EE327 Digital Signal Processing Statistics and Noise 2

Yasser F. O. Mohammad

REMINDER 1: Mean and Variance

- Mean

$$\mu = \frac{1}{N}\sum_{i=0}^{N-1} x[i] = \frac{1}{N}\sum_{i=0}^{N-1} x_i \qquad \mu = \lim_{T\to\infty}\frac{1}{T}\int_0^T x(t)dt$$

- Variance

$$\sigma^2 = \frac{1}{N-1}\sum_{i=0}^{N-1}(x_i - \mu)^2 \qquad \sigma^2 = \lim_{T\to\infty}\frac{1}{T}\int_0^T (x(t)-\mu)^2 dt$$

- Root Mean Square (quadratic mean)

$$RMS = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1}(x_i)^2} \qquad RMS = \lim_{T\to\infty}\sqrt{\frac{1}{T}\int_0^T x(t)^2 dt}$$

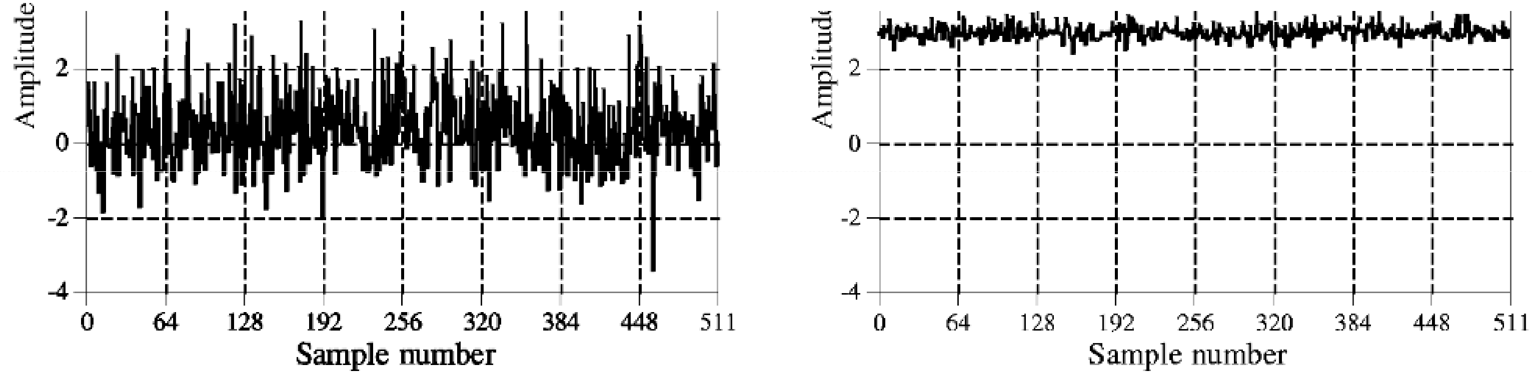
All continuous formulas assume that x(t) is zero for t<o or x(-t)=x(t)

Self test: Find the relation between RMS and variance for signals with zero mean

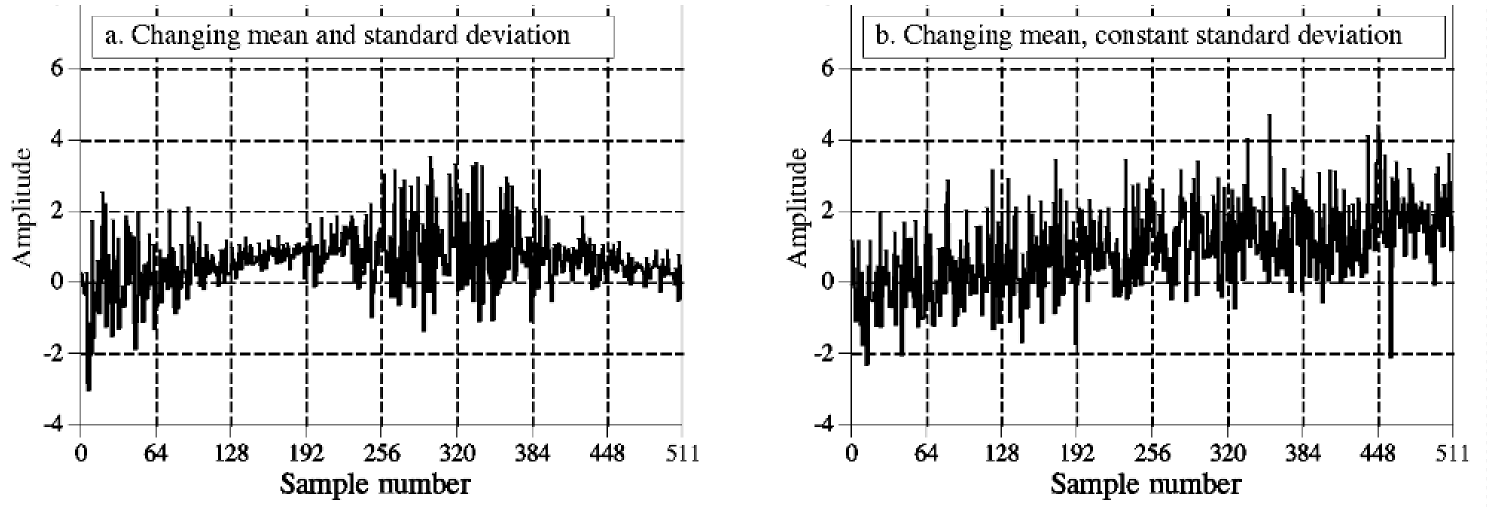
REMINDER 2: Examples of mean and variance

