# Power Spectra

In the previous chapter we sort of off-handedly related the power $\mathcal{P}$ to the amplitude $\psi$ by:

$$\mathcal{P} \propto |\psi|^2$$

Here we will investigate power spectra more carefully. Our approach will be sort of a spiral, and each time around the cycle we increase our sophistication.

The contents of this chapter are:

## ■ Correlations and Autocorrelations

First we review and extend some ideas from much earlier. Recall that the convolution of two functions **f** and **g** is:

$$h[t] = f * g = \int_{-\infty}^{\infty} f[u] \, g[t - u] \, du$$

This is sometimes called the *cross-correlation function*, since it correlates the signal **f** at some time **i** with a second signal **g** at earlier times **i - j**. We have avoided using this term for reasons that are about to become obvious.

Similar to the *cross-correlation function* is the *correlation*:

$$Corr[f, g] \equiv \int_{-\infty}^{\infty} f[u] \, g[t + u] \, du$$

For a finite time series, normalisation becomes an issue since later terms of the correlation will have fewer members of the series **f** and **g** to add up. Thus:

$$\text{Corr}[f, g]_i = \sum_{j=0}^{m} f_j\, g_{i+j}$$

Consider two simple time series:

```
f = {f₀, f₁, f₂}
g = {g₀, g₁}
```

$$f = \{f_0, f_1, f_2\}$$
$$g = \{g_0, g_1\}$$

Then:

$$\text{Corr}[f, g] = \{(f_0\, g_0 + f_1\, g_1),\ f_0\, g_1\}$$
$$\text{Corr}[g, f] = \{(f_0\, g_0 + f_1\, g_1),\ (f_1\, g_0 + f_2\, g_1),\ f_2\, g_0\}$$

Thus the correlation operation is not commutative!

You will recall that the convolution of **f** and **g** is the inverse Fourier transform of the product of the Fourier transforms **F** and **G:**

$$f * g \leftrightarrow F[\omega]\, G[\omega]$$

Similarly for the correlation:

$$\text{Corr}[f, g] \leftrightarrow F[\omega]\, G[-\omega]$$

In this course we restrict ourselves to real input. Thus:

$$G[-\omega] = G^*[\omega]$$

This means that there is no additional information for the negative frequency components. We can now write:

$$\text{Corr}[f, g] \leftrightarrow F[\omega]\, G^*[\omega]$$

The *autocorrelation* is the correlation of a function with itself. Thus, the above relation becomes the "Wiener-Khinchin Theorem:"

$$\text{AutoCorr}[f] = \text{Corr}[f, f] \leftrightarrow |\,F[\omega]\,|^2$$

We wish to have the autocorrelation normalised so that the the magnitude of each term is independent of the number of values in it. Thus, the individual terms of the autocorrelation function of a time series are:

$$a[\tau] = a[i\,\Delta] = \text{AutoCorr}[f]_i = \frac{1}{m - i + 1} \sum_{j=0}^{m} f_j\, f_{i+j}$$

Often $\tau$ is called the *lag*.

In general the Fourier transform $\mathbf{F[\omega]}$ can be written as an amplitude times a phase:

$$\mathbf{F}[\omega] = \mathbf{A}[\omega]\, \mathbf{e}^{-i\,\phi[\omega]}$$

The *total* power then is:

$$\mathcal{P}_{\text{tot}} = \int_{-\infty}^{\infty} |\,\mathbf{F}[\omega]\,|^2\, d\omega = \int_{-\infty}^{\infty} |\,\mathbf{A}[\omega]\,|^2\, d\omega$$

Note that the total power then is independent of the phase. The total power is the same whether we calculate it in the frequency domain or the time domain:

$$\mathcal{P}_{\text{tot}} = \int_{-\infty}^{\infty} |\,f[t]\,|^2\, dt$$

The above result is known as *Parseval's theorem.*

## ■ Power and the Z Transform

So far we are assumign that the power at a frequency $\omega$ is:

$$\mathcal{P}[\omega] = |\,\mathbf{F}[\omega]\,|^2 = \mathbf{F}^*[\omega]\, \mathbf{F}[\omega]$$

or in terms of the Z Transform:

$$\mathcal{P}[z] = \mathbf{F}^*[z]\, \mathbf{F}[z]$$

But since:

$$z \equiv \mathbf{e}^{-i\,\omega}$$

$$z^* = \frac{1}{z}$$

Thus if:

$$\mathbf{F}[z] = f_0 + f_1\, z + f_2\, z^2 + \ldots$$

$$F^*[z] = f_0^* + f_1^* \frac{1}{z} + f_2^* \frac{1}{z^2} + \ldots$$

Consider a Z Transform for which the three terms shown above are the only terms. Then:

$$F^*[z]\, F[z] = f_0\, f_2^* \frac{1}{z^2} + (f_0\, f_1^* + f_1\, f_2^*)\, \frac{1}{z} +$$
$$(f_0\, f_0^* + f_1\, f_1^* + f_2\, f_2^*) + (f_0^*\, f_1 + f_1^*\, f_2)\, z + f_0^*\, f_2\, z^2$$

Think of this as the Z Transform of some time series **r**:

$$r = \{r_{-2},\ r_{-1},\ r_0,\ r_1,\ r_2\}$$

If all the terms in the time series are real, so are the coefficients **f**, so:

$$r_{-n} = r_n$$

But if **F[z]** is just:

$$\{f0,\ f1,\ f2\}$$

then its autocorrelation is:

$$\text{AutoCorr}[F] = \{\ f_0^2 + f_1^2 + f_2^2,\ f_0\, f_1 + f_1\, f_2,\ f_0\, f_2\}$$

But these are just the terms:

$$\{r_0,\ r_1,\ r_2\}$$

Thus, as we suggested in the previous section, there is a close relation between the power spectrum and the autocorrelation function. We shall explore this further soon.


