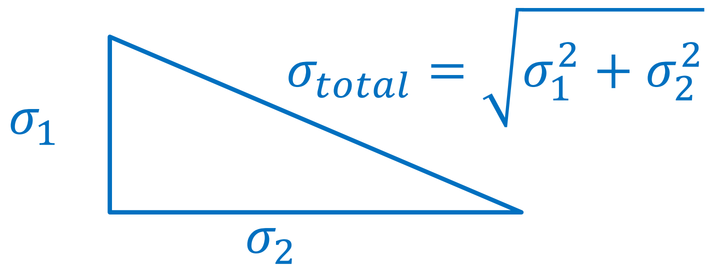
The Sum of Random Signals

- If two random signals are added their variances add:

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

- The standard deviation of the combined signal is:

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



- The standard deviation is said to add "in quadrature" as in the sides of a right triangle

Adding Two Signals with Noise

- The mean of the sum of two signals will be the sum of the means

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

- The standard deviation of the noise of the two signals will add in quadrature

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

In Class Problem
Adding Two Signals with Noise

- Assume two signals with the following probabilities:

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

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

- Compute their individual SNR's
- Compute the SNR of the sum of the 2 signals

In Class Problem
Adding Two Signals with Noise

- Assume two signals with the following probabilities:

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

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

- Compute their individual SNR's

$$SNR_1 = \frac{\mu_1}{\sigma_1} = {}^{2}/_{.5} = 4$$

$$SNR_2 = \frac{\mu_2}{\sigma_2} = {}^{1}/_{.125} = 8$$

In Class Problem
Adding Two Signals with Noise

- Compute the SNR of the sum of the 2 signals

Signal 1 $\mu_1 = 2, \sigma_1 = .5$ $SNR_1 = {}^{2}/_{.5} = 4$

Signal 2 $\mu_2 = 1, \sigma_2 = .125$ $SNR_2 = {}^{1}/_{.125} = 8$

- Then:

$\mu_{total} = \mu_1 + \mu_2 = 3$ $\sigma_{total} = \sqrt{\sigma_1^2 + \sigma_2^2} = .5154$

$$SNR_{total} = \frac{3}{.5153} = 5.821$$

NOPRINT

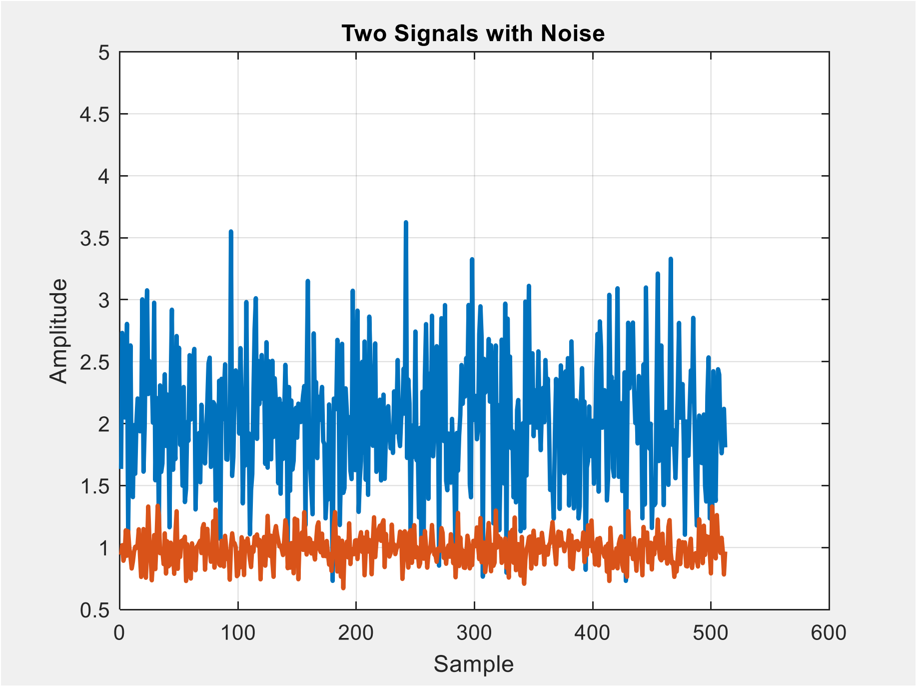
In Class Problem
Adding Two Signals with Noise

