4 The Periodogram

4.1 Motivation

A basic idea in mathematics and statistics is to take a complicated object (such as a time series) and break it up into the sum of simple objects that can be studied separately, see which ones can be thrown away as being unimportant, and then adding what's left back together again to obtain an approximation to the original object. The periodogram of a time series is the result of such a procedure.

4.2 Sinusoidal Decomposition of a Time Series

Given a time series data set of length n (assumed to be even for now for convenience), it is possible to find cosines and sines of periods $n, n/2, \ldots, n/(n/2) = 2$ that when added together along with $\bar{x}$ gives the data set back again.

Mathematically, we have

$$x(t) - \bar{x} = \sum_{k=2}^{n/2+1} \gamma_k \left[a_k \cos(2\pi(t-1)\omega_k) + b_k \sin(2\pi(t-1)\omega_k) \right],$$

where $\omega_k = (k-1)/n$, γ_k is two unless $k = n/2+1$ when it is one, and a_k and b_k are the real and imaginary parts of the discrete Fourier transform $z(1), \ldots, z(n)$ of the data:

$$z(k) = \sum_{j=1}^{n} x(t) e^{2\pi i (t-1)\omega_k}, \quad k = 1, \ldots, n,$$

If we define

$$C_k^2 = a_k^2 + b_k^2, \qquad \phi_k = \arctan(b_k / a_k),$$

we can also write

$$x(t) - \bar{x} = \sum_{k=2}^{n/2+1} \gamma_k C_k \cos(2\pi(t-1)\omega_k - \phi_k).$$

Further,

$$\hat{R}(0) = \frac{1}{n} \sum_{t=1}^{n} (x(t) - \bar{x})^2 = \sum_{k=2}^{n/2+1} \gamma_k C_k^2.$$

There are similar formulas for when n is odd (see Theorem 1.4.2 of the text).

4.3 Definition of Periodogram

Def: A plot of nC_k^2 versus $\omega_k = (k-1)/n$ for $k = 1, \ldots, [n/2]+1$ is called the periodogram of a data set. The function

$$\hat{f}(\omega) = \frac{1}{n}\left|\sum_{t=1}^{n} x(t)e^{2\pi i(t-1)\omega}\right|^2, \quad \omega \in [0, .5],$$

and $\hat{f}(\omega) = \hat{f}(1-\omega)$ for $\omega \in [0.5, 1]$ is called the sample spectral density of the data set.

Note: It is customary to plot $\log(nC_k^2/\hat{R}(0))$ vs. ω_k rather than nC_k^2.

4.4 Interpreting the Periodogram

1. The periodogram is a very useful tool for describing a time series data set. We will see it is much more useful than the correlogram but it does require some training to interpret properly.

