Lecture 19-23

Spectral Analysis
Textbook Sections: Chapter 4

Two Domains

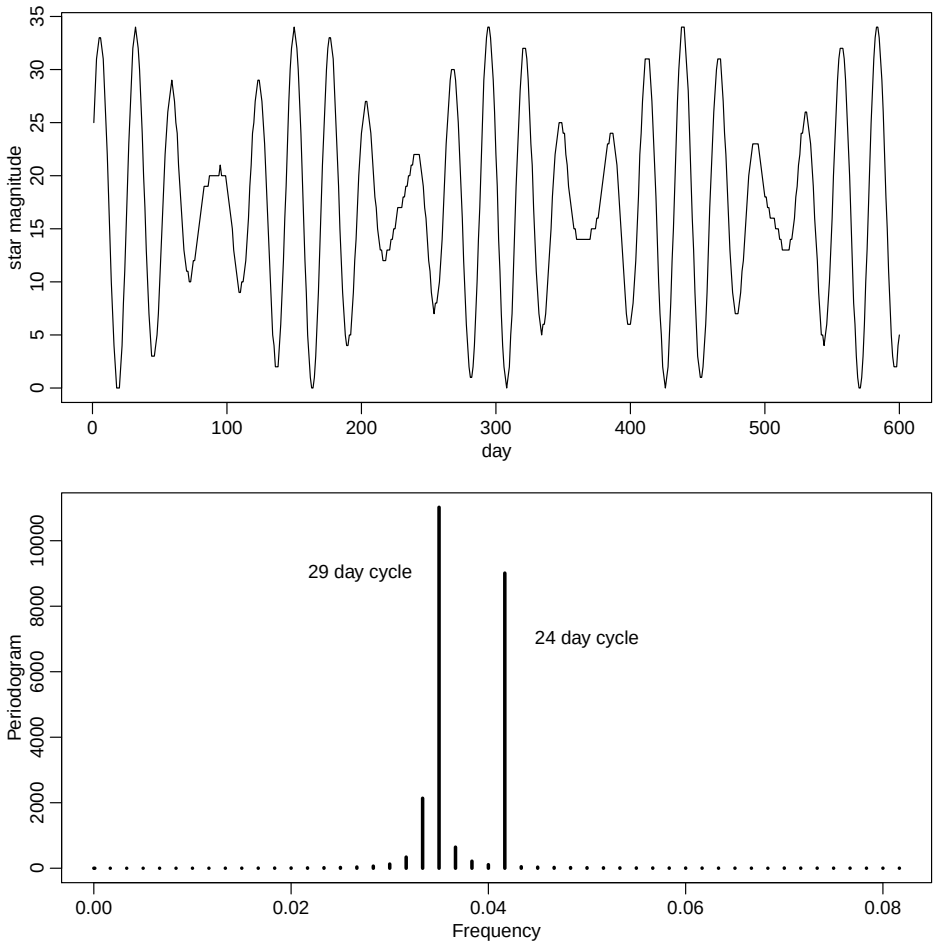
Everything we have covered so far breaks down the structure of $\{X_t\}$ in terms of time - lags, differences, ACVF and ACF, etc. This time series framework is known as the **time domain**.

Another approach to time series analysis is the **frequency domain**. In this framework, we decompose $\{X_t\}$ into uncorrelated random sinusoids of different frequencies. Then we analyze the decomposition. Knowing whether there are more high frequency or low frequency components can help us understand the general behavior of the data. Knowing if there are frequencies that dominate all others in the decomposition can help us figure out the periodic structure of the process. There is a corresponding decomposition of the ACVF.

Work in the **frequency domain** is concerned with the *spectrum* of frequencies present in a process, and is often called **spectral analysis**.

The two approaches are mathematically equivalent. However, they offer two different perspectives. Each of these is useful in some situations.