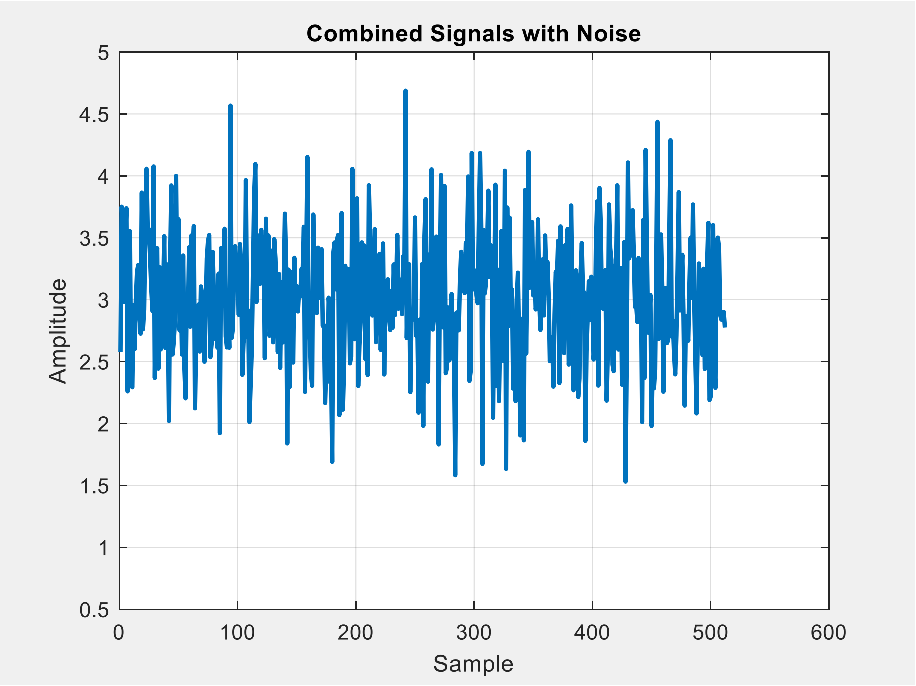
$$\mu_1 = 2, \sigma_1 = .5$$

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

$$\mu_{total} = \mu_1 + \mu_2 = 3$$

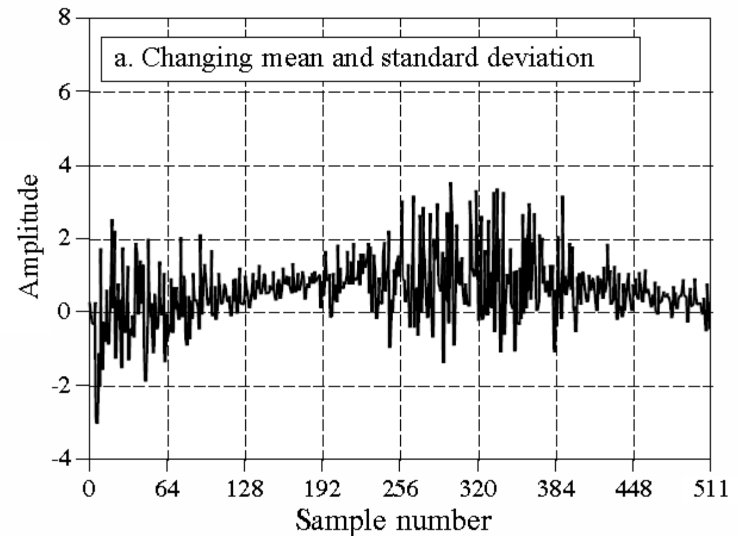
$$\sigma_{total} = \sqrt{\sigma_1^2 + \sigma_2^2} = .5154$$

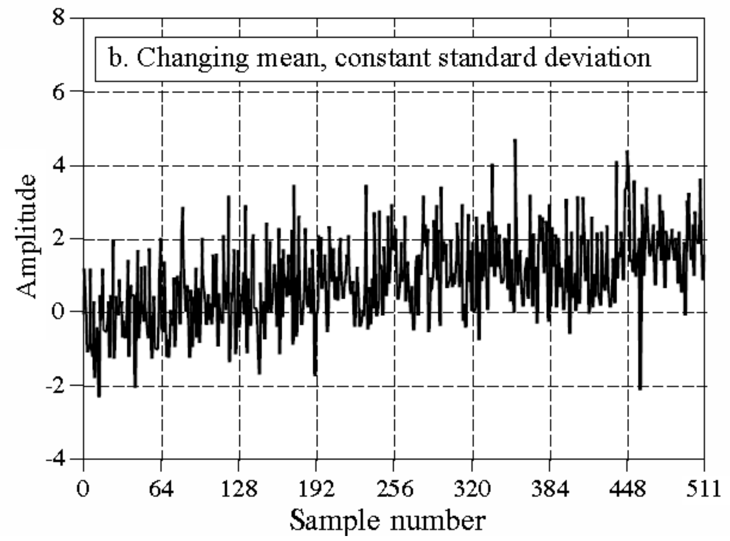




Process Stationarity

Non-stationary Processes



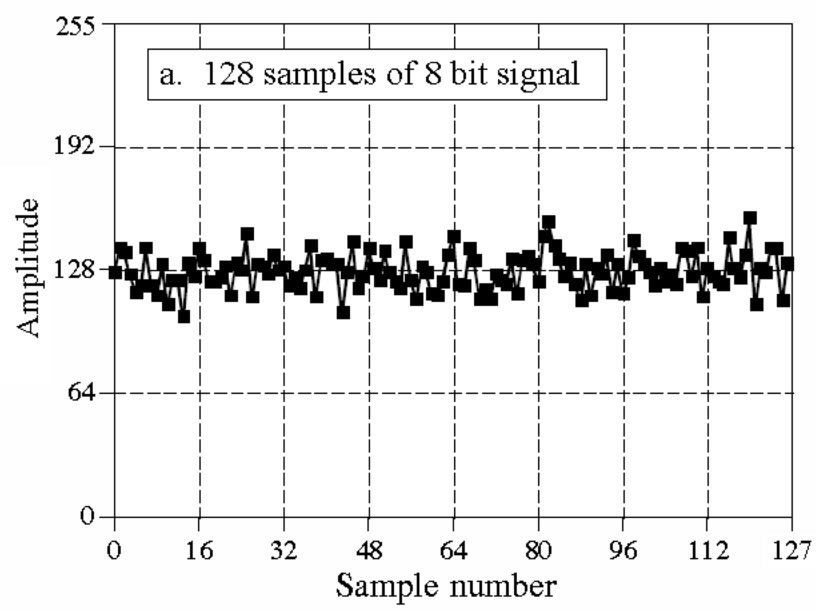


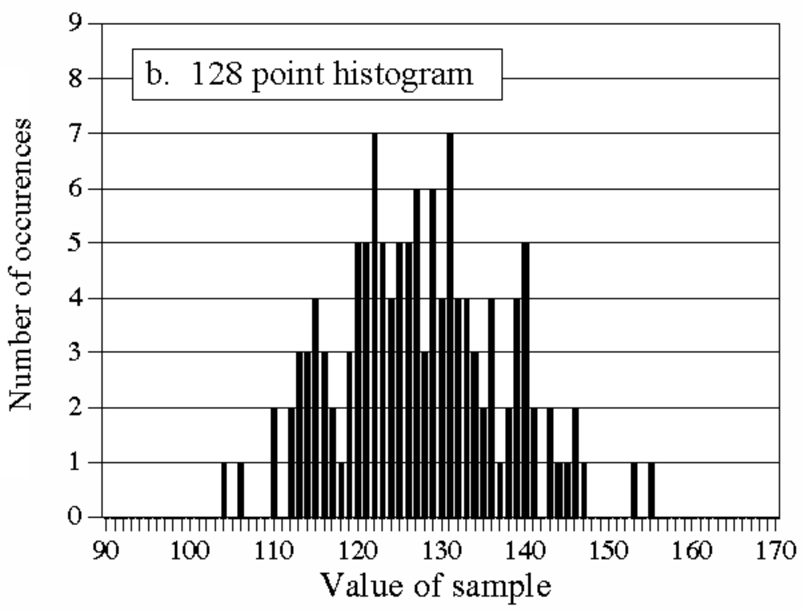
- If the underlying probabilities change over time the process is said to be <u>non-stationary</u>
- What is the impact of using short segments (fewer samples) on the typical error of the mean and standard deviation?

