DIGITAL SIGNAL PROCESSING: COSC390

Jordan Hanson

January 9, 2019

Whittier College Department of Physics and Astronomy

Previous lectures covered:

- Complex numbers 2: The Fourier series and Fourier transform (continuous and discrete)
- *Time-permitting*: The Laplace transform (continuous and discrete)

This lecture will cover: (Reading: **Chapter 2**)

- Statistics and probability: the normal distribution and other useful distributions
- Noise: digitization and sampling
- Noise: Spectral properties of noise, ADC and DAC

STATISTICS AND PROBABILITY: THE NORMAL DISTRIBUTION

The *mean*, μ, and *standard deviation*, σ, of a data set $\{x_i\}$ are defined as

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

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

Octave commands:

```
x = randn(100,1);
mean(x)
std(x)
```

One nice theorem: *The variance is the average of the squares minus the square of the average.* Let $\langle x \rangle$ represent the average of the quantity or expression x. We have

$$\sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2 \tag{3}$$

Proof: observe on board.

Note: There is a distinction between the *process or signal process* and the *the data*. Just because the data has a given μ and σ does not imply that the signal process has or will continue to have the exact same values of μ and σ. The underlying process could be *non-stationary*.

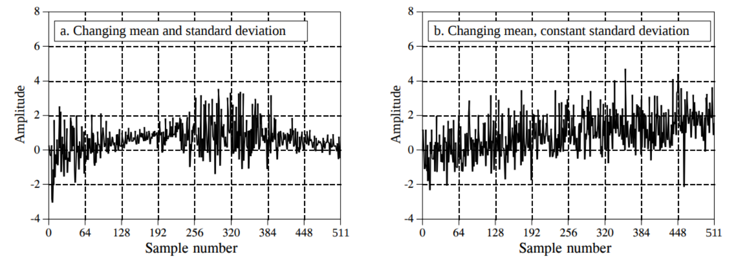

Figure 1: Signal processes in (a) and (b) are considered non-stationary because one or both of μ and σ depend on time.

A histogram is an object that represents the frequency[1] of particular values in a signal. For example, below is a histogram of 256,000 numbers drawn from a probability distribution:

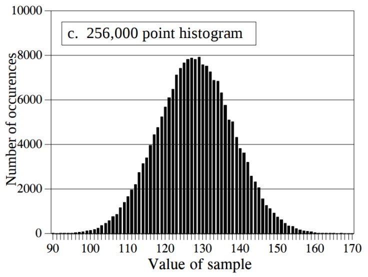

Figure 2: The histogram contains counts versus sample values.

[1] Careful: the word frequency refers to the number of occurences in the data, not a sinusoidal frequency.

The following octave code should reproduce something like Fig. 2 from the textbook:

```
x = randn(256000,1)*10.0+130.0;
[b,a] = hist(x,100);
plot(a,b,'o');
```

The function *randn(N,M)* draws $N \times M$ numbers from a normal distribution and returns them in the size the user desires. The function *hist(x,N)* creates N bins and sorts the data x_i into them.

For data that is appropriately stationary, we can use histograms to estimate μ and σ faster, since we only have to loop over bins rather than every data sample. Let H_i represent the counts in a given bin, and i represent the bin sample. We have:

$$\mu = \frac{1}{N} \sum_{i=1}^{M} i H_i \tag{4}$$

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{M} (i - \mu)^2 H_i \tag{5}$$

(To obtain the mean in signal *amplitude*, you'd have to convert bin number to amplitude - more on that in a moment).

Some vocabulary:

- **normalization** - Total probability is 1.0. For pdf - the integral from $[-\infty, \infty]$ is 1.0. For pmf - the sum from $[-\infty, \infty]$ is 1.0.
- **pmf** - Probability mass function: A *normalized continuous function* that gives the probability of a value, given the value.
- **histogram** - Histograms are an attempted measurement of the pmf by breaking the data into discrete bins. Histograms can be *normalized* as well.
- **pdf** - Probability density function: A *normalized continuous function* that gives the probability density of a value, given the value. Integrating the *normalized* pdf between two values gives the probability of observing data between the given values.

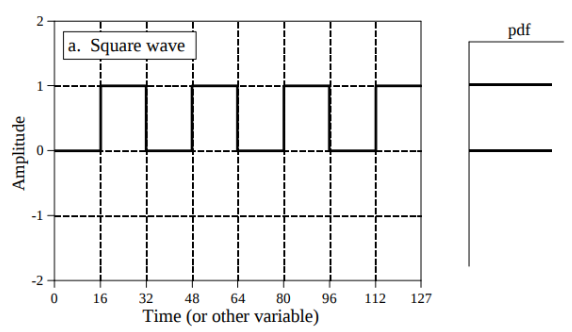

Figure 3: The square-wave signal spends equal time at 0.0 and 1.0, and the probability density function reflects that.