## ■ Stochastic Processes

As our undergraduate laboratories keep telling students over and over and over, virtually every physical measurement has an uncertainty associated with it. Thus if one measures, for example, a time series of the voltage from a signal generator and then repeats the measurement, the two sets of numbers will differ somewhat from each other. Nonetheless, the *average* behavior of repeated measurements can be similar.

For example, here is data on the number of female births per day in California in 1959:

Female Births

(Source: http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/)
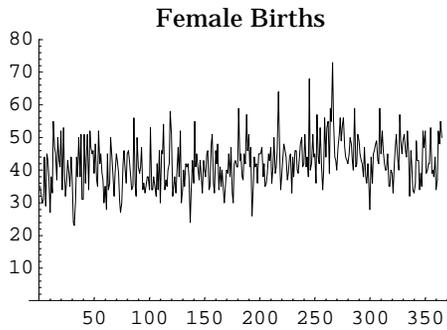
The possible values for the time series at a particular time is a *random* variable **X[t]**, which has an associated probability distribution. The observed value **x[t]** at that time is then in general only one of an infinity of possible values it might have taken.

For each time $t_i$ there is a probability function $P_i$ and a random variable $X_i$. Thus for times $\{t_0, t_1, t_2, \ldots, t_n\}$ there is an ordered list of the probabilities and another ordered list of random variables, and these two lists completely describe the statistical properties of the series. This is called a *stochastic process*.

Given $P_i$ one may generate an ensemble of $X_i$ using Monte Carlo techniques; a brief introduction to this topic appears at http://www.upscale.utoronto.ca/GeneralInterest/Harrison/MonteCarlo.pdf.

In the birth data above, there is some hint that the births per day is increasing throughout the year. This is called a *trend*, and such data is said to be *non-stationary*.

Say you are attempting to determine a relation between the number of births and the time of year. Then you will wish to remove any trends before doing further analysis. A simple way to *de-trend* the birth data might be to fit it to a straight line:

```
Births = m * Day + b
```

and then adjust the number of births:

```
adjustedBirths = Births - m * Day
```

If you are looking for periodic structures in the birth data, you might also consider subtracting the intercept **b** from the data so that it oscillates about zero. Here is the result of these two operations on the birth data:

adjustedBirths
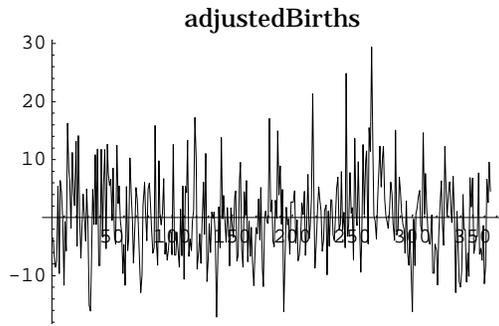
Here is the autocorrelation of this de-trended birth data:



Autocorrelation of births
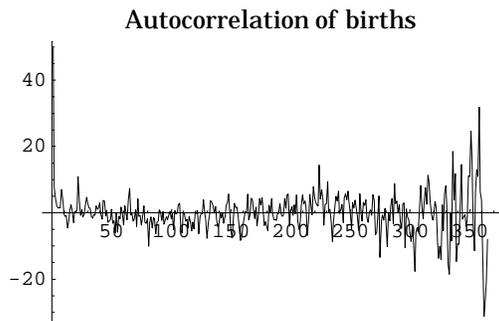
The above plot tells us there is little if any correlation between the number of births on a given day with the number of births on later days, although there may be some signs of a yearly cycle.

For comparison, here is some data on the brightness of a variable star on 600 successive nights:



Variable star brightness
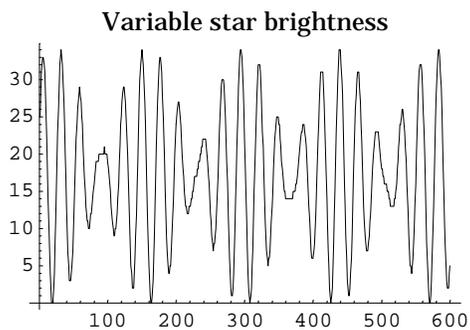
After de-trending with a straight line fit, the autocorrelation looks like this:

Autocorrelation of star brightness

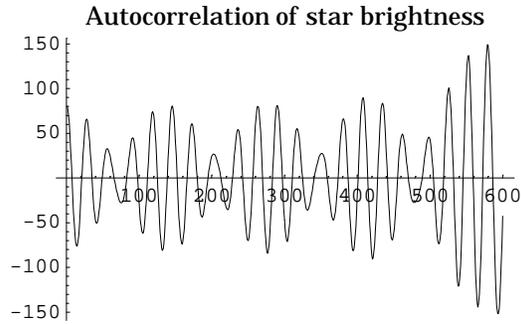This shows clearly that the value of the brightness at later times depends on the value at earlier times. This is no surprise, since the brightness is clearly periodic.