Histograms

Histograms

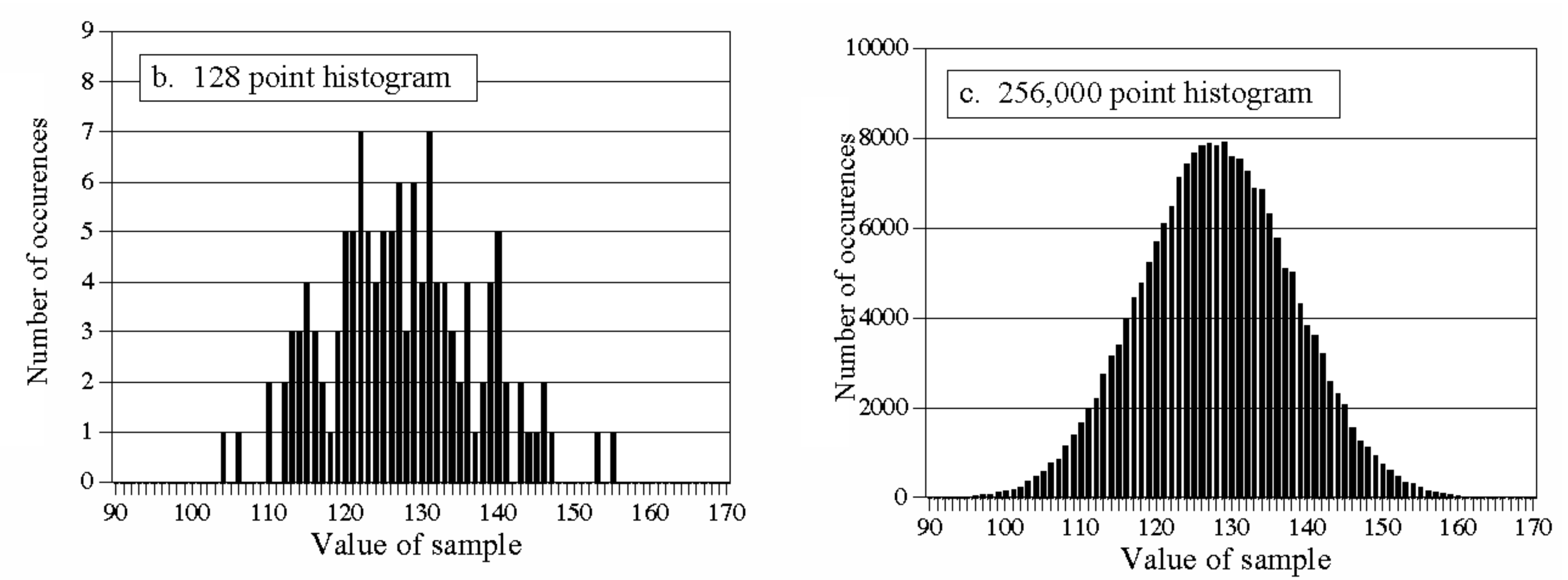
- Describes the number of samples in the data set that have the given value.
- Example: Samples from an 8-bit A/D converter



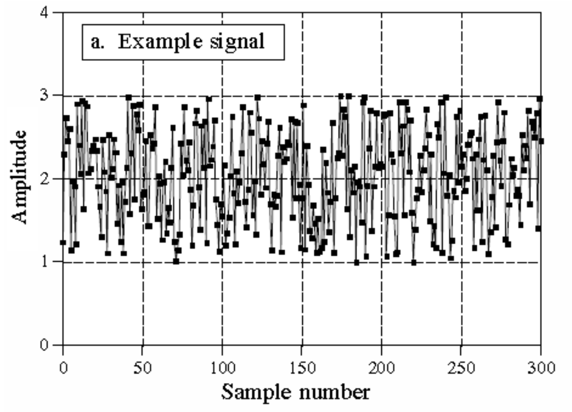


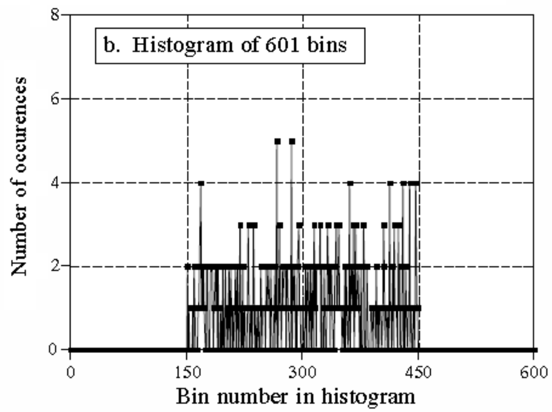
Histogram With More Data Samples

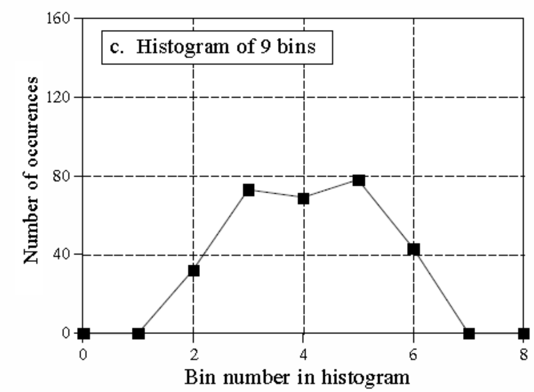
- Increasing the number of samples reveals the underlying distribution, smoother.