The next figure is a plot taken from *Time Series Analysis and Applications* by Shumway and Stoffer. It demonstrates the usefulness of studying the periodic structure of data. The R code for creating the figure is included in the appendix at the end of the notes.



Overview

The main objects of discussion will be the following. We will get familiar with these terms, their properties, and related applications.

- The **spectral density** of a process $\{X_t\}$ contains information about which frequencies are present in a process, and in what proportions. This is a population construct.
- The **(raw) periodogram** is the basic sample estimate of the spectral density. It is fairly easy to compute in practice, but it has some unfavorable theoretical properties.
- The **smoothed periodogram** is a modification of the periodogram with nicer theoretical properties.
- **Filtering** can be used to remove (filter out) certain frequencies.

Sinusoid Terminology

Since we'll work with sinusoidal functions, it's good to know the terminology. We will consider a general cosine function $Acos(2\pi\omega t + \phi)$. This is a function of variable t. The other parts are interpreted as follows.

- A is the amplitude. It determines how far the peaks and the troughs of the sinusoid reach. The function $cos(2\pi t)$ has phase $A = 1$, and its peaks reach to 1.
- ϕ is the phase. It determines how far the function is shifted horizontally. The function $cos(2\pi t)$ has phase $\phi = 0$, and has a peak at $t = 0$. For other values of ϕ, this peak will be shifted ϕ units to the left.
- ω is the frequency. It determines how fast the sinusoid oscillates, and how long the period is. The function $cos(2\pi t)$ has phase $\omega = 1$, and has a period of 1. Other values of ω would lead to a period of $1/\omega$.

Trig Identity

Recall the trigonometric identity $cos(x + y) = cos(x)cos(y) - sin(x)sin(y)$.
Let's apply it to $Acos(2\pi\omega t + \phi)$, setting $x = 2\pi\omega t$, and $y = \phi$.

$$\begin{aligned} A\cos(2\pi wt + \phi) &= A[\cos(2\pi wt)\cos(\phi) - \sin(2\pi wt)sin(\phi)] \\ &= \underbrace{A\cos(\phi)}_{U_1}\cos(2\pi wt)\underbrace{-A\sin(\phi)}_{U_2}\sin(2\pi wt) \\ &= U_1\cos(2\pi wt) + U_2\sin(2\pi wt) \end{aligned}$$

This form is preferable, as it makes some calculations easier.

Sinusoid with Random Coefficients

Since we are modelling random processes, we need to introduce randomness into our sinusoid components. We will continue with the above expression, but instead of constants $U_1 = Acos(\phi)$, and $U_2 = -Asin(\phi)$, we will use random coefficients A and B. This is equivalent to constructing sinusoids of a fixed frequency, but with random amplitude and phase. On the other hand, the constants U_1 and U_2 correspond to a fixed amplitude A and phase ϕ.

Define a sinusoid with random coefficients to be a random sequence $\{X_t\}$ of the form:

$$X_t = A\cos(2\pi wt) + B\sin(2\pi wt),$$

where A and B are *uncorrelated* random variables with mean 0 and common variance.

Spectral Density

The following four facts are meant to motivate the study of the spectral density function. We won't concern ourselves with the technical details. They are beyond the scope of this course. However, we will need to understand the meaning and importance of the below four facts. **This section is very important for our coverage of spectral analysis (frequency domain).**

1. Any stationary sequence (causal or non-causal) $\{X_t\}$ with zero mean can be approximated by a linear combinations of a large number m of sines and cosines with random coefficients,

$$X_t \approx \sum_{j=1}^{m} \{A_j \cos(2\pi w_j t) + B_j \sin(2\pi w_j t)\}, \tag{1}$$

where m is large, w_j's are distinct frequencies, $\{A_j\}$ and $\{B_j\}$ are random with zero means and $Var(A_j) = Var(B_j) = \sigma_j^2$. Moreover, $\{A_j\}$ and $\{B_j\}$ are all uncorrelated with each other. We will often take m to be equal to $n/2$ where n is the length of the observed series.

