Lecture 5: Spectral Analysis and Filtering

Introduction to Time Series, Fall 2023

Ryan Tibshirani

Related reading: Chapters 4.1–4.3 and 4.7–4.8 of Shumway and Stoffer (SS).

1 Periodic processes

- Consider a periodic process of the form

$$x_t = A\cos(2\pi\omega t + \phi) \tag{1}$$

- Importantly, the quantity ω in the above definition is called *frequency* of the process; and the quantity $1/\omega$ is called the *period*. As t varies from 0 to $1/\omega$, note that the process goes through one complete cycle (it ends up back where it started). See Figure 1