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*

$$K = 1 + 3.322\log_e N$$

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

* https://www.statisticshowto.datasciencecentral.com/choose-bin-sizes-statistics/

Computing Mean and Variance from Histogram

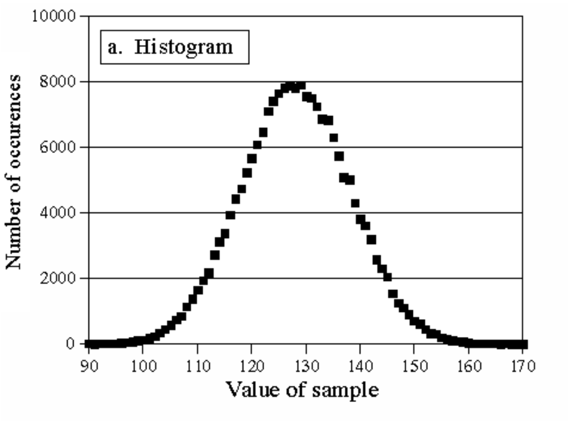
- One can estimate the mean and variance from parameters of the histogram
- H_i is the number samples in the i^{th} bin

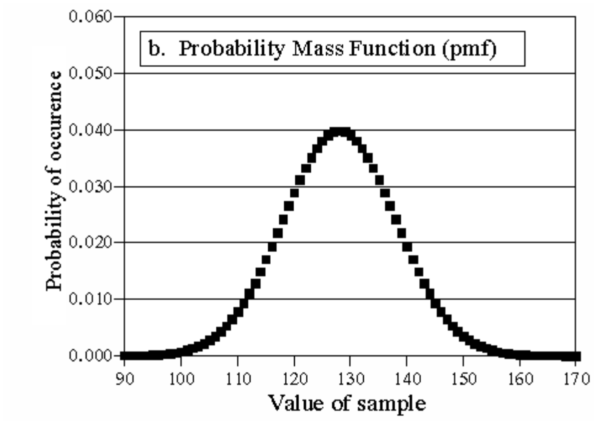
$$N = \sum_{i=0}^{M-1} H_i$$ Total samples N

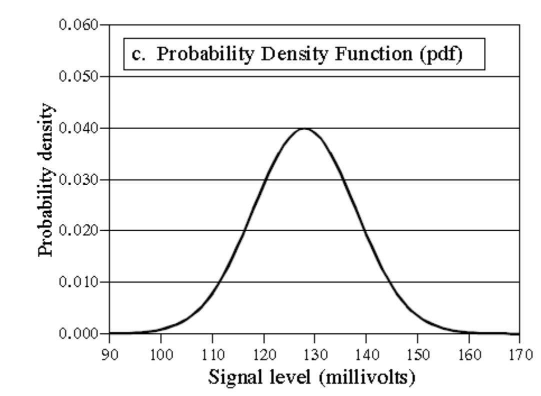
$$\hat{\mu} = \frac{1}{N} \sum_{i=0}^{M-1} i H_i$$

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

Probability Mass Function Vs Histogram







Histogram:

Based on finite number of samples – a statistical estimate of the underlying probability

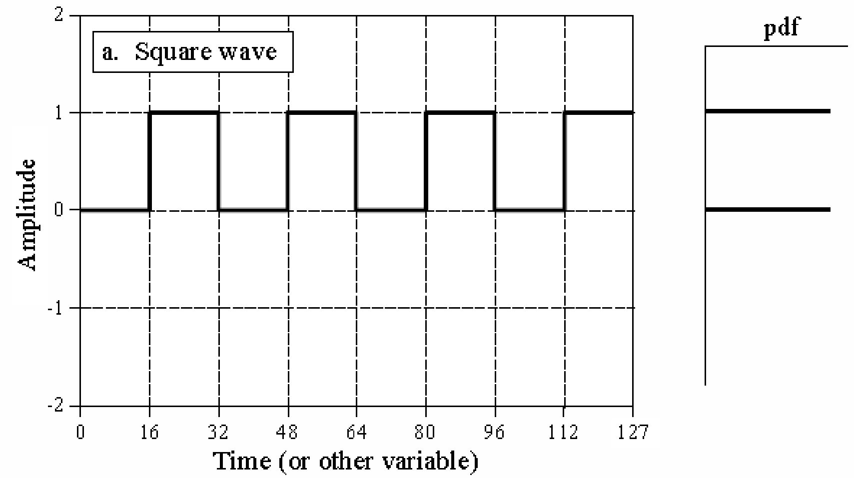
Probability Mass Function :

The underlying probability for a signal that takes on discrete

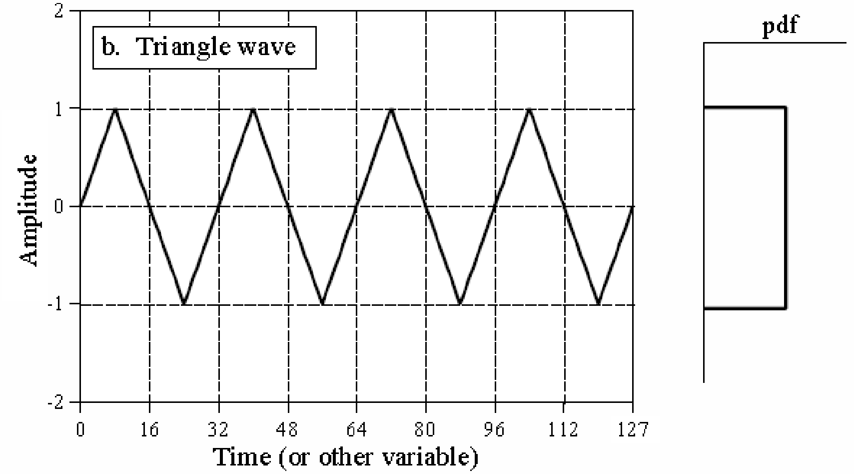
Probability Density Function:

The underlying probably for signal that is a continuous function

PDF For Square and Triangle Waves



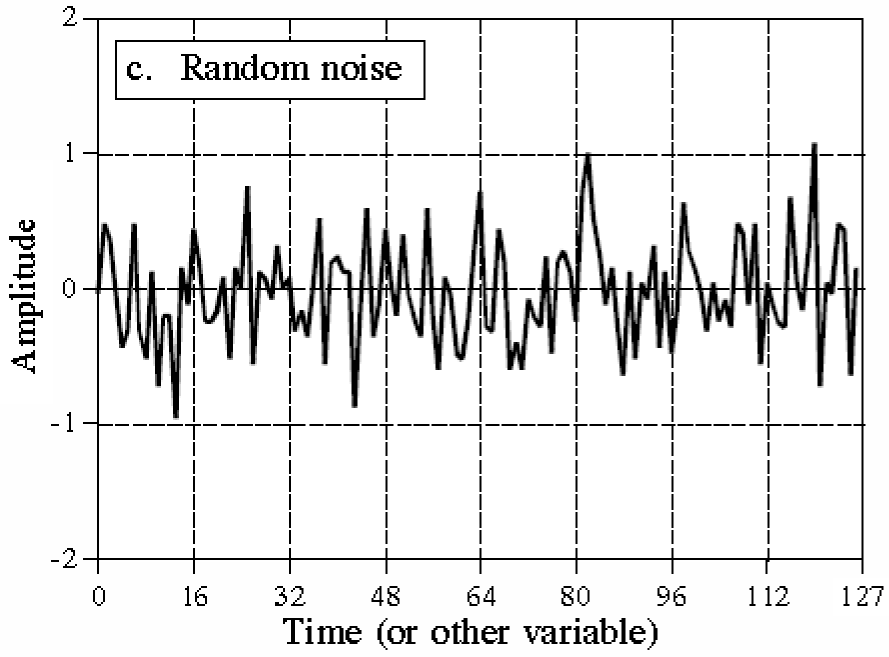
Two discrete values with equal probability

