Lecture Notes on Characterizing Noise

1 Introduction

Ultimately, fundamental sources of noise and the discrete nature of particles limit our ability to probe nature. For instance, say we are trying to determine the color of a distant galaxy. The galaxy may emit just a few photons per second and, if the noise in our detector is not too big, we might be able to identify those photons. On a smaller scale, we might be looking for a rare decay of a subatomic particle. To see it, we would have to keep the background counts in the detector at a very low level. There are a lot of techniques in experimental physics aimed at optimizing the signal to noise ratio. To understand how these work, we must know how to characterize sources of noise.

As with signals, it is instructive to look at noise in both the time and frequency domain. This will give us an opportunity to introduce two of the three fundamental theorems that link the two domains. They are Parseval's theorem and the Wiener-Khinchin Theorem (the third is the convolution theorem). In the following, we consider the data as continuous function of time and we will concentrate on just the noise component. In general,

$$h(t) = s(t) + n(t) \quad \text{with Fourier transform} \quad H(f) = S(f) + N(f), \tag{1}$$

where $h(t)$ is the total signal, $s(t)$ is the part that contains the information we are trying to get at, and $n(t)$ is the noise. To be concrete, we will give $h(t)$ the units of volts and concentrate just on $n(t)$.

There are a number of references on digging signals out of noise and the characterization of noise. My two favorites, as usual, are *Numerical Recipes*, and Chapter 15 of *The Art of Electronics*.

2 Time Stream

A noise waveform can have a wide variety of shapes. In the upper left hand panels of Figures 1, 2, and 3, the time streams of three different types of noise are plotted, "Gaussian white noise" low-pass filtered "Gaussian noise" and "uniform noise." By time stream, we mean that if you hooked up an oscilloscope to a noise source you would see what is plotted. These plots were made using computer generated random numbers with $\bar{n}(t) = 0$. We may think of the noise as being "sampled" or measured every second (see lecture notes on sampling). In the cases of Gaussian white noise and uniform noise, each sample is completely independent from the previous one. This is not true for the filtered Gaussian noise. It was made by low-pass filtering the Gaussian noise with a time constant of 5 seconds (see lecture notes on signals). As we will show below, this introduces correlations into the noise. These three waveforms all look like noise. How do we tell them apart? In other words, how do we characterize noise?

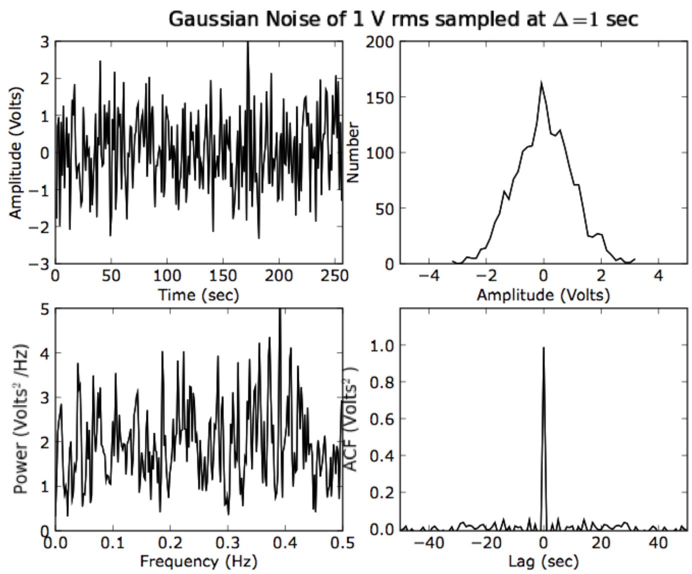

Figure 1: Gaussian-distributed, or sometimes called normally-distributed, noise with 1 V rms. The upper left shows the time stream samples at 1 second. The upper right shows a histogram of the time stream. This histogram has a Gaussian profile with a "sigma" of 1 Volt. The lower left plot shows the power spectral density (the integral of which is the variance, or rms^2). Since this power spectrum is on average flat, the noise is often termed "white." The lower right panel shows the autocorrelation function. For uncorrelated samples, there are no correlations. The value at zero lag is variance.