Often we assume that the probability distribution in a stochastic process is *normal*, so that if we repeated a particular measurement enough times a histogram of the results would be a Gaussian. Then the probability distribution $P_X[\mathbf{x}]$ can be completely characterised by its mean and its standard deviation. The mean $\bar{\mathbf{x}}$ is the expectation value of $\mathbf{X}$, $<\mathbf{X}>$, and is given by:

$$\bar{\mathbf{x}} \; = \; < \mathbf{X} > \; = \; \int_{-\infty}^{\infty} \mathbf{x}\, P_X[\mathbf{x}]\; d\mathbf{x}$$

and the standard deviation $\sigma$ is calculated from its square, the *variance*:

$$\sigma^2 = \; < (\mathbf{X} - \bar{\mathbf{x}})^2 > \; = \; \int_{-\infty}^{\infty} (\mathbf{x} - \bar{\mathbf{x}})^2\, P_X[\mathbf{x}]\; d\mathbf{x}$$

This assumption of a normal distribution, or any other probability function, is usually only approximate. For example, there were 44 female births in California on January 5, 1959. If the probability function is a Gaussian, then there is a small but non-zero probability that there could have been, say, 5017 births on that date; this is clearly unreasonable. For this particular data there is a further problem with assuming a normal distribution since the number of births is constrained to be an integer. Thus, when I earlier wrote that a datum is "in general only one of an infinity of possible values it might have taken" we conclude that to be too strong a statement.

Another common probability function is the *Poisson* distribution; this distribution governs, for example, radioactive decay. For this distribution, the standard deviation is the square root of the mean value. Consider again the 44 births on January 5. The previous day there were 31 births; the following day there were 29 births. If we were to estimate the uncertainty in the number of births on January 5, we should take the square root of what number? The point is that we don't know what the "mean value" is in this case.

Some people make a distinction between *experimental* and *non-experimental* data. The former is the result of some data collection which at least in principle is repeatable. The birth data are non-experimen-

tal, since we can't repeat January 5, 1959. However, even for experimental data one seldom has the time or resources to repeat measurements many many times. And without many many repeated measurements we don't have much of a handle on what the actual mean value is.

So, our best guess of the uncertainty in the 44 births on January 5 is $\sqrt{44} = 6.6 \simeq 7$. In fact for *any* number a reasonable guess of its uncertainty is often its square root. For example, as of March 2, 1999 Mats Sundin of the Toronto Maple Leafs has scored 22 goals; some of these goals were lucky and he also had bad luck and missed some opportunities. Thus, we expect he could have easily scored as few as $22 - \sqrt{22} \simeq 17$ goals or as many as $22 + \sqrt{22} \simeq 27$.

You should be aware that there is a further refinement of the sorts of statistical analysis we have been describing, called *Bayesian* after Reverend Thomas Bayes (1702-1761). Essentially Bayesian analysis divides the probability distribution into two parts, the purely statistical fluctuations from nature and the uncertainties in our knowledge of what those fluctuations are. The Bayesians are somewhat of a cult, and their detractors tend to use words like *voodoo* in describing this analysis. We will not be discussing Bayesian analysis in this course, but further information is available from the **International Society for Bayesian Analysis** web site at http://www.bayesian.org/.

## ■ Power: Once More Around the Cycle

We have just argued that almost all real physical processes are stochastic. This means that when we talk about the power being related to the Fourier transforms we should be talking about the *expectation value* of those transforms:

$$\mathcal{P}[\omega] \;=\; <\; \mathbf{F}^*[\omega]\; \mathbf{F}[\omega]\; >$$

Further, we don't measure real time series for infinitely long periods of time, but instead some time interval **T**. So for a time series of a stochastic time series the power becomes:

$$\mathcal{P}[\omega] \;=\; \lim_{\mathbf{T} \to \infty}\; \frac{1}{\mathbf{T}}\; <\; \mathbf{F}_{\mathbf{T}}^*[\omega]\; \mathbf{F}_{\mathbf{T}}[\omega]\; >$$

If we have measured for an infinitely long period of time, then the statistical variations will cancel out.

Since we have in general **N** points in such a time series, where **T** equal **N∆**:

$$\mathcal{P}[\omega] \;=\; \lim_{\mathbf{N} \to \infty}\; \frac{1}{\mathbf{N}}\; <\; \mathbf{F}_{\mathbf{N}}^*[\omega]\; \mathbf{F}_{\mathbf{N}}[\omega]\; >$$

Consider:

$$< F_T^*[\omega_1] \ F_T[\omega_2] \ >$$

This equals:

$$\int_{-T/2}^{T/2} \int_{-T/2}^{T/2} < f[t_1] \ f[t_2] > \ e^{i \ (\omega_1 \ t_1 - \omega_2 \ t_2)} \ dt_1 \ dt_2$$

Note we do not need to worry about expectation values for the exponential, since it is a numerical quantity.