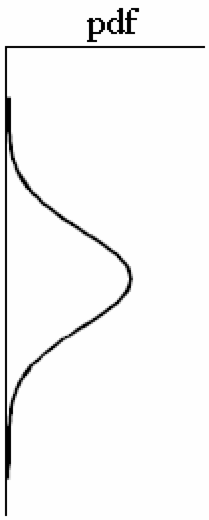


Equal probability for all values over a range

PDF for Random Noise

- PDF for this signal is a *normal* distribution

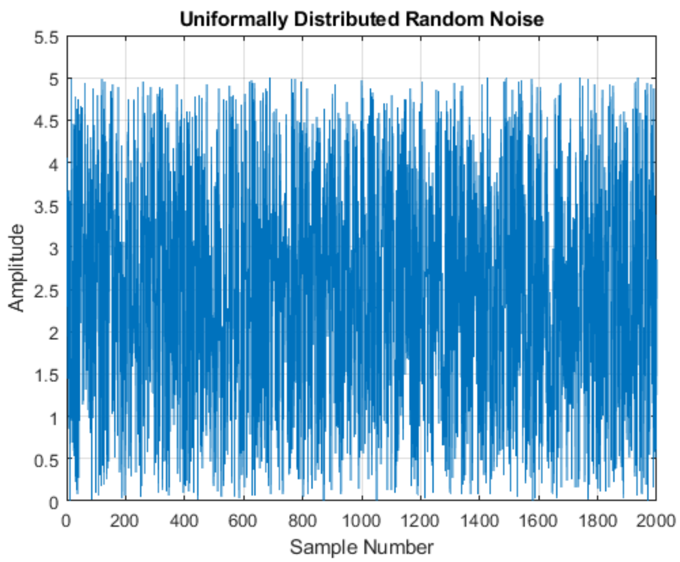


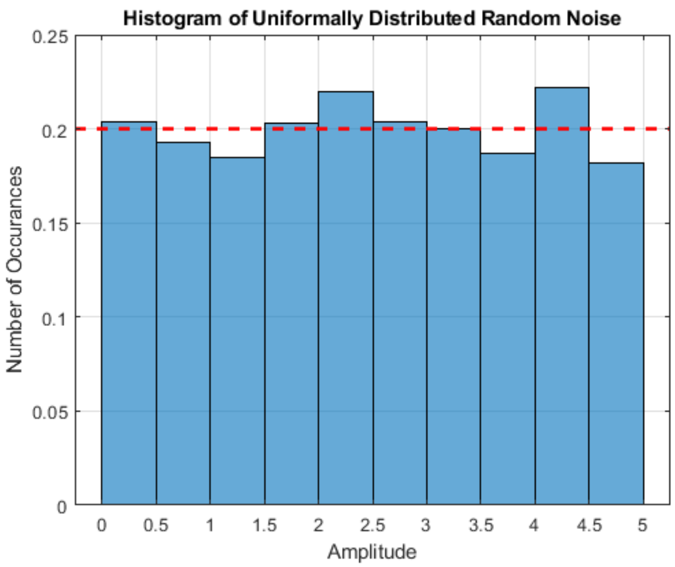


Varying probability density centered on the mean value

PDF for Uniform Random Noise

- PDF for this signal is a *uniform* distribution





The PDF is a flat across the amplitude range

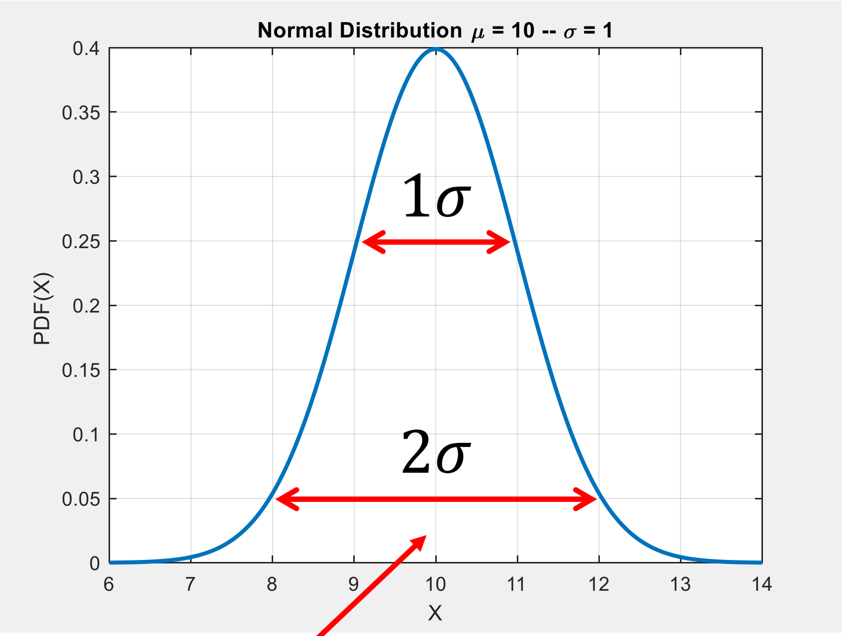
The Normal Distribution

Normal Distribution Example

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

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

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



Area under the curve = 1

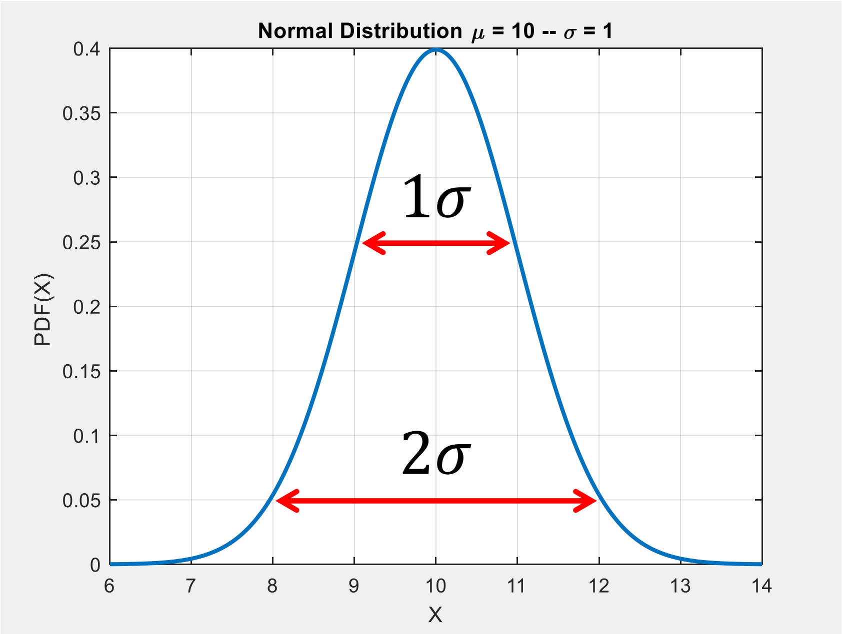
68.3% of values fall within $\pm 1\sigma$.

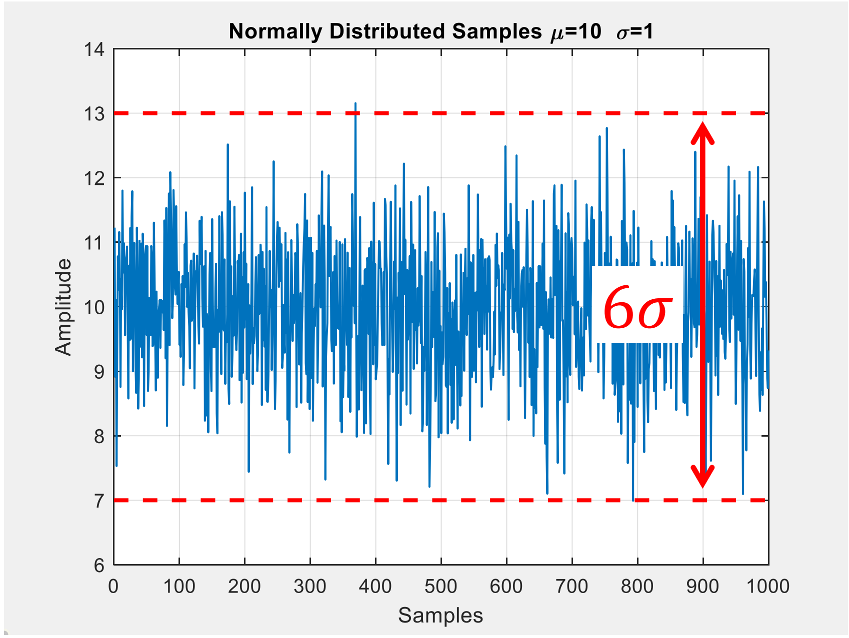
95.5% of values fall within $\pm 2\sigma$

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 ($\pm 3\sigma\ to\ \pm 4\sigma$)