One immediate way to characterize the noise is to find the variance (or square of the root-mean-square), which is nearly the same as the sample variance. This is a convenient measure of the spread in the data. The formula for $N >> 1$ is

$$rms^2 = \sigma^2 = \frac{1}{N}\sum_i (n_i - \bar{n})^2 = \frac{1}{N}\sum_i n_i^2, \tag{2}$$

where n_i is the noise sample at time i and as we said above $\bar{n} = 0$. For the Gaussian white noise this is 0.99 Volts^2, for the filtered noise it is 0.088 Volts^2, and for the uniform noise it is 0.086 Volts^2.

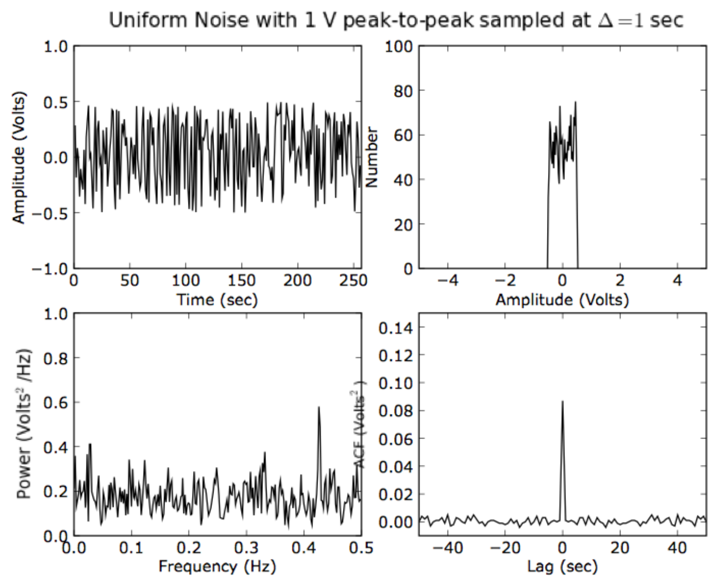

Figure 2: Uniform-distributed noise with an amplitude of 1 V peak to peak. The upper left shows the time stream samples at 1 second. The upper right shows a histogram of the time stream. This histogram has a top-hat profile with a width of 1 Volt. The lower left plot shows the power spectral density (the integral of which is the variance). The lower right panel shows the autocorrelation function. Again, for uncorrelated samples, there are no correlations. The value at zero lag is variance.

3 Distribution

It is difficult to get much quantitative information by looking directly at the time stream. One way to characterize the noise is by finding the *distribution* of the amplitudes of the waveform. When we say "Gaussian" or "uniform" we are talking about the distribution.

To find the distribution, we ask for how many samples, for instance, is the signal between 0.2 and 0.22 volts and compare that to the number of samples for which the signal is between 0.22 and 0.24 volts. The results are shown in the upper right panel in each plot. Note that if the samples were shuffled around, you would get the same distribution. The distribution is completely insensitive to the time ordering of the data; it just accounts for the distribution of amplitudes.

Consider the case for which the distribution is Gaussian. This is the most common case. If you were to fit a Gaussian profile, $\exp(-V^2/2\sigma^2)$, to the distribution the σ^2 you'd get would be the same as the variance you'd get from the time stream using equation 2. It turns out that filtering data does not change the distribution; note that the low-passed Gaussian noise is still Gaussian. Finally, You can see where uniform noise gets its name. Each value of the amplitude is equally likely.

4 Power Spectral Density

Next we move on to the *power spectral density.* This tells us how the variance is distributed amongst the various frequency components. The power spectral density is in essence the square of the magnitude of the Fourier transform. This means that it does not contain any phase information. For the moment, we will ignore the subtleties associated with sampling and the finite duration of the time stream and consider the mathematical form of the Fourier Transform.

$$N(f) = \int_{-\infty}^{\infty} n(t) \exp(i2\pi f t) dt \qquad \text{Volts/Hz.} \tag{3}$$

$N(f)$ tells us the amount of "signal" at each frequency as we discussed in class. It contains phase information and thus is a complex quantity. Now, form the magnitude squared of this,

$$P'(f) = |N(f)|^2 + |N(-f)|^2 \qquad \text{Volts}^2/\text{Hz}^2. \tag{4}$$

$P'(f)$ is the power spectral density and it tells us about the squared amplitude of each Fourier component; in other words, the variance at each frequency. If the signal is in volts, the variance must have units of V^2; we will deal with the Hz^2 below.