FAST TEST: These signals are discrete or continuous?
What is the domain of these signals?

Stationary



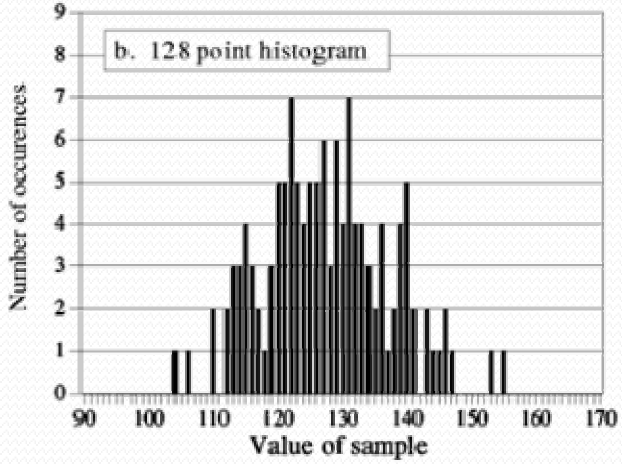
nonstationary

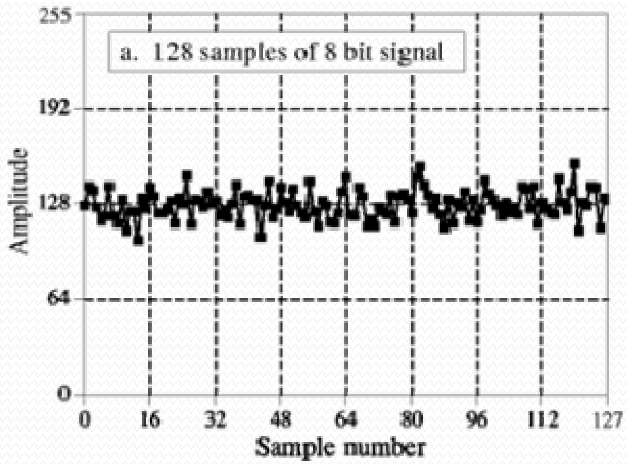


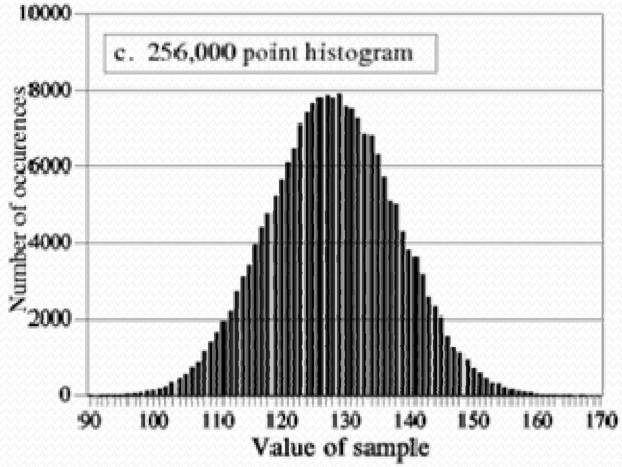
Histogram (acquired signal)