The Central Limit Theorem

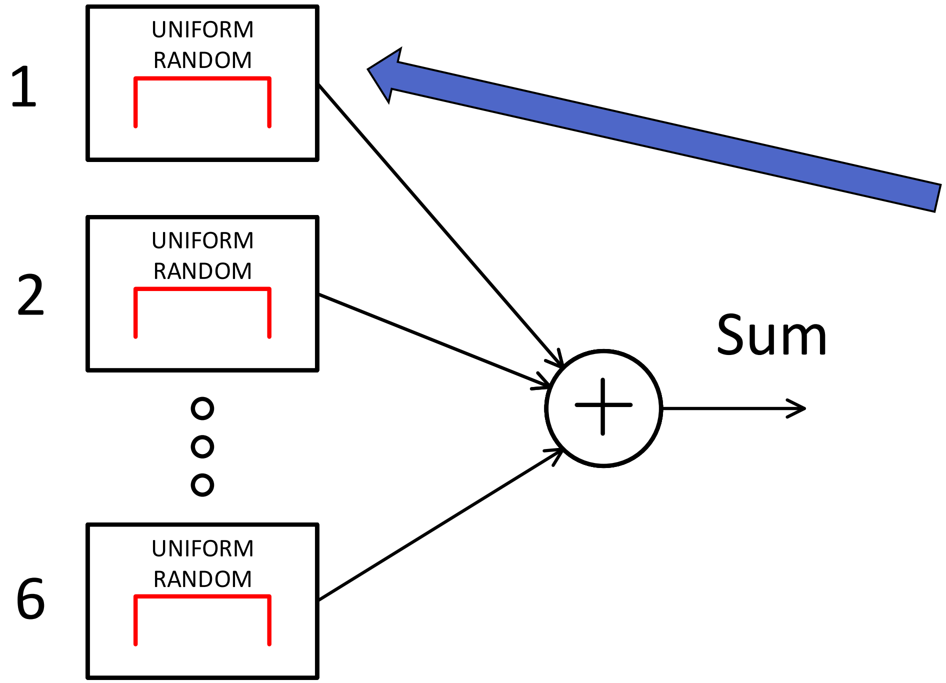
The Central Limit Theorem

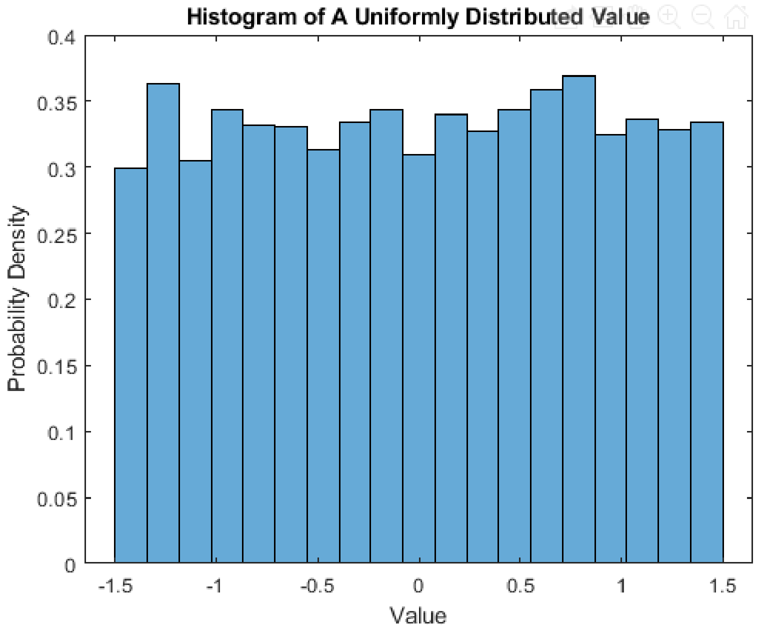
- The sum of random processes becomes normally distributed as more and more of the random numbers are added together.

- 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. Add them. What is the distribution of the sum?