But $< f[t_1] \ f[t_2] >$ is the autocorrelation function of **f**, **a[$\tau$]**, for a lag $\tau = t_1 - t_2$.

Note that **a[$\tau$]** equal **a[-$\tau$]**: the autocorrelation is symmetric.

Write:

$$t_1 \ = \ \tau \ + \ t_2$$

and:

$$< F_T^*[\omega_1] \ F_T[\omega_2] \ > \ = \ \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} a[\tau] \ e^{i\omega_1 \ \tau} \ e^{i \ (\omega_1 - \omega_2) \ t_2} \ d\tau \ dt_1 \ =$$

$$\int_{-T/2}^{T/2} a[\tau] \ e^{i\omega_1 \ \tau} \ d\tau \int_{-T/2}^{T/2} e^{i \ (\omega_1 - \omega_2) \ t_2} \ dt_2$$

The second integral above is just:

$$\frac{2 \ \text{Sin}[\frac{(\omega_1 - \omega_2) \ T}{2}]}{(\omega_1 - \omega_2)}$$

This is, of course, a Sinc function, which we saw much earlier was one way of representing a Dirac delta function. The maximum height of this fucntion is **T**, and the width is **1/T** in Hz, or **2$\pi$/T** in radians/sec. Thus, the "area" under the curve is **2$\pi$** and we write:

$$\int_{-T/2}^{T/2} e^{i \ (\omega_1 - \omega_2) \ t_2} \ dt_2 \ = \ 2 \ \pi \ \delta[\omega_1 \ - \ \omega_2]$$

We shall use the result for $< F_T^*[\omega_1] \ F_T[\omega_2] \ >$ to get two important results. First, let $\omega_1 = \omega_2$. Then the second integral:

$$\int_{-T/2}^{T/2} e^{i \ (\omega_1 - \omega_2) \ t_2} \ dt_2$$

is just **T** and:

$$< F_T^*[\omega] \, F_T[\omega] \, > \; = \; T \int_{-T/2}^{T/2} a[\tau] \, e^{i\omega\tau} \, d\tau$$

Thus, the power is:

$$\mathcal{P}[\omega] \; = \; \lim_{T \to \infty} \frac{1}{T} < F_T^*[\omega] \, F_T[\omega] \, > \; = \; \int_{-\infty}^{\infty} a[\tau] \, e^{i\omega\tau} \, d\tau$$

Above we called this relation the "Wiener-Khinchin Theorem."

The second case we want to consider is when $\omega_1 \neq \omega_2$. Then the integral:

$$\int_{-T/2}^{T/2} e^{i \, (\omega_1 - \omega_2) \, t_2} \, dt_2 \; = \; 0$$

so we have proved that:

$$\lim_{T \to \infty} < F_T^*[\omega_1] \, F_T[\omega_2] \, > \; = \; 0$$

Note that the Wiener-Khinchin Theorem says that for power spectra, just as for filters, we can work in the frequency domain:

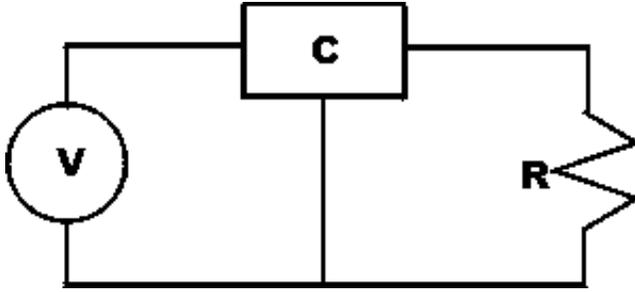$$\mathcal{P}[\omega] \; = \; \lim_{T \to \infty} \frac{1}{T} < F_T^*[\omega] \, F_T[\omega] \, >$$

or we can work in the time domain:

$$\mathcal{P}[\omega] \; = \; \lim_{T \to \infty} \int_{-T/2}^{T/2} a[\tau] \, e^{i\omega\tau} \, d\tau$$

$$a[\tau] \; \equiv \; < f[t] \, f[t - \tau] \, >$$

■ **An Example**

We close this section with an example.

We have a voltage source that includes noise in the output, so it is stochastic. We filter it with a narrow band filter **C**, and then will measure the power dissipated by the resistor **R**.