$$H[i] = H_i = \text{no. of samples that have the value } x$$

$$\text{where } i \times \Delta \le x < (i+1) \times \Delta$$







Calculating the Histogram

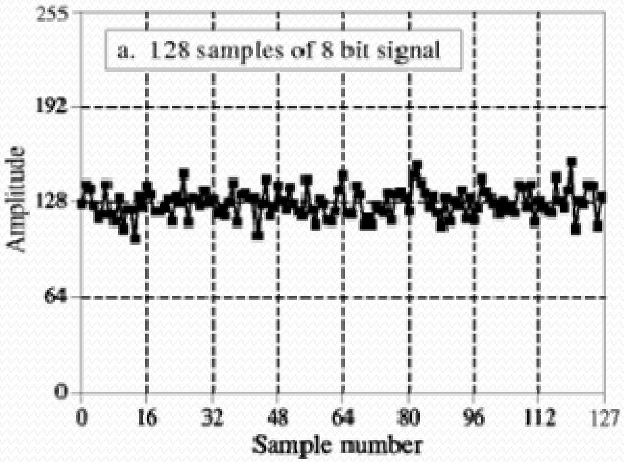
```
100 'CALCULATION OF BINNED HISTOGRAM
110 '
120 DIM X[25000]                    'X[0] to X[25000] holds the floating point signal,
130 '                               'with each sample having a value between 0.0 and 10.0.
140 DIM H%[999]                     'H%[0] to H%[999] holds the binned histogram
150 '
160 FOR I% = 0 TO 999               'Zero the binned histogram for use as an accumulator
170   H%[I%] = 0
180 NEXT I%
190 '
200 GOSUB XXXX                      'Mythical subroutine that loads the signal into X%[ ]
210 '
220 FOR I% = 0 TO 25000 '           'Calculate the binned histogram for 25001 points
230   BINNUM% = INT( X[I%] * 100 )
240   H%[ BINNUM%] = H%[ BINNUM%] + 1
250 NEXT I%
260 '
270 END
```

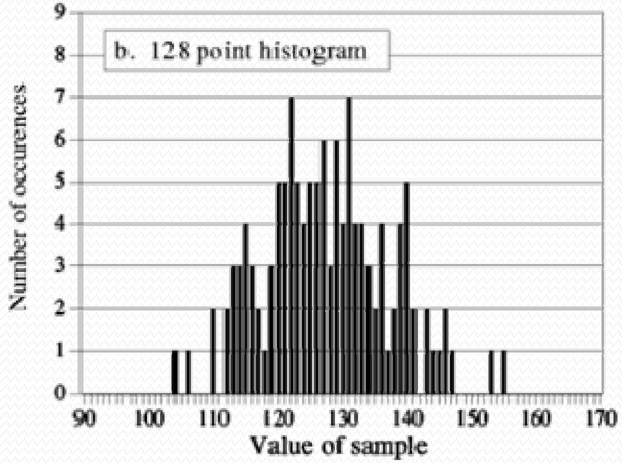
Histogram (acquired signal)

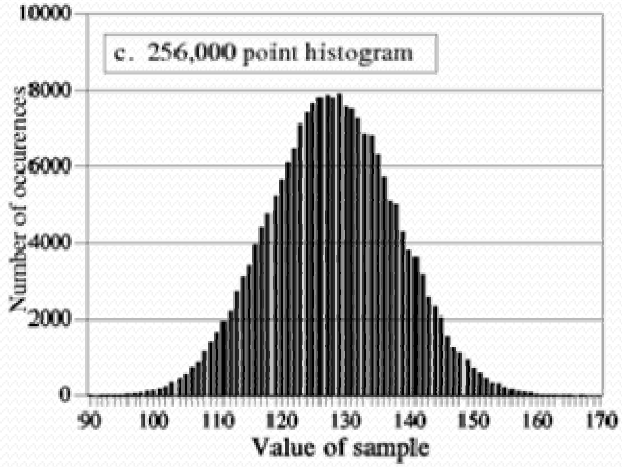
$$N = \sum_{i=0}^{M-1} H_i$$

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

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







Calculating histogram, mean and variance

```
100 'CALCULATION OF THE HISTOGRAM, MEAN, AND STANDARD DEVIATION
110 '
120 DIM X%[25000]                 'X%[0] to X%[25000] holds the signal being processed
130 DIM H%[255]                   'H%[0] to H%[255] holds the histogram
140 N% = 25001                    'Set the number of points in the signal
150 '
160 FOR I% = 0 TO 255             'Zero the histogram, so it can be used as an accumulator
170   H%[I%] = 0
180 NEXT I%
190 '
200 GOSUB XXXX                    'Mythical subroutine that loads the signal into X%[ ]
210 '
220 FOR I% = 0 TO 25000  'Calculate the histogram for 25001 points
230   H%[ X%[I%] ] = H%[ X%[I%] ] + 1
240 NEXT I%
250 '
260 MEAN = 0                      'Calculate the mean via Eq. 2-6
270 FOR I% = 0 TO 255
280   MEAN = MEAN + I% * H%[I%]
290 NEXT I%
300 MEAN = MEAN / N%
310 '
320 VARIANCE = 0                  'Calculate the standard deviation via Eq. 2-7
330 FOR I% = 0 TO 255
340   VARIANCE = VARIANCE + H%[I%] * (I%-MEAN)^2
350 NEXT I%
360 VARIANCE = VARIANCE / (N%-1)
370 SD = SQR(VARIANCE)
380 '
390 PRINT  MEAN  SD               'Print the calculated mean and standard deviation.
400 '
410 END
```

