

# DIGITAL SIGNAL PROCESSING: COSC390

---

Jordan Hanson

January 10, 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*,  $\mu$ , and *standard deviation*,  $\sigma$ , of a data set  $\{x_i\}$  are defined as

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (1)$$

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

Octave commands:

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

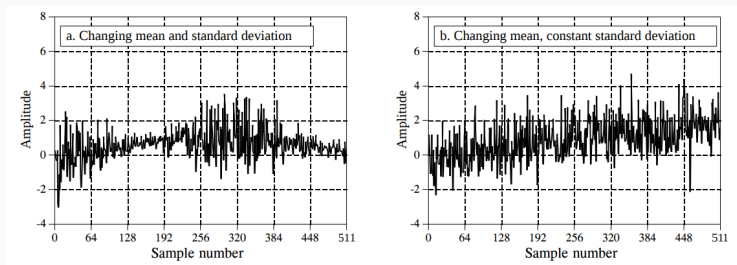
One nice theorem: *The variance is the average of the squares minus the square of the average.* Let  $\langle x \rangle$  represent the average of the quantity or expression  $x$ . We have

$$\sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2 \quad (3)$$

Proof: observe on board.

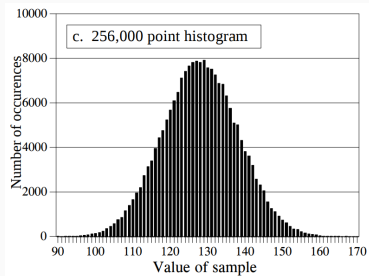
# STATISTICS AND PROBABILITY: THE NORMAL DISTRIBUTION

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



**Figure 1:** Signal processes in (a) and (b) are considered *non-stationary* because one or both of  $\mu$  and  $\sigma$  depend on time.

A **histogram** is an object that represents the frequency<sup>1</sup> 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.

---

<sup>1</sup>Careful: the word frequency refers to the number of occurrences 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  $\mu$  and  $\sigma$  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:

$$\mu = \frac{1}{N} \sum_{i=1}^M i H_i \quad (4)$$

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

To obtain the mean in signal *amplitude*, you'll have to convert bin number to amplitude.

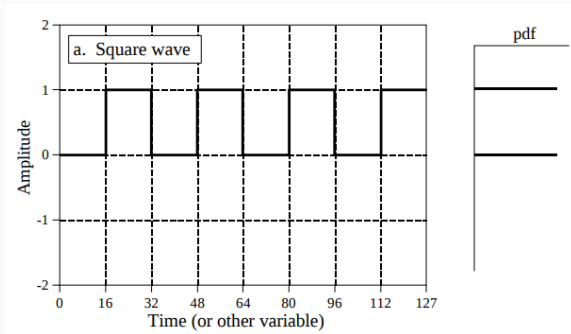
3.1  
-0.03  
1.2  
0.2  
-0.7  
-1.45  
2.2  
-0.05  
0.93  
0.21

**Table 1:** Using Eq. 4 and 5, find estimates of  $\mu$  and  $\sigma$  for this data.

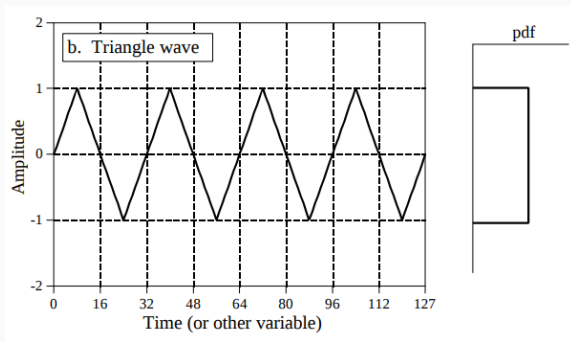
```
x = [...];  
[b,a] = hist(x,4); %(How many bins?)
```

Some vocabulary:

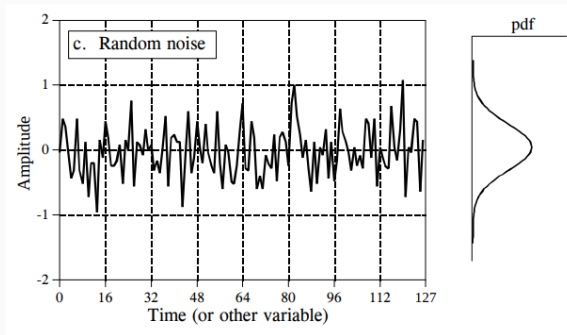
- **normalization** - Total probability is 1.0. For pdf - the integral from  $[-\infty, \infty]$  is 1.0. For pmf - the sum from  $[-\infty, \infty]$  is 1.0.
- **pmf** - Probability mass function: A *normalized continuous function* that gives the probability of a value, given the value.
- **histogram** - 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.



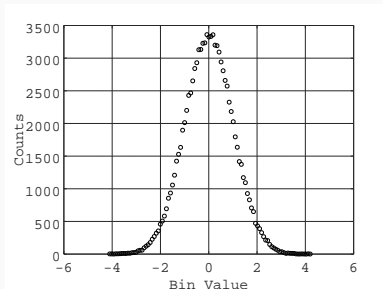
**Figure 4:** The triangle-wave signal spends equal time at all values *between 0.0 and 1.0*, and the probability density function reflects that.



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

# NORMAL DISTRIBUTION

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  $\mu$ . 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.

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



Rearranging Eq. 6, we have

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

Integrating both sides gives

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

Exponentiating,

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

Ensuring that the PDF is *normalized* requires

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

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

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

Squaring both sides, we have

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

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

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

Now we have

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

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

$$C_1^2 \frac{2}{k} \int_0^{\infty} \int_0^{2\pi} r \exp(-r^2) dr d\phi = 1 \quad (15)$$

## NORMAL DISTRIBUTION

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

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

Solving for  $C_1$ , we find

$$C_1 = \sqrt{\frac{k}{2\pi}} \quad (17)$$

Thus the pdf of normally distributed data is

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

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

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

# STATISTICS AND PROBABILITY: PROGRAM- MING WITH OCTAVE

---

More on the *hist* function in octave<sup>2</sup>

```
pkg install -forge io
pkg install -forge statistics
pkg load statistics
pkg help histfit
histfit(randn(1000,1))
histfit(rand(1000,1))
```

Let's work out the  $\sigma$  of a *flat* distribution between  $[0, 1]$ . What is it for a flat distribution between  $[-1, 1]$ ? (We can derive this by hand as well if we cannot access statistics package).

---

<sup>2</sup>I hope this works, but if not, it's ok.

Some interesting notation for normal distributions:

$$N(\mu, \sigma) = \sqrt{\frac{1}{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \quad (20)$$

Let's write a function **NGaus.m** that produces the Gaussian probability given  $\mu$  and  $\sigma$ :

```
function ret = NGaus(mu,sigma,x)
    ...
endfunction
```

Now let's write a function *NRand* that sums  $N$  uniformly-distributed (flat) random variables  $x$ :

```
function ret = NRand(n)
    ret = sum(rand(n,1));
endfunction
```

Create a histogram of a few hundred outputs of *NRand*. What do you notice about the pmf? Let's plot *NGaus* on the same axes as the histogram of *NRand*. How do they compare?

We are on our way to producing  $N(0,1)$  distributed numbers, and therefore our first **noise** signals...



More vocabulary:

- **cdf** - Cumulative distribution function: Probability that a continuous random variable  $X$  is less than some value  $x$ . For a given pdf, the cdf  $\Phi(X)$  is the integral of the total probability on  $[-\infty, x]$ . The derivative of the pdf is related to the pdf via the fundamental theorem of calculus.

If the pdf follows  $f(x)$ , then

$$\Phi(X \leq x) = \int_{-\infty}^x f(x) dx \quad (21)$$

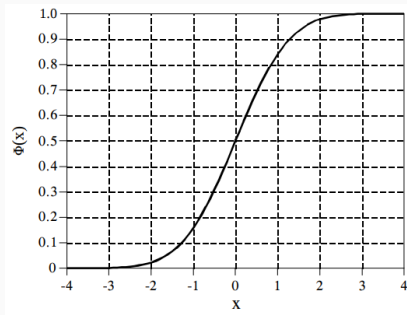
The Box-Muller method for  $N(0, 1)$  distruted numbers:

$$X_1 = \sqrt{-2 \ln(U)} \cos(2\pi V) \quad (22)$$

$$X_2 = \sqrt{-2 \ln(U)} \sin(2\pi V) \quad (23)$$

Try this in octave...

The cdf of  $N(0, 1)$  has an expected shape, but can't be expressed with elementary functions.



**Figure 7:** The cumulative distribution of the normal distribution. Although we can plot it, it's hard to write. We will discuss the *erf* and *erfc* functions in the near future.

# STATISTICS AND PROBABILITY: OTHER USEFUL DISTRIBUTIONS

---

We now know how to obtain random uniform numbers (**rand**) in octave, and have algorithms (Box-Muller) and functions (**randn**) in octave for  $N(0, 1)$ <sup>3</sup>. What if we require a *different pdf*? One technique is to use *inverse transform sampling*:

1. For the pdf  $p(x)$ , work out the cdf  $\Phi(x)$ .
2. Generate a sample of uniform random numbers  $u_i \in [0, 1]$ .
3. Call  $\Phi(u_i)$ .

**Write an octave script that generates exponentially-distributed numbers. Be careful to normalize when comparing to the expected pdf  $\propto \exp(-x)$ .**

---

<sup>3</sup>This can be scaled to any  $\mu$  and  $\sigma$  values we need.

Octave has many pre-programmed distributions. Although system noise is usually normally distributed, it's good to know some of these:

<https://octave.org/doc/v4.2.0/Distributions.html>

## NOISE: DIGITIZATION AND SAMPLING

---

## CONCLUSION

---

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