We will assume that the filter is:

$$C[\omega] \; = \; 1 \quad \omega_0 \; - \; \Delta\omega \; \leq \; \omega \; \leq \; \omega_0 \; + \; \Delta\omega$$

and is zero otherwise. Let the voltage from the source be $v_{in}$ and the voltage from the filter, which is the voltage across the resistor be $v_{out}$. Now:

$$v_{in}[t] \; = \; \frac{1}{2\,\pi} \int_{-\infty}^{\infty} V[\omega] \; e^{i\,\omega t} \; d\omega$$

$$v_{out}[t] \; = \; \frac{1}{2\,\pi} \int_{-\infty}^{\infty} C[\omega] \; V[\omega] \; e^{i\,\omega t} \; d\omega$$

Thus the power through the resistor is:

$$\mathcal{P}_R \; = \; \frac{< \, v_{out}[t]^2 \, >}{R} \; =$$

$$\frac{1}{4\,R\,\pi^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C^*[\omega_1] \; C[\omega_2] \; < V^*[\omega_1] \; V[\omega_2] > e^{i\,(\omega_2 - \omega_1)\,t} \; d\omega_1 \, d\omega_2 \; =$$

$$\frac{1}{4\,R\,\pi^2} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C^*[\omega_1] \; C[\omega_2] \; \left( 2\,\pi\,\delta[\omega_1 \, - \, \omega_2] \int_{-\infty}^{\infty} a[\tau] \; e^{i\,\omega_1\,\tau} \; d\tau \right)$$

$$e^{i\,(\omega_2 - \omega_1)\,t} \; d\omega_1 \, d\omega_2 \; = \; \frac{1}{2\,R\,\pi} \int_{-\infty}^{\infty} \left| \, C[\omega] \, \right|^2 \, \mathcal{P}[\omega] \; d\omega$$

But the filter **C** is non-zero only for frequencies $\omega_0 \; - \; \Delta\omega \; \leq \; \omega \; \leq \; \omega_0 \; + \; \Delta\omega$. Thus:

$$\mathcal{P}_R \; = \; \frac{1}{2\,R\,\pi} \int_{\omega \, - \, \Delta\omega}^{\omega \, + \, \Delta\omega} \mathcal{P}[\omega] \; d\omega$$

If $\Delta\omega$ is small, then this becomes:

$$\mathcal{P}_R \; \simeq \; \frac{1}{2\,R\,\pi} \, \mathcal{P}[\omega] \; \Delta\omega \; = \; \frac{1}{2\,\pi\,R} \, \mathcal{P}[\omega \, = \, 2\,\pi\nu] \; 2\,\pi\,d\nu \; = \; \frac{1}{R} \, \mathcal{P}[\nu] \; d\nu$$

# ■ Power and Stochastic Processes

In the previous section, we generalised to stochastic processes only in that we changed all the terms involving the time series and its Fourier transform into expectation values. Here we look into this more carefully. We will be working throughout with true time series, for which of course:

$$F_j = \left( \sum_{k=0}^{n-1} f_k \, E^{-I \, 2 \, \pi \, j \, k / n} \right) \Delta = \sum_{k=0}^{n-1} f_k \, E^{-I \, 2 \, \pi \, j \, k / n}$$

$$f_k = \left( \frac{1}{n} \sum_{j=0}^{n-1} F_j \, E^{I \, 2 \, \pi \, j \, k / n} \right) \frac{1}{\Delta} = \frac{1}{n} \sum_{j=0}^{n-1} F_j \, E^{I \, 2 \, \pi \, j \, k / n}$$

$$t_k = k \Delta = k$$

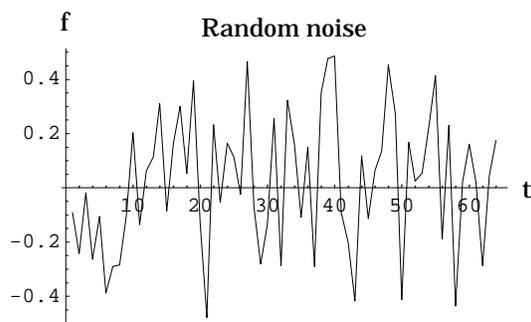$$\omega_j = j \left( \frac{2 \, \pi}{n \Delta} \right) = j \left( \frac{2 \, \pi}{n} \right)$$

$$T = n \Delta = n$$

where the second form is for the usual choice that $\Delta = 1$.

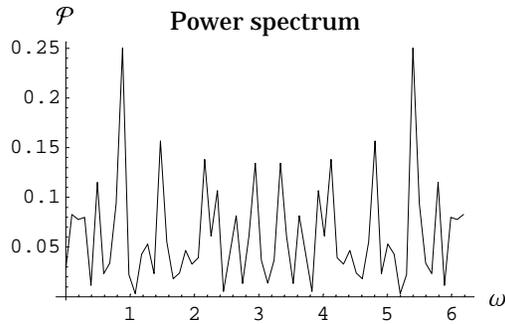The power for frequency **j** is:

$$\mathcal{P}_j = \mathbf{Re}[F_j]^2 + \mathbf{Im}[F_j]^2$$

*White noise* is noise that has a flat frequency power spectrum. Thus, if the time series looks like:



the the power spectrum looks like:

There are a couple of problems with the above plot. First, the wide variation in the values of the power make it hard to see what is going on. Thus, one commonly plots the logarithm of the power. The second problem is exactly what we meant by saying that white noise has a flat frequency spectrum. What we mean is that there is equal power in each *octave*. As you probably know, the octaves are also a logarithmic scale: for example concert A is 440 Hz, the note A that is one octave higher is 880 Hz, the note A one octave higher than that is 1760 Hz, the next octave begins at 3520 Hz, etc. Thus, we commonly display power spectra as a log-log plot:
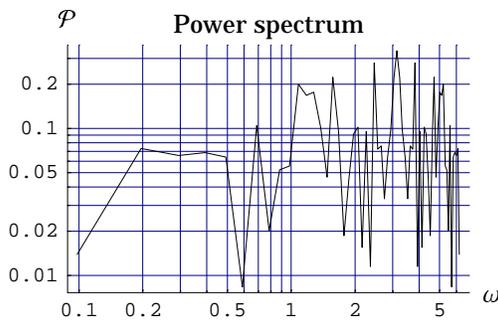


Using the logarithm of the power is equivalent to measuring the power in a decibel scale:

$$\beta \equiv 10 \log_{10} \frac{\mathcal{P}}{\mathcal{P}_0}$$

where $\mathcal{P}_0$ is the power of some standard reference level. For sound intensities in W/$m^2$ one decibel is roughly the minimum limit of human audibility. The unit is named for Alexander Graham Bell.

The standard deviation in the power for a frequency $\omega_j$ is $\mathcal{P}_j/\sqrt{2}$. This is bad news: the uncertainty is about the size of the value itself! To see why, consider the *drunkard's walk*:
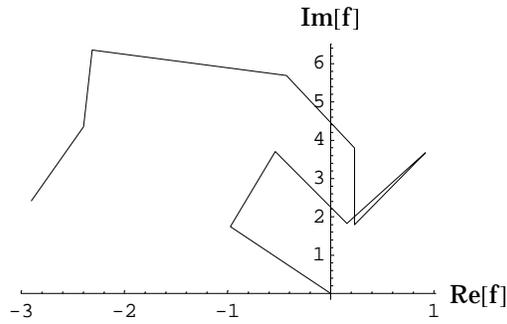
We start at the origin and take a step of length **f** in a random direction. As you probably know, the average distance we end up from the origin after **n** steps is:

$$\bar{\mathbf{f}} = \sqrt{\mathbf{n}}\ \mathbf{f}$$

Thus,

$$< \left| \mathbf{F_j} \right|^2 > \ \propto \ \bar{\mathbf{f}}^2 = \mathbf{n}\, \mathbf{f}^2$$

We can not improve the situation by sampling for longer intervals and/or changing the sampling interval. If we sample at a higher value of **n** with the same sampling interval, the Nyquist frequency is unchanged but we have more frequency components between 0 and the critical frequency. If we use a smaller sampling interval with the same total number of points in the time series, our frequency resolution is not changed but the Nyquist frequency is larger.

If experimentally possible, we can repeat the measurement of the time series, say, **m** times. This will reduce the error by $\sqrt{m}$.

Other times, it is not possible or practical to repeat the measurement. There are still some things that can be done, however.

One possibility it to partition the time series into **m** series each of length **n/m**. Each segment is separately Fourier transformed. Note that the **j**-th element of each Fourier transform corresponds to the same frequency. Thus we can form an average of the **m** different Fourier transforms, which will reduce the error by $\sqrt{m}$. This is sometimes called "ensemble averaging" or *Barlett's method*.

Another possibility is to average **m** adjacent frequencies. This reduces the resolution in frequency, since you end up with **n/m** different frequency values, each corresponding to the mid-point of the frequencies used in computing each average. This technique, sometimes called "frequency averaging" or *Danielli's method*, is particularly useful if the original time series has a discrete frequency spacing greater than what we really need.

Both ensemble and frequency averaging sacrifice frequency resolution to reduce the standard deviation of the power. However, recall that the algorithm to do discrete Fourier transforms on a time series of length **n** has a speed proportional to:

$$O \ (\texttt{n} \ \texttt{ln}_2 \ \texttt{n})$$

For ensemble averaging we take **m** different Fourier transforms of time series of length **n/m** while for frequency averaging we take one Fourier transform of the entire time series of length **n**. It turns out that ensemble averaging is a bit more efficient.

Finally, imagine that the power spectrum is not "white:"
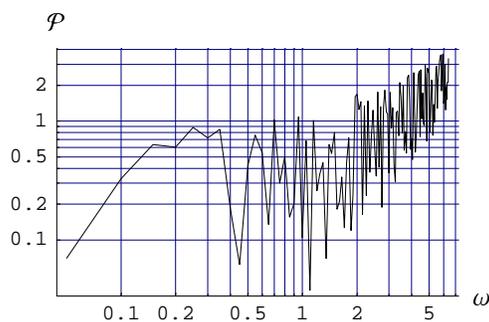


The above spectrum is increasing with increasing frequency. Such a spectrum is called "blue" since large frequencies correspond to small wavelengths, and for light small wavelengths are the blue-violet end of the spectrum. If one wishes to "whiten" the spectrum, one common technique is to do something similar to when we were detrending a time series: we replace each value of the power by the slope. But, the derivative at each point is approximately:

$$\frac{d \mathcal{P}}{d \omega} \bigg|_j \ \simeq \ \frac{\mathcal{P}_{j+1} - \mathcal{P}_j}{\Delta \omega}$$