Effect of binning

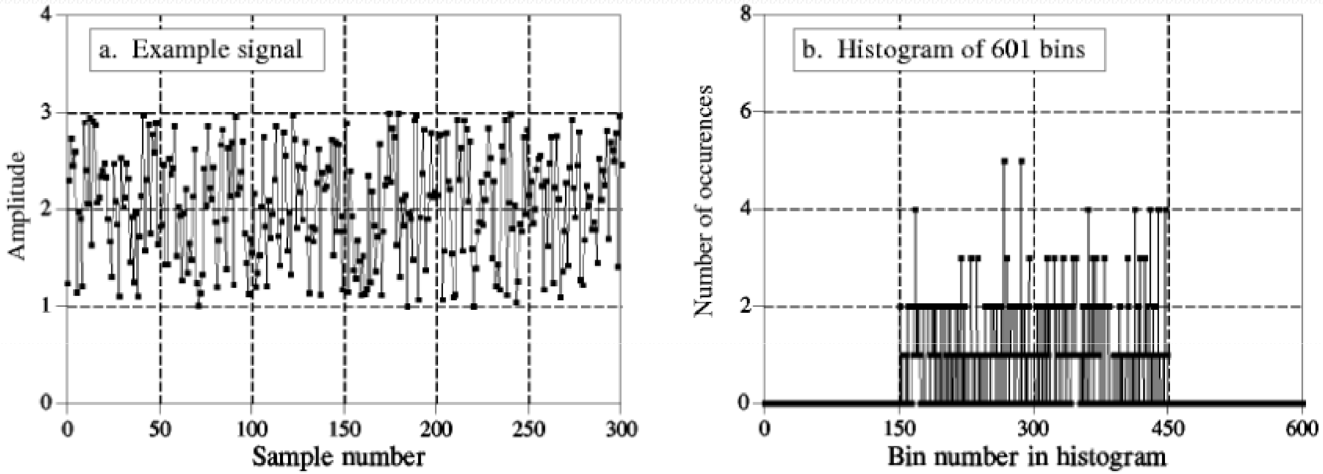
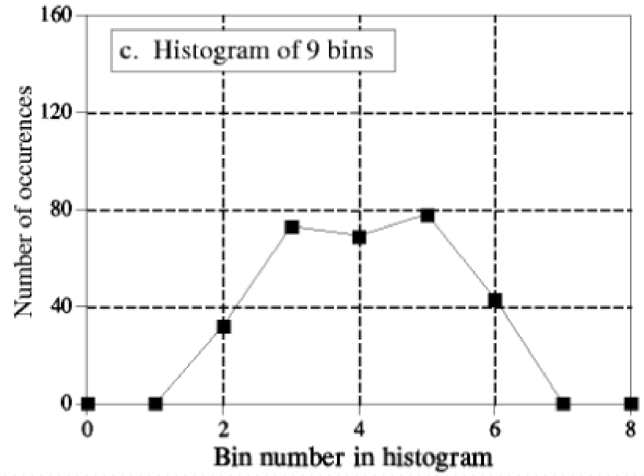


FIGURE 2-7
Example of binned histograms. As shown in (a), the signal used in this example is 300 samples long, with each sample a floating point number uniformly distributed between 1 and 3. Figures (b) and (c) show binned histograms of this signal, using 601 and 9 bins, respectively. As shown, a large 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.



Probability mass function (Pmf) Probability density function (Pdf)