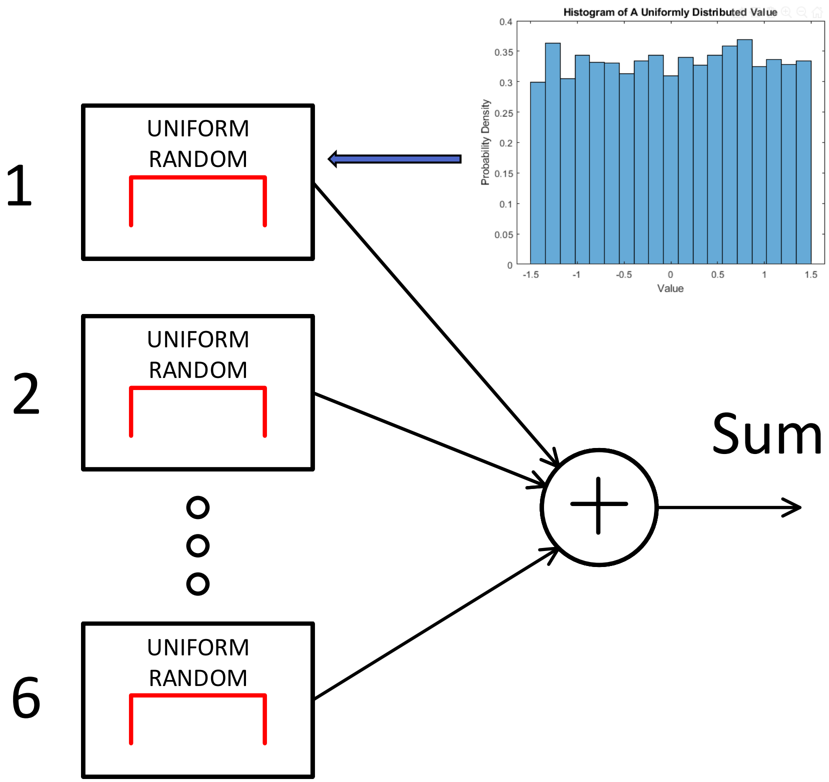


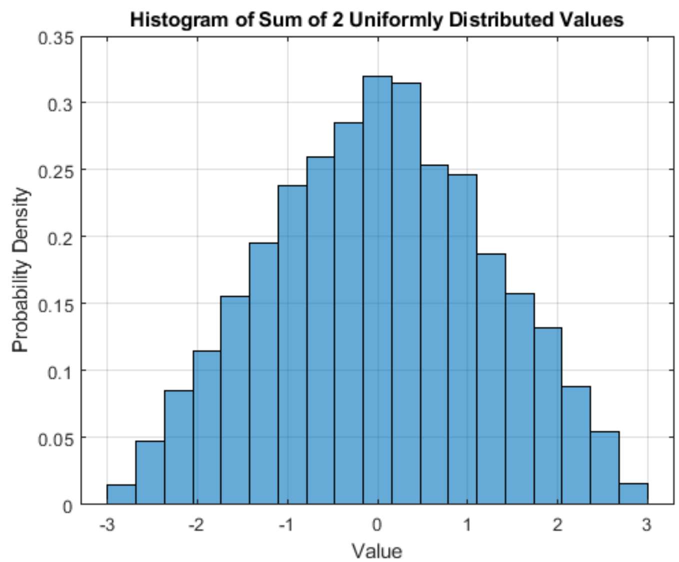


Distribution of 1 Uniform random variable

Central Limit Theorem MATLAB Example

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

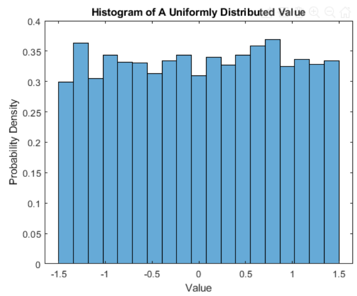


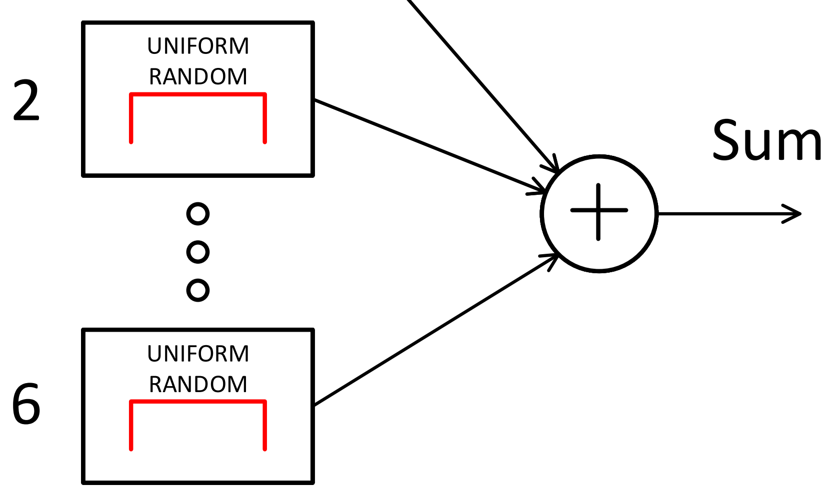


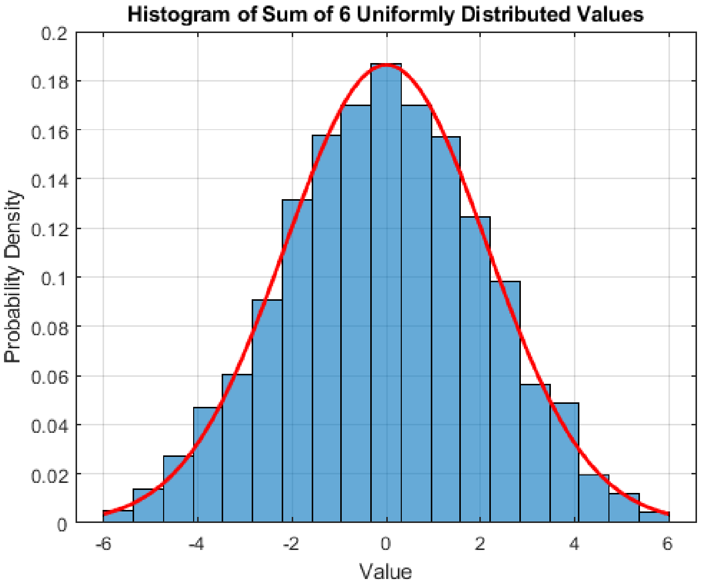
Distribution of sum of 2 Uniform random variables

Central Limit Theorem MATLAB Example

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







Distribution of sum of6 Uniform random variables

Digital Signal Processing

Precision and Accuracy

Precision and Accuracy

- Accuracy
  - How close is the estimated mean $\hat{\mu}$ to the true mean?

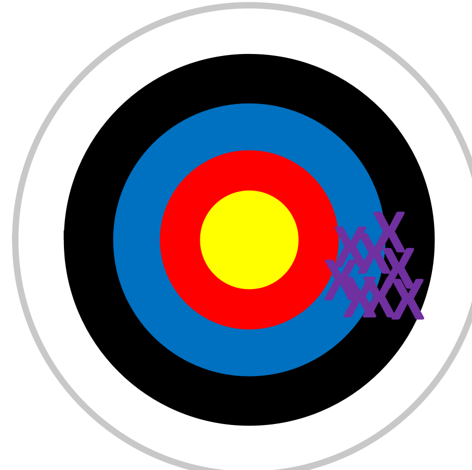
$$Accuracy = \hat{\mu} - \mu$$

- Precision
  - How well do the individual measurements or samples compare with each other?
  - It is expressed by the Signal to Noise Ratio (SNR) or by the Coefficient of Variation (CV)

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

Precision and Accuracy

Neither Accurate
Nor Precise