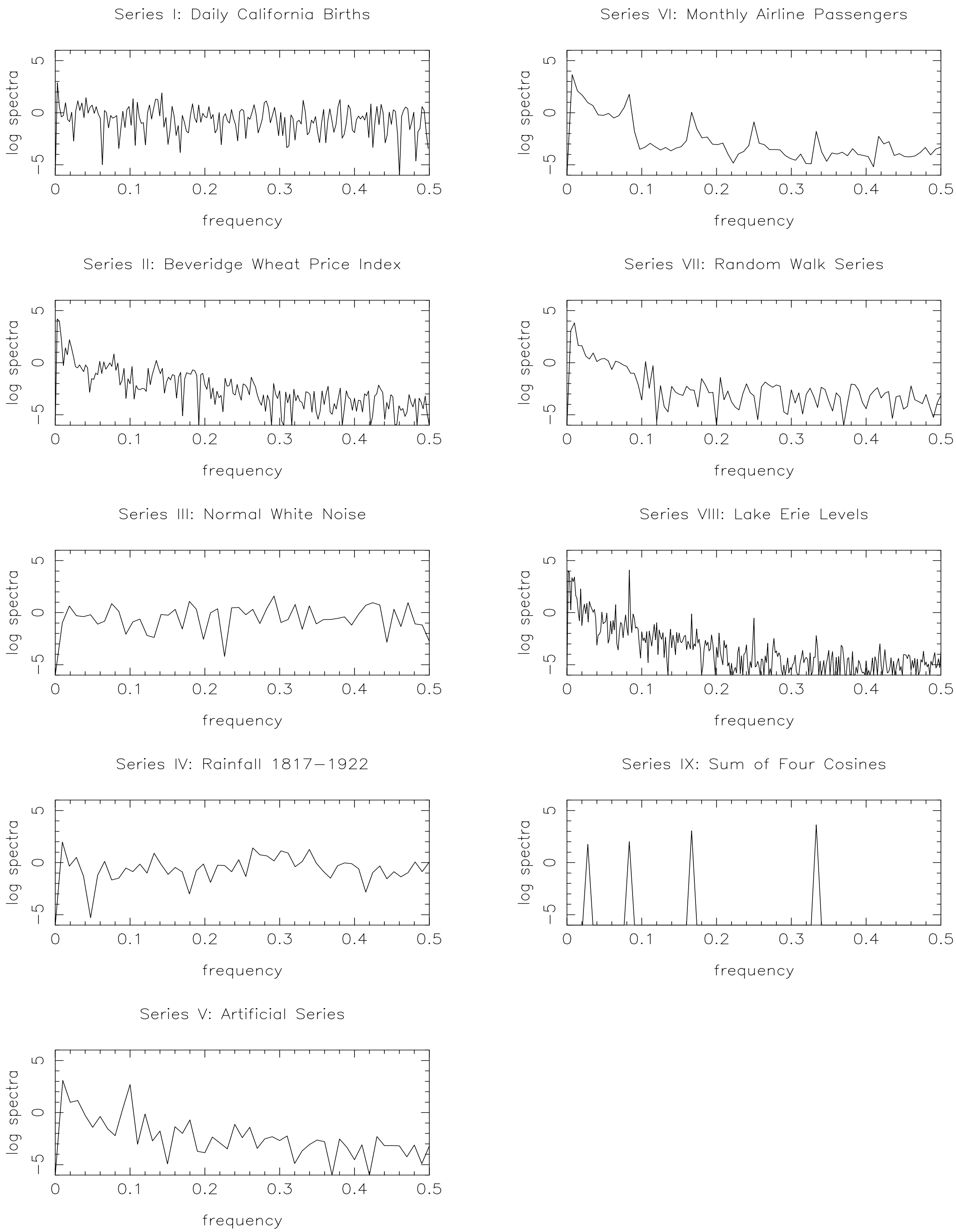
2. We will use the terms low frequency and high frequency extensively. The basic idea is that sinusoids of low frequency (or equivalently long period) are smooth in appearance whereas those of high frequency (or short period) are very wiggly. Thus if a time series appears to be very smooth (wiggly), then the values of the periodogram for low (high) frequencies will be large relative to its other values and we will say that the data set has an excess of low (high) frequency. For a purely random series, all of the sinusoids should be of equal importance and thus the periodogram will vary randomly around a constant.

3. If a time series has a strong sinusoidal signal for some frequency, then there will be a peak in the periodogram at that frequency.

4. If a time series has a strong nonsinusoidal signal for some frequency, then there will be a peak in the periodogram at that frequency but also peaks at some multiples of that frequency. The first frequency is called the fundamental frequency and the others called harmonics.

5. One of the aims of this course is to have you describe what a data set must look like when you can only see its periodogram (see the old exams for the course for examples).

4.5 Examples of the Periodogram



4.6 Cumulative Periodogram

A useful tool for describing the overall behavior of the periodogram (and thus the data set) is the cumulative periodogram

$$\hat{F}(\omega_k) = \frac{\sum_{j=1}^{k} \hat{f}(\omega_j)}{\sum_{j=1}^{q} \hat{f}(\omega_j)}, \quad k = 1, \ldots, q = [n/2] + 1.$$

Note that $\hat{F}$ starts out near zero at $\omega_1 = 0$ and must grow to be equal to one at $\omega_q = 0.5$ since we are accumulating quantities that are nonnegative.

Note that for a purely random series, $\hat{F}$ should follow along a line from $(0, 0)$ to $(0.5, 1)$. For a series having excess of low (high) frequency, $\hat{F}$ will start out above (below) that $y = 2x$ line, while there will be a jump in $\hat{F}$ at any frequency where $\hat{f}$ has a peak.

4.7 Examples of Cumulative Periodogram