- Pmf characterizes underlying process
- N→ ∞ ---------> Hist → Pmf

$$\sum Pmf = 1$$

$$\int_{-\infty}^{\infty} Pdf = 1$$

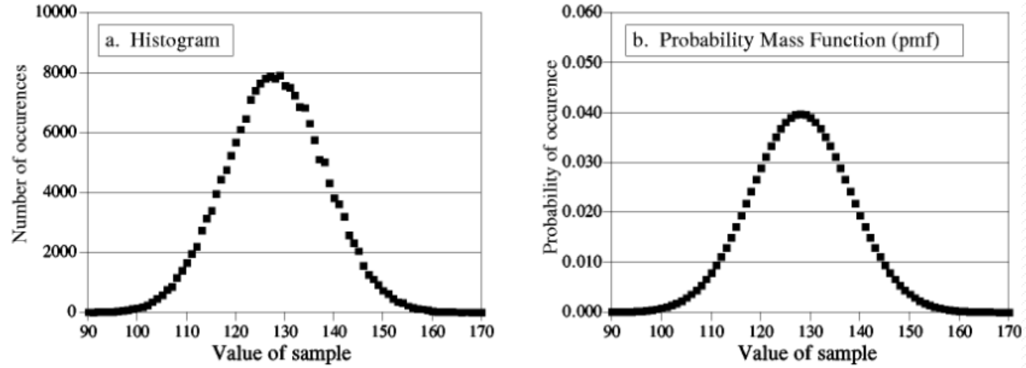
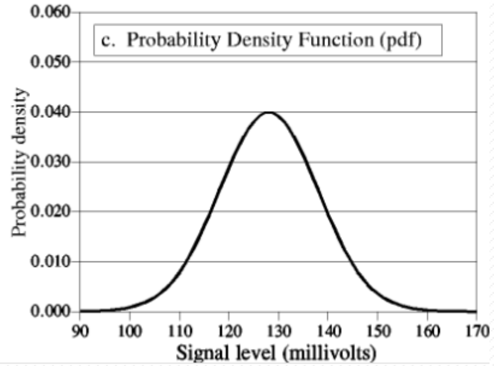
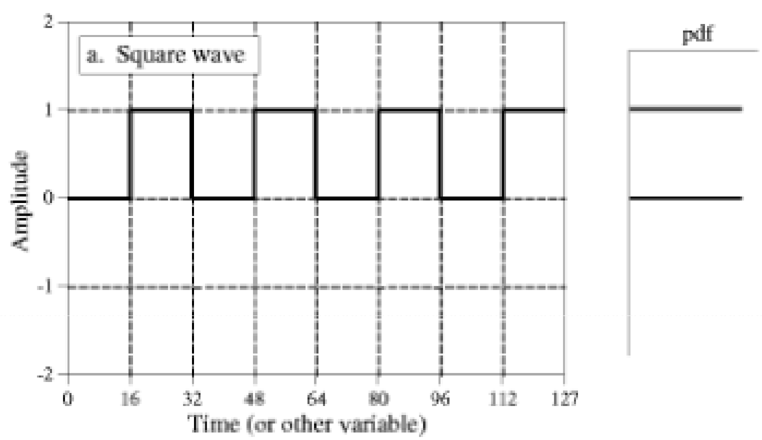
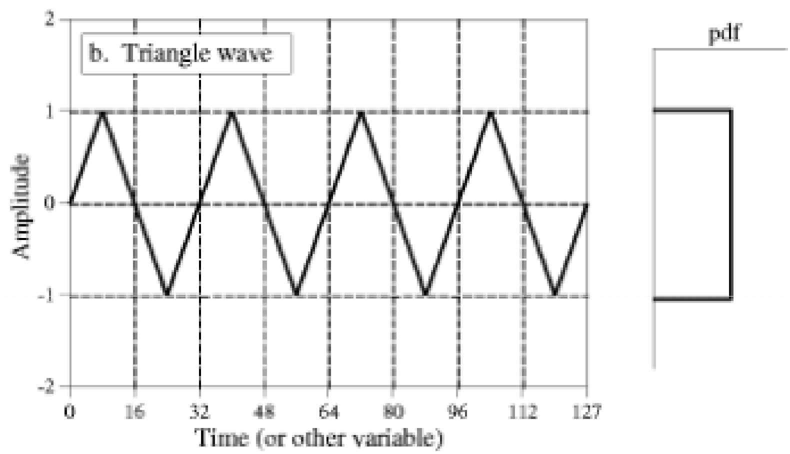


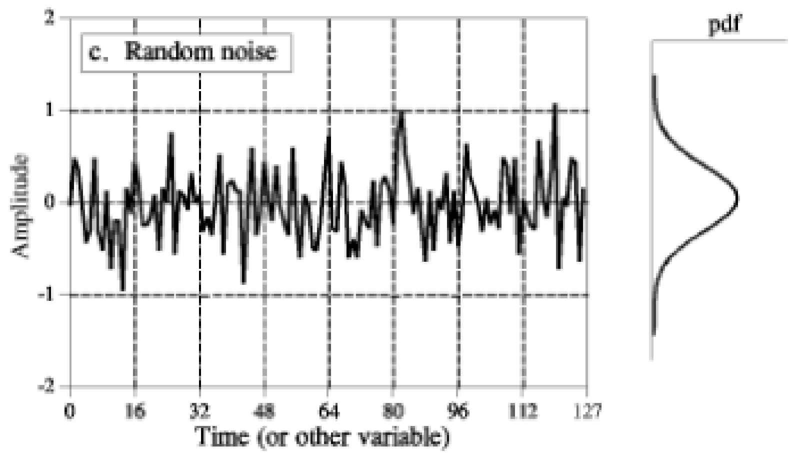
FIGURE 2-5
The relationship between (a) the histogram, (b) the probability mass function (pmf), and (c) the probability density function (pdf). The histogram is calculated from a finite number of samples. The pmf describes the probabilities of the underlying process. The pdf is similar to the pmf, but 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.