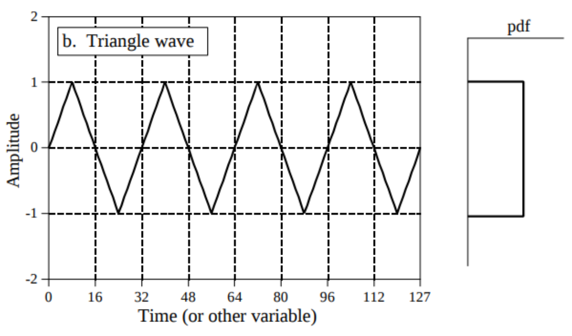

Figure 4: The triangle-wave signal spends equal time at all values *between* 0.0 and 1.0, and the probability density function reflects that.

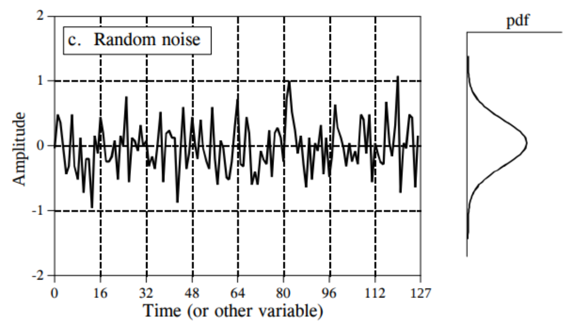

Figure 5: The random noise *usually* spends time near 0.0, but rarely it fluctuates to larger values.

Normally distributed data decreases in probability at a rate that is proportional (1) to the *distance from the mean*, and that is proportional (2) to the *probability itself*.

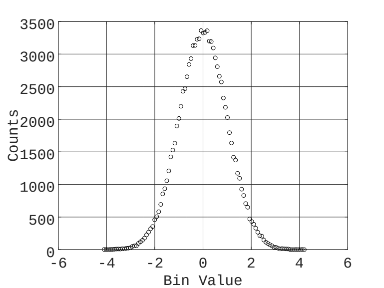

Figure 6: Normally distributed data counts decrease as measured further from the mean for *two reasons*.

Normal Distribution PDF

Let $p(x)$ be the PDF of normally distributed data x with mean μ. In order to obey conditions (1) and (2), the function $p(x)$ must be described by the following differential equation, where k is some constant.

$$\frac{dp}{dx} = -k(x - \mu)p(x) \tag{6}$$

Rearranging Eq. 6, we have

$$\frac{dp}{p} = -k(x-\mu)dx \tag{7}$$

Integrating both sides gives

$$\ln(p) = -\frac{1}{2}k(x-\mu)^2 + C_0 \tag{8}$$

Exponentiating,

$$p(x) = C_1 \exp\left(-\frac{1}{2}k(x-\mu)^2\right) \tag{9}$$

Ensuring that the PDF is *normalized* requires

$$\int_{-\infty}^{\infty} p(x)dx = 1 \tag{10}$$

But how do we integrate Eq. 9? First, a change of variables. Let $s = \sqrt{k/2}(x - \mu)$, so $ds = \sqrt{k/2}dx$. Then, we have

$$C_1\sqrt{\frac{2}{k}}\int_{-\infty}^{\infty}\exp(-s^2)ds = 1 \tag{11}$$

Squaring both sides, we have

$$C_1^2\frac{2}{k}\left(\int_{-\infty}^{\infty}\exp(-s^2)ds\right)^2 = 1 \tag{12}$$

Let's pretend the two factors of the integral involve different variables:

$$C_1^2 \frac{2}{k} \left(\int_{-\infty}^{\infty} \exp(-x^2) dx \right) \left(\int_{-\infty}^{\infty} \exp(-y^2) dy \right) = 1 \tag{13}$$

Now we have

$$C_1^2 \frac{2}{k} \int_{-\infty}^{\infty} \exp(-(x^2 + y^2)) dx dy = 1 \tag{14}$$

Change to polar coordinates ($x^2 + y^2 = r^2$)

$$C_1^2 \frac{2}{k} \int_{0}^{\infty} \int_{0}^{2\pi} r \exp(-r^2) dr d\phi = 1 \tag{15}$$

One more substitution: $u = r^2$, and $du = 2rdr$:

$$-\frac{C_1^2}{k}\int_0^\infty \int_0^{2\pi} \exp(-u)dud\phi = 1 \tag{16}$$

Solving for C_1, we find

$$C_1 = \sqrt{\frac{k}{2\pi}} \tag{17}$$

Thus the pdf of normally distributed data is

$$p(x) = \sqrt{\frac{k}{2\pi}} \exp\left(-\frac{1}{2}k(x-\mu)^2\right) \tag{18}$$

Let's defined $k = \frac{1}{\sigma_x^2}$ so that it's clear the exponent has the proper ratio of units:

$$\boxed{p(x) = \sqrt{\frac{1}{2\pi\sigma_x^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma_s}\right)^2\right)} \tag{19}$$

STATISTICS AND PROBABILITY: OTHER USEFUL DISTRIBUTIONS

CONCLUSION

UNIT 1.3 OUTLINE

Previous lectures covered:

- Complex numbers 2: The Fourier series and Fourier transform (continuous and discrete)
- *Time-permitting*: The Laplace transform (continuous and discrete)

This lecture will cover: (Reading: **Chapter 2**)

- Statistics and probability: the normal distribution and other useful distributions
- Noise: digitization and sampling
- Noise: Spectral properties of noise, ADC and DAC