Although the distribution of the data is insensitive to the time ordering of the points, the Fourier transform is very sensitive to the ordering. In fact, it sniffs out any periodicity in the ordering that it can find.

In every physics class you learn that the "power" in a wave is proportional to its amplitude squared. Since a description of a signal in Fourier space contains no new information – the frequency domain just lets us think about the signal in a different way – you might suspect, and rightfully so, that the total power in the time domain is the same as the total power in the frequency domain. This is called Parseval's theorem. It is one of the fundamental theorems in signal processing. It says that

$$\begin{aligned}
\text{Total power} &= \int_{-\infty}^{\infty} n(t)^2 dt && (5)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} N(f)N(f')\exp(-i2\pi(f+f')t)dt\,df\,df' && (6)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} N(f)N^*(-f')\exp(-i2\pi(f-f')t)dt\,df\,df' && (7)\\
&= \int_{-\infty}^{\infty} N(f)^* N(f)df && (8)\\
&= \int_{0}^{\infty} P'(f)df && (9)
\end{aligned}$$

Note that in the second line we used the fact that $n(t)$ is a real function, in other words $N(f) = N^*(-f)$. The derivation did not rely on the fact that that waveform was noise; the relation is generally true.

Now we consider the units. In the above derivation, we used total power and relied on the fact that a noise train could not be infinitely long, or else we would get infinite power. The usual practice is to say that the total variance $1/T \int_0^T n(t)^2 dt$ must be the same in both domains. When this is done, the units of $P'(f)$ change slightly. $P'(f)$ gets multiplied by the what's called the resolution bandwidth or $1/T$. The result, which we will denote $P(f)$, is the power spectral density (PSD); it now has units of Volts^2/Hz. Parseval's Theorem now takes the form

$$\text{Total variance} = \frac{1}{T}\int_0^T n(t)^2 dt = \int_0^{f_N} P(f)df. \qquad (10)$$

The upper limit on the frequency interval tells us to integrate up to the Nyquist frequency (recall the discussion of sampling).

For Gaussian white noise, we see that the PSD is flat. There is equal *power* or *variance* at all frequencies. This is what is meant by white. The uniform noise is also white. The filtered Gaussian noise is clearly not white; the high frequency components have been filtered out. There is a lot more *power* or *variance* at low frequencies. In the notes on signals, a filter was characterized by the "3dB point." Recall that $f_{3dB} = 1/2\pi\tau = 1/2\pi RC$. The τ for these data is 5 seconds. So we expect the PSD to be a factor of 2 below the unfiltered value at 0.03 Hz. You can tell that this happens. If we integrate under each of the three curves (you can do this pretty well by eye), the result is 0.97, 0.083, 0.088 Volts^2 for Gaussian, filtered Gaussian, and uniform noise respectively.

5 Auto-correlation Function

There is one more way to characterize noise that I am aware of and this involves the auto-correlation function, or just the correlation function for short. We can ask ourselves, if the signal is high at time t_1, what is the chance that the signal is also high a short time later, at $t_1 + \tau$? If there is some finite chance, we say that the data are correlated. Noise can be either correlated or uncorrelated.