Example Pdfs







The Mighty Gaussian

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

Example Gaussians

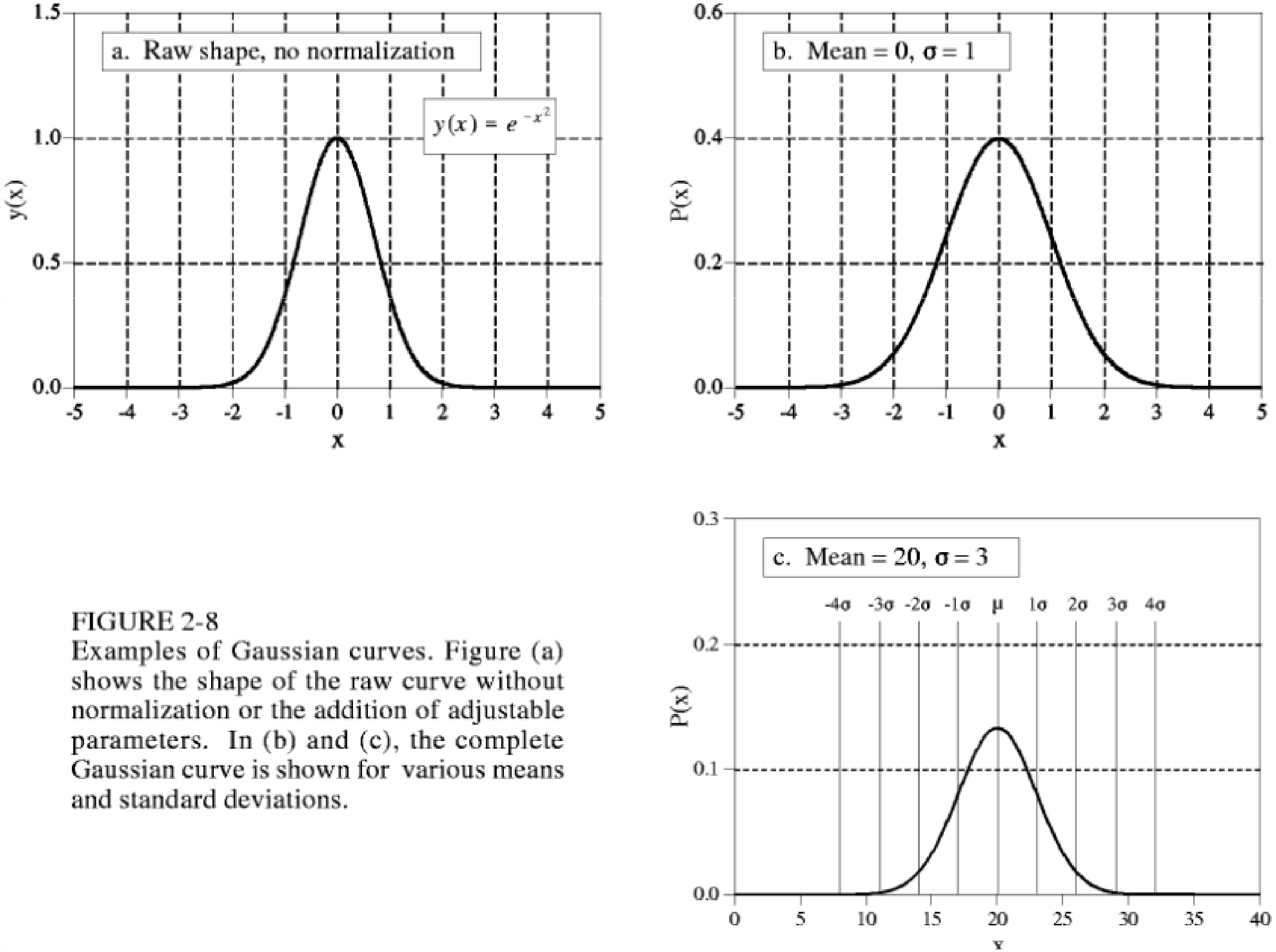
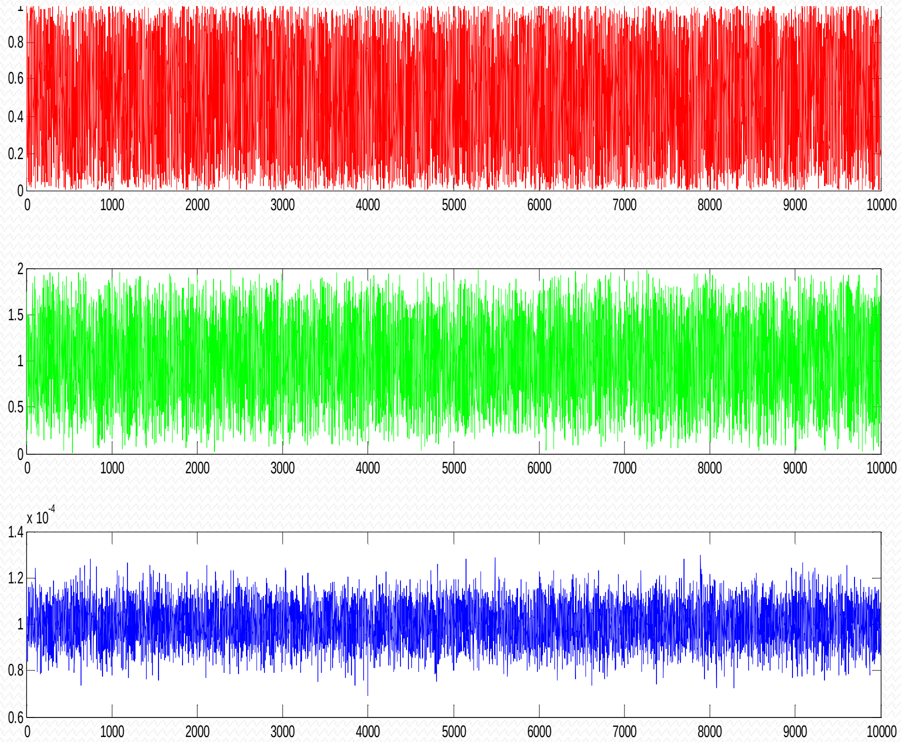
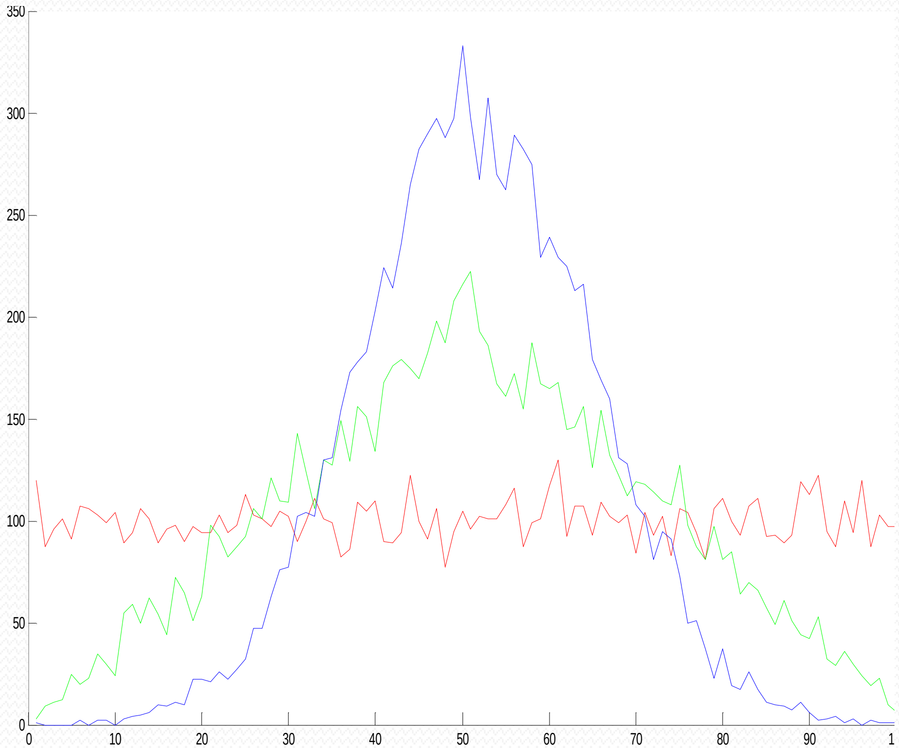


FIGURE 2-8
Examples of Gaussian curves. Figure (a) shows the shape of the raw curve without normalization or the addition of adjustable parameters. In (b) and (c), the complete Gaussian curve is shown for various means and standard deviations.