- If a sequence has mean μ, then the centered sequence $\{X_t - \mu\}$ has the above expression.
- In some trivial cases, m may be small. In general, m is large.

2. The approximate variance and the autocovariances of a stationary sequence $\{X_t\}$ are

$$Var(X_t) \approx \sigma_1^2 + \cdots + \sigma_m^2,$$

$$\gamma(h) \approx \sum_{j=1}^{m} \sigma_j^2 \cos(2\pi w_j h).$$

As a good exercise, verify this using the above expression.

Note that σ_j^2 determines the contribution of variability at frequency ω_j to the total variability.

It is useful to find out which frequencies contribute more than others. For this reason, we want to estimate the σ_js.

3. It can be shown mathematically (under certain technical conditions), that there is a function $f(\omega)$ on $[-0.5, 0.5]$ which is symmetric about zero, such that $\sigma_j^2 \approx 2f(j/n)/n$ when m is taken to be equal to $n/2$, where n is the length of the observed series. This function f is called the **spectral density function**.

Note that $f(j/n)$ is the value of the spectral density function at frequency $\omega = j/n$ and it is a rescaled version σ_j^2. Estimating the values of the spectral density function at frequencies j/n for $j = 1, 2, \cdots, m$ is equivalent to estimating the σ_js.

4. For a given stationary, zero-mean process $\{X_t\}$, the spectral density $f_X(\omega)$ $(-1/2 \leq \omega \leq 1/2)$ and the autocovariance function $\gamma_X(h)$ $(h = 0, \pm 1, \pm 2, \cdots)$ are related in the following ways.

- For any positive integer h,

$$\gamma_X(h) = \int_{-1/2}^{1/2} \exp(2\pi i \omega h) f_X(\omega) d\omega = 2 \int_0^{1/2} \cos(2\pi i \omega h) f_X(\omega).$$

- For any $-1/2 \leq \omega \leq 1/2$,

$$f_X(\omega) = \sum_{h=-\infty}^{\infty} \gamma_X(h) \exp(-2\pi i \omega h) = \gamma_X(0) + 2 \sum_{h=1}^{\infty} \gamma_X(h) \cos(2\pi \omega h)$$

Spectral Densities of ARMA

The spectral densities of $ARMA(p,q)$ have been studied, and we have exact expressions for them.

- **White noise**
  The spectral density of white noise with variance σ^2 is a constant and it given by
  $$f(\omega) = \sigma^2.$$

- $\mathbf{MA(1)}$
  The spectral density of an $MA(1)$ sequence is given by
  $$f(\omega) = \sigma^2(1 + \theta^2 + 2\theta\cos(2\pi\omega)).$$

- $\mathbf{AR(1)}$
  The spectral density of a stationary $AR(1)$ sequence is given by
  $$f(\omega) = \sigma^2/[1 + \phi^2 - 2\phi\cos(2\pi\omega)].$$

- $\mathbf{ARMA(1,1)}$
  The spectral density of an $ARMA(1,1)$ sequence is given by
  $$f(\omega) = \sigma^2\frac{1 + \theta^2 + 2\theta\cos(2\pi\omega)}{1 + \phi^2 - 2\phi\cos(2\pi\omega)}.$$

Note the structure of the spectral density for an $ARMA(1,1)$ sequence. The numerator comes from the MA part and the denominator comes from the AR part. The general result for ARMA sequences is exactly the same. Let us recall the polynomials used in the discussion on issues of stationarity, invertibility etc.