If we replace each value for the power $\mathcal{P}_j$ by $\mathcal{P}_{j+1} - \mathcal{P}_j$ the result is:
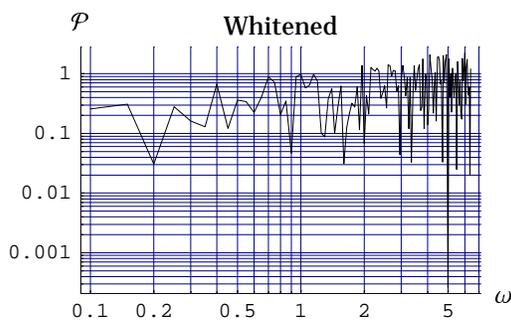
## ■ Leakage

In the **Sampling & Windowing** chapter we discussed the fact that if we start a time series at time zero and stop it at time **t Δ**, this is equivalent to windowing the time series with a square window. Since the Fourier transform of this window is a *Sinc* function, this means that the value of the transform at some frequency **j** contains components that are artifacts of the fact that we have chopped the time series at small and large times. Since this is true of the Fourier transform, it is also true of the power spectrum. This is usually called *leakage*.

The solution, just as when we first discussed windowing, is to smooth the abrupt transitions at the edges of the time series by multiplying it by a window function. Here, as before, common choices for windows include Hanning, Hamming, Welch and Parzen.

## ■ Power in the Time Domain

We saw above that, just as for filters, we can calculate the power in the frequency domain or the time domain. In the time domain the power is related to the autocorrelation:

$$\mathcal{P}[\omega] = \lim_{T \to \infty} \int_{-T/2}^{T/2} a[\tau] \, e^{i \omega \tau} \, d\tau$$

$$a[\tau] \equiv \, < f[t] \, f[t - \tau] >$$

Now, the autocorrelation is an even function of the lag $\tau$, so the integral is also equal to:

$$\mathcal{P}[\omega] = \lim_{T \to \infty} \int_{-T/2}^{T/2} a[\tau] \, e^{-i \omega \tau} \, d\tau$$

This makes it clear that the power is just the Fourier transform of the autocorrelation function.

For a real time series, we weight each term in the autocorrelation by the total number of elements used in its computation:
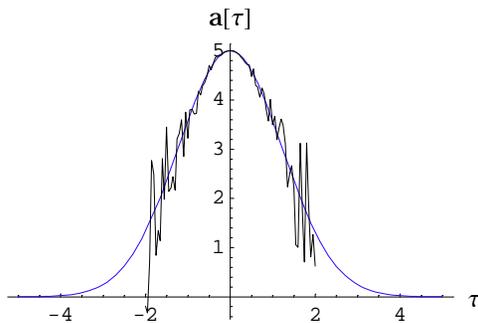
$$a_0 = \frac{1}{n} \sum_{k=0}^{n-1} f_k^2$$

$$a_1 = \frac{1}{n - 1} \sum_{k=0}^{n-2} f_k \, f_{k+1}$$

$$a_2 = \frac{1}{n-2} \sum_{k=0}^{n-3} f_k\, f_{k+2}$$

$$a_n = \frac{1}{1}\, f_0\, f_{n+1}$$

If the time series is stochastic, this means that in general the terms in the autocorrelation for higher lags will exhibit much more noise. Below we show an example where the theoretical autocorrelation is a Gaussian, but the measured autocorrelation exhibits noise for large lags.



The solution may well be obvious to you by now: window the autocorrelation to suppress the longer lag times.

$$a_k' = w_k\, a_k$$

## ■ Maximum Entropy Method

In the **Stochastic Processes** section of this chaper, we referred to Bayesian statistical analysis. Here we discuss a technique for power spectra estimation inspired by this approach.

We write once more the power as the square of the Fourier transform, but this time write the transform in terms of the Z transform:

$$\mathcal{P}_\omega \simeq \left| \sum_{k=0}^{n-1} f_k\, z^k \right|^2$$

where the approximation emphases that this is only an estimate for a finite number **n** of frequency components, and:

$$z \equiv e^{-i\omega}$$

Although formally identical to the previous representation in terms of $\mathbf{F}[\omega]$, in this guise it is called various the *direct method*, the *moving average (MA) method*, or the *all-zero model*. The reason for the latter name is because the power can have zeroes in the $\mathbf{z}$ plane but no poles.

Imagine we are dealing with a power spectrum, such as the light intensity from a pure element lamp, that has "lines" at discrete frequency values. In the above representation we would have to try to represent those lines in terms of a large number of frequency components that add up to approximate the line shape.

However, just as when we were building inverse filters by taking the reciprocal of the Z transform, we can represent the power as:

$$\mathcal{P}_\omega = \mathcal{P}_j \simeq \left| \frac{1}{\sum_{k=0}^{n-1} d_k \, z^k} \right|^2 = \left| \frac{b_0}{1 + \sum_{k=1}^{n-1} b_k \, z^k} \right|^2$$

This representation clearly can have explicit poles. The approximation is called the *all-poles model*, or the *maximum entropy method*.