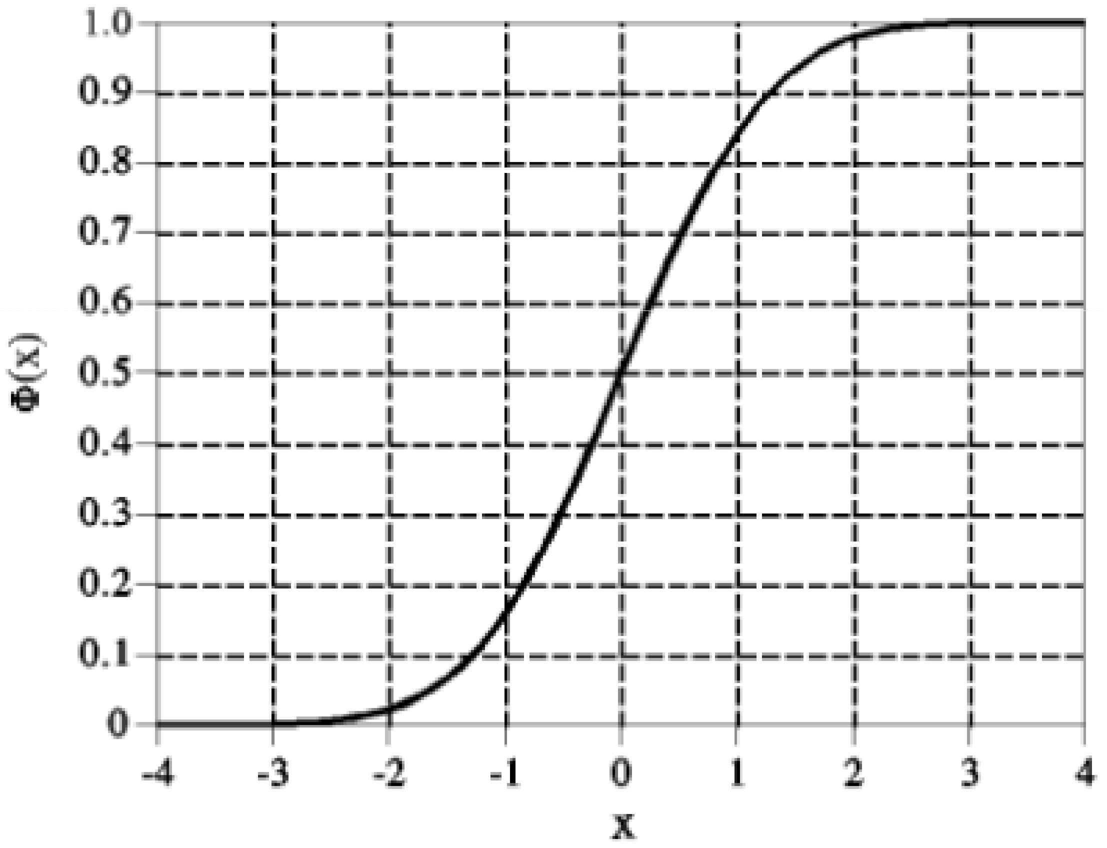
Why Gaussians

- Central Limit Theory
  - Effects of large number of uniform noise factors approximate a Gaussian





Cumulative Distribution Function



How to generate random numbers

- Uniform 0-1
  - X=RND
  - Variance is 1/12
  - Mean is 1/2

- Uniform from *a* to *b*
  - X=RND*(b-a)+a

SELF Test: What is the mean and variance of this signal

- Normal
  - $R_1 = RND, \quad R_2 = RND$

$$X = \mu + \sigma\left(\sqrt{-2\log(R_1)}\cos(2\pi R_2)\right)$$

How to combine random signals

- Assume *X*, *Y* are INDEPENDENT random signals with mean μ_x , μ_y and std. dev. σ_x, σ_y :

$$\mu_{aX \pm b} = a\mu_X \pm b \qquad \sigma^2_{aX \pm b} = a^2\sigma^2_X$$

$$\mu_{aX \pm bY} = a\mu_X \pm b\mu_Y \qquad \sigma^2_{aX \pm bY} = a^2\sigma^2_X + b^2\sigma^2_Y$$

SELF Test: Prove these identities

Chebyshev's Theorem

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

This is a lower bound

Precision and Accuracy

- Precision = repeatability
- Accuracy = bias

  → systematic errors
- The two questions:
  - Repeating will remove the error → precision
  - Calibration will remove the error → accuracy