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q,$$
$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p.$$

- $\mathbf{MA(q)}$
  The spectral density of a stationary $MA(q)$ sequence is given by
  $$f(\omega) = \sigma^2|\theta(z)|^2, \text{where } z = \exp(-2\pi i\omega).$$

- $\mathbf{AR(p)}$
  The spectral density of a stationary $AR(p)$ sequence is given by
  $$f(\omega) = \sigma^2/|\phi(z)|^2, \text{where } z = \exp(-2\pi i\omega).$$

- $\mathbf{ARMA(p,q)}$ The spectral density of a stationary $ARMA(p,q)$ sequence is given by
  $$f(\omega) = \sigma^2[|\theta(z)|^2/|\phi(z)|^2], \text{where } z = \exp(-2\pi i\omega).$$

The function `arma.spec()` can be used to compute the spectral density function values for a specified $ARMA(p,q)$ process. I'll go over a few plots in lecture. Try it out for yourself.

If you fit an $ARMA(p,q)$ model to a data set, you can use `arma.spec()` with the estimated parameter values to obtain the theoretical spectral density of your model. You can then compare it to the sample-based (smoothed) periodogram. The periodogram can be obtained and smoothed with the function `spec.pgram()`, which will be covered shortly.

Periodogram

Suppose we have n observations of a stationary zero-mean process $\{X_t\}$.
Set $m = n/2$ if n is even, and $m = (n-1)/2$ is n is odd.
Recall that the Fourier frequencies are $\omega_j = j/n$ for $j = 1, 2, \cdots, m$.
In this setting, the periodogram is most basic estimate of the spectral density. We will use the notation $I_j = I(\omega_j)$.
There are three ways to approach it.

1. Look back at the decomposition of X_t into random sinusoids in step 1 of the Spectral Density section. Suppose we approach it as a regression problem, in which X_t is regressed on the sinusoids of different frequencies. In this setting, it is natural to estimate the coefficients A_j and B_j $(j = 1, \cdots, m)$ using the method of least squares. This leads to the estimates

$$\hat{A}_j = \frac{\sum_{1\le t\le n} X_t \cos(2\pi\omega_j t)}{\sum_{1\le t\le n} \cos^2(2\pi\omega_j t)} = (2/n)\sum_{1\le t\le n} X_t \cos(2\pi\omega_j t),$$
$$\hat{B}_j = \frac{\sum_{1\le t\le n} X_t \sin(2\pi\omega_j t)}{\sum_{1\le t\le n} \sin^2(2\pi\omega_j t)} = (2/n)\sum_{1\le t\le n} X_t \sin(2\pi\omega_j t).$$

   Above, we use the equalities $\sum_{1\le t\le n} \cos^2(2\pi\omega_j t) = n/2$ and $\sum_{1\le t\le n} \sin^2(2\pi\omega_j t) = n/2$.

   Consider the quantity $P(\omega_j) = \hat{A}_j^2 + \hat{B}_j^2$. This is called the **scaled periodogram** value at frequency ω_j. The **raw periodogram** $I(\omega)$ can be obtained as

$$I_j = I(\omega_j) = (n/4)P(\omega_j).$$

2. The **discrete Fourier transform** at frequency ω_j (for $j = 1, \cdots, m$) is defined as

$$d(\omega_j) = (1/\sqrt{n})\sum_{1\le t\le n} x_t \exp(-2\pi i\omega_j t),$$

   where $i = \sqrt{-1}$ is the imaginary number and

$$\exp(-2\pi i w_j t) = \cos(2\pi\omega_j t) - i\sin(2\pi\omega_j t).$$

   The **raw periodogram** can also be obtained as the squared modulus of the discrete Fourier transform:

$$I(\omega_j) = |d(\omega_j)|^2 = (1/n)[\{\sum_{1\le t\le n} x_t \cos(2\pi\omega_j t)\}^2 + \{\sum_{1\le t\le n} x_t \sin(2\pi\omega_j t)\}^2]$$

   The above expression can be obtained by applying Euler's formula to get $\exp(-2\pi i w_j t) = \cos(2\pi\omega_j t) - i\sin(2\pi\omega_j t)$, and using the fact that $|a+bi|^2 = a^2+b^2$ for any complex number $a + bi$ with real a, b.