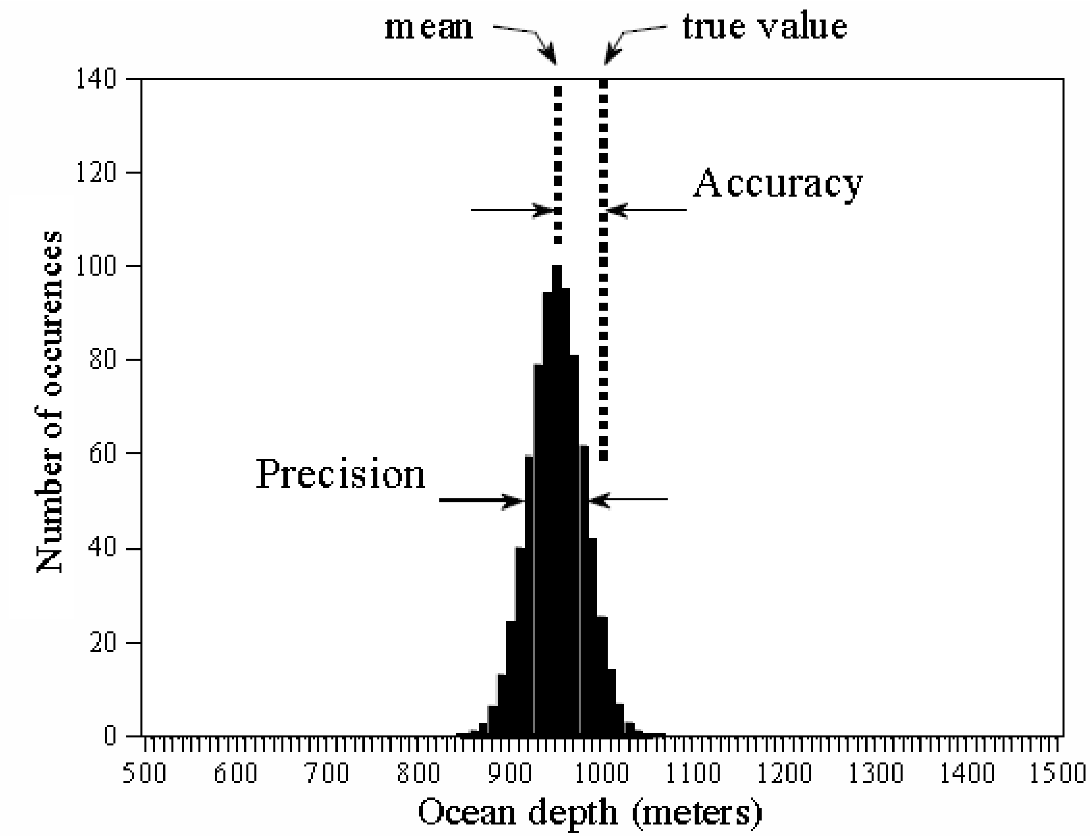
Not Accurate
But Precise

Accurate but
not Precise

Accurate
And Precise

Precision and Accuracy

- For a normally distributed signal



Digital Noise Generation

Digital Noise Generation

- Generating random noise is helpful for testing how DSP algorithms operate in noise
- Most programming languages can produce uniformly distributed random numbers
- By adding uniformly distributed random numbers you can create normally (Gaussian) distributed random numbers

Digital Noise Generation

- Another approach is to start with a seed value S
- Then apply the equation shown here

$$R = (aS + b)\ modulo\ c$$

- The modulo function takes the remainder after dividing by c
- The next seed is the last random value generated
- a, b and c are chosen to give good random characteristics

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 of a 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
- Precision and Accuracy
  - Precision is related to standard deviation, 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.