Recall the autocorrelation:

$$a_i = \frac{1}{n - i + 1} \sum_{k=0}^{m} f_k \, f_{k+i}$$

whose Fourier transform is also an estimate of the power spectrum. We write that Fourier transform as a Z transform and equate to the all poles model:

$$\sum_{i=0}^{n-1} a_i \, z^i \simeq \left| \frac{b_0}{1 + \sum_{k=1}^{n-1} b_k \, z^k} \right|^2$$

You will recall from our discussion of filters and inverse filters that in general the right hand side of the above equation will have a large number of terms, and in fact as $\mathbf{n}$ decreases the number of terms will increase. However, the left hand side has exactly $\mathbf{n}$ terms. Thus, the approximate equal sign in the above must be interpreted to mean that some terms on the right hand side will correspond to its counterpart on the left hand side, but there will also be additional terms on the right hand side with no corresponding part on the left.

Thus, the right hand side of the above relation defines a set of extrapolations of the autocorrelation function, but also extrapolates to lags greater than the original time series can actually measure!

Now we shall explore why the word *entropy* appears above. As you may already know, entropy is a measure of the *disorder* of a physical system; it is also a measure of the *probability* of occurence of a

particular configuration. For example, imagine we have two black marbles and two white marbles in a box that forces exactly two of the marbles to be on the left and two on the right. There are three possible ways for the marbles to be arranged: all black on the left and all white on the right, one white and one black on each side of the box, or all white on the left and all black on the right.
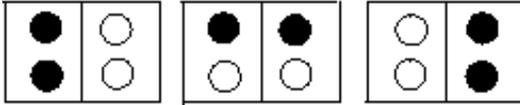


There is only one combination of the marbles to produce the configuration on the left, and similarly only one combination of marbles to produce the configuration on the right. But there are *four* possible combinations that can produce the middle arrangement. The entropy is:

**S ≡ k ln (number of combinations)**

Thus the middle arrangement has the highest entropy.

If all possible arrangement of the marbles are equally likely, subject to the constraint that there are two and only two marbles on each side of the box, then the probability of each of the three posssible arrangements is:

$$P = \frac{\textbf{number of combinations}}{\textbf{total combination}}$$

so the entropy is related to the probability according to:

**S = k ln ( P * total combinations)**

Thus the highest entropy state is the one with the highest probability.

We also say that the arrangement where we have a black marble and a white marble on each side of the box contains less *order* than when there is complete differentiation between the two sides.

In information theory, the entropy plays a crucial role. If we have a signal containing a lot of entropy, it has a lot of noise. But good information has a high signal to noise ratio, which means it should have low entropy. In fact, many people talk about the *negentropy*, which is just the negative of the entropy. Thus a good information system has high negentropy.

Recall that for the Bayesians, probability distributions are divided into two parts, the statistical fluctuations from nature and our uncertainties in our knowledge of what those fluctuations are. Given a set of constraints and an ensemble of possible probability distributions, the most un-biased estimate we can make of the correction probability distribution is the one containing the least information, i.e. the one with the maximum entropy. In this way we pick the distrubtion that gives us the least additional informa-

tion beyond the constraints themselves. Often the principle is credited to E.T. Jaynes in the 1950's, although Gibbs used it in the nineteenth ceentury.

Above we realised that:

$$\left| \frac{b_0}{1 + \sum_{k=1}^{n-1} b_k\, z^k} \right|^2$$

defines a set of extrapolations of the autocorrelation function. It turns out that this particular extrapolation has, of all possible extrapolations, the one with the maximum entropy.

As with all things Bayesian, admirers of the principle of maximum entropy apply it not only to time series, but also to linguisitic analysis of natural languages, techniques of image restoration, and a great deal else. Detractors dismiss it all as a cult.

In order to actually solve:

$$\sum_{i=0}^{n-1} a_i\, z^i \simeq \left| \frac{b_0}{1 + \sum_{k=1}^{n-1} b_k\, z^k} \right|^2$$

it turns out that the equation can be re-written as:

$$\begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_m \\ a_1 & a_0 & a_1 & \cdots & a_{m-1} \\ a_2 & a_1 & a_0 & \cdots & a_{m-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_m & a_{m-1} & a_{m-2} & \cdots & a_0 \end{pmatrix} \begin{pmatrix} 1 \\ b_1 \\ b_2 \\ \cdots \\ b_m \end{pmatrix} = \begin{pmatrix} b_0 \\ 0 \\ 0 \\ \cdots \\ 0 \end{pmatrix}$$

The matrix has the same elements along all its diagonals; it is called a *Toeplitz* matrix. Efficient algorithms to solve the above equation have been given by Burg and also by Anderson; see Press et al., **Numerical Receipes**, Section 12.8 for details.


## ■ Authors

This document is Copyright © 1999 Richard C. Bailey and David M. Harrison. It was prepared using *Mathematica,* which produced a PostScript file; *Adobe Distiller* was then used to prepare the PDF file. This is version 1.6 of the document, date (m/d/y) 04/07/99.