3. From the Spectral Density section that

$$f_X(\omega) = \gamma_X(0) + 2\sum_{h=1}^{\infty}\gamma_X(h)\cos(2\pi\omega h)$$

   if the autocovariance function is absolutely summable.

   The periodogram is obtained by simply replacing $\gamma_X(\cdot)$ with $\hat{\gamma}_X(\cdot)$:

$$I(\omega_j) = \hat{\gamma}_X(0) + 2\sum_{h=1}^{\infty}\hat{\gamma}_X(h)\cos(2\pi\omega h).$$

Distribution of the Periodogram Ordinates

The individual quantities $I_j = I(\omega_j)$ are called **periodogram ordinates**, and their distribution is of interest.

When n is large, $I(\omega_j)$ behaves approximately like $(\xi_j/2)f(j/n)$, where ξ_j has a chi-square distribution with 2 degrees of freedom. Moreover, the $I(\xi_j)$'s are independent for different j.

A random variable that has the Chi-squared distribution iwht k degrees of freedom has expected value k, and variance $2k$. From this we can deduce that when n is large,

$$\mathrm{E}(I_j) \approx 2f(\omega_j), \qquad \mathrm{Var}(I_j) \approx 4f^2(\omega_j).$$

The expression for the variance presents a problem. As the sample size increases, the variance of I_j does not shrink. This is an undersirable property for an estimator, and leads to something called **non-consistency**. An estimator is called not consistent if it does not converge in probability to the parameter it estimates. Having variance that does not decrease with n leads to this. Luckily there is a way out through smoothing the periodogram.

Smoothing the Periodogram

The easiest way to smooth a periodogram is to take a local (weighted) average. There are many methods for smoothing, but we will present only two.

Notation:

- Fourier frequencies: $\omega_j = j/n$, where $j = 1, \cdots, m$.
- $f_j = f(\omega_j)$ are the spectral density values at the Fourier frequencies.
- $I_j = I(\omega_j)$ are the raw periodogram values at the Fourier frequencies.
- $\hat{f}_{j,L}$ are the smoothed periodogram values at the Fourier frequencies, where smoothing is done by averaging over L values. Here $L = 2k + 1$ is odd, so averaging is done over k neighbors to the left, and k neighbors to the right.

The simplest approach is smoothing with **Daniell's kernel**. This is equivalent to just taking a local moving average over L values (k neighbors to the left, k neighbors to the right, and the raw value itself for a total of $L = 2k + 1$). For example, if we want to smooth over the periododogram values at four nearest frequencies (and at the frequency itself), then the smoothed periodogram is

$$\hat{f}_{j,5} = (1/5)[I_{j-2} + I_{j-1} + I_j + I_{j+1} + I_{j+2}].$$

A slightly improved version is called the **modified Daniell's kernel**. The only difference is that here the two farthest frequencies are given half of the weight. Suppose we again want to smooth over the values at four nearest frequencies (using $L = 5$). Now the estimate is

$$\hat{f}_{j,5} = (1/5)[(1/2)I_{j-2} + I_{j-1} + I_j + I_{j+1} + (1/2)I_{j+2}].$$

This is the default smoothing method for R's function 'textttspec.pgram()'. To use the modified Daniell's kernel, just specify the value of L using the argument 'spans'.

Choosing the Amount of Smoothing

The choice of L is important in practice, as it determines the shape of the smoothed plot. If the chosen L is too small, the plot will remain too rough to discern important features. On the other hand, if L is too large, the plot will be smooth, but the important features may be flattened out. It is possible to choose the amount of smoothing by eye, after looking of plots of smoothed periodograms with different values of L. Here is a more systematic approach.