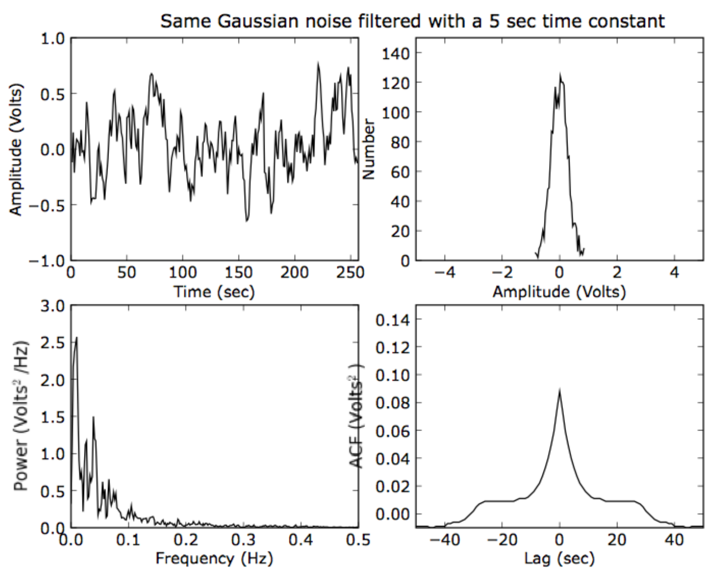

Figure 3: Filtered Gaussian-distributed noise. The upper left shows the time stream samples at 1 second. The upper right shows a histogram of the time stream. Note that even though the data are filtered, the histogram, or distribution, is still Gaussian in shape. The lower left plot shows the power spectral density (the integral of which is the variance). The lower right panel shows the autocorrelation function. Because there are correlations, the autocorrelation function has some width. The somewhat funny shape is due to the finite length of the time stream which was not corrected for. The value at zero lag is the variance.

How do we quantify the tendency to be correlated? Well, we do just what our intuition tells us. Step 1. Take the waveform whose correlations you are interested in, $n(t)$, and shift it slightly, say by τ, to form $n(t+\tau)$. Step 2., multiply these two together. If $n(t+\tau)$ tends to the same value as $n(t)$ then both the positive and negative parts will on average be the same. Step 3., integrate, or average, over the product to get the net effect. Mathematically we write this as

$$A(\tau) = \int_{-\infty}^{\infty} n(t)n(t+\tau)dt. \tag{11}$$

This is awfully similar to the convolution, there the τ in the integrand is replaced by $-\tau$. What happens to the autocorrelation when $\tau = 0$? Well, we just get the total power in the signal. If we were to divide by the total time of the integration, then we would get the variance. In the other extreme, if τ is long compared to any correlation time, then the places with a high signals will on average not correlate to other places with high signals and the integral will be zero. The correlation time, τ_c, is the time for which $A(\tau)$ drops to e^{-1} its value (though this definition is somewhat arbitrary).

In the lower right hand panel of figures 1, 2 and 3, we have plotted the correlation function of the time stream. For the first case, Gaussian white noise, $\tau_c = 0$. There are no correlations in the data. The data have been normalized so that $A(\tau = 0)$ is the variance of the time stream. This is done by dividing $A(\tau)$ in equation (7) by T. Analogous to what was done in equation (6). The uniform noise also has no correlations. But, look what happens when we filter the data. We introduce correlations. This makes sense. When we take away the high frequency components, the signal cannot change rapidly. Thus, if one point is high, the next will, on average, be high too. For the low pass filter, $\tau_c = RC$.

And now we get to a beautiful theorem called the Wiener-Khinchin Theorem, named after its co-discoverers. It says that the Fourier transform of the power spectral density is the auto-correlation function. Awesome!

$$A(\tau) = \int_{-\infty}^{\infty} P'(f)exp(-i2\pi f\tau)df. \tag{12}$$

You might have suspected this connection. The Fourier transform of white noise is a delta function at $\tau = 0$. So, of course the correlation function for any white noise will be zero everywhere but the origin. To be complete, and get some practice with Fourier math, we will prove the W-K theorem. Again, though we use $n(t)$ the result is generally true.

$$\begin{aligned}
A(\tau) &= \int_{-\infty}^{\infty} n(t)n(t+\tau)dt & (13)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} N(f)\exp(-i2\pi ft)\, N(f')\exp(-i2\pi f'(t+\tau)dt\, df\, df' & (14)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} N(f)N(f')\exp(-i2\pi f'\tau)\, \exp(-i2\pi(f+f')t)dt\, df\, df' & (15)
\end{aligned}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} N(-f)N(f')\exp(-i2\pi f'\tau)\ \exp(-i2\pi(f'-f)t)dt\,df\,df' \quad (16)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} N(-f)N(f')\exp(-i2\pi f'\tau)\,\delta(f'-f)df\,df' \quad (17)$$

$$= \int_{-\infty}^{\infty} N(-f)N(f)\exp(-i2\pi f\tau)\,df \quad (18)$$

$$= \int_{-\infty}^{\infty} P'(f)\exp(-i2\pi f\tau)\,df \quad (19)$$