1. Come up with candidate L values. It is possible to simply use all odd numbers from 3 to $2m - 1$. These are the largest and smallest possible values of L in a sample of size n.
2. Compute the raw (unsmoothed) periodogram values I_j for $j = 1, 2, \cdots, m$.
3. For each candidate L, do the following:
   a) Compute the smoothed periodogram values $\hat{f}_{j,L}$ for $j = 1, 2, \cdots, m$.
   b) Compute and record the value of the criterion
      - $Q(L) = \sum_{j=1}^{m}(I_j - \hat{f}_{j,L})^2 + \frac{1}{L}\sum_{j=1}^{m} I_j^2$ if using Daniell's kernel.
      - $Q(L) = \sum_{j=1}^{m}(I_j - \hat{f}_{j,L})^2 + \frac{1}{L-1}\sum_{j=1}^{m} I_j^2$ if using modified Daniell's kernel.
4. Choose the value of L that had the smallest value of the criterion $Q(L)$.

The R script *sta137_smoothingPgrm.R* contains code for utilizing this approach. The script is available on Smartsite.

Filtering

A process $\{Y_t\}$ is said to be a **linear filter** of a process $\{X_t\}$ if Y_t can be expressed as a linear combination of the elements of $\{X_t\}$. In other words, $\{Y_t\}$ is a linear filter of a process $\{X_t\}$ if

$$Y_t = \sum_{j=-\infty}^{\infty} \beta_j X_t,$$

where the $\{\beta_j\}$ are real constants satisfying $\sum_{j=-\infty}^{\infty} |\beta_j|$.

We have seen several examples of linear filters already.

- first difference
- MA process
- causal process

Suppose we know the properties of $\{X_t\}$, and that $\{Y_t\}$ is a linear filter of $\{X_t\}$ with known coefficients $\{\beta_j\}$. Here's how we can obtain the properties of $\{Y_t\}$.

1. If $\{X_t\}$ is stationary, then a linear filter of $\{Y_t\}$ is also stationary.
2. If $\{X_t\}$ is stationary, and has spectral density $f_X(\omega)$ on $[-0.5, 0.5]$, then the spectral density of $\{Y_t\}$ is
$$f_Y(\omega) = |g(\omega)|^2 f_X(\omega),$$
where $g(\omega)$ is the **frequency response function** defined as
$$g(\omega) = \sum_{j=-\infty}^{\infty} \beta_j e^{-2\pi i \omega j}.$$
The squared absolute value of the frequency response function, $|g(\omega)|^2$, is called the **power transfer function**.
df
3. If $\{X_t\}$ is stationary, and has ACVF $\gamma_X(h)$ for $h = 0, \pm 1, \pm 2, \cdots$, then the ACVF of $\{Y_t\}$ is
$$\gamma_Y(h) = \sum_{k=-\infty}^{\infty} \left(\sum_{j=-\infty}^{\infty} \beta_j \beta_{j+k} \right) \gamma_X(h-k).$$

Here's a useful consequence of the second result above.
If $\{X_t\}$ is stationary, and $\{Y_t\}$ is a local average, $Y_t = (X_t + X_{t-1} + \cdots + X_{t-(L-1)})/L$, then

$$f_Y(\omega) = \frac{1 - cos(2\pi\omega L)}{1 - cos(2\pi\omega)} f_X(\omega).$$

Appendix

Code for First Figure

```
library(astsa)
n = length(star)
par(mfrow=c(2,1), mar=c(3,3,1,1), mgp=c(1.6,.6,0))
plot(star, ylab="star magnitude", xlab="day")
Per = Mod(fft(star-mean(star)))^2/n
Freq = (1:n -1)/n
plot(Freq[1:50], Per[1:50], type='h', lwd=3, ylab="Periodogram",
     xlab="Frequency")
u = which.max(Per[1:50]) # 22 freq=21/600=.035 cycles/day
uu = which.max(Per[1:50][-u]) # 25 freq=25/600=.041 cycles/day
1/Freq[22]; 1/Freq[26] # period = days/cycle
text(.05, 7000, "24 day cycle"); text(.027, 9000, "29 day cycle")
### another way to find the two peaks is to order on Per
y = cbind(1:50, Freq[1:50], Per[1:50]); y[order(y[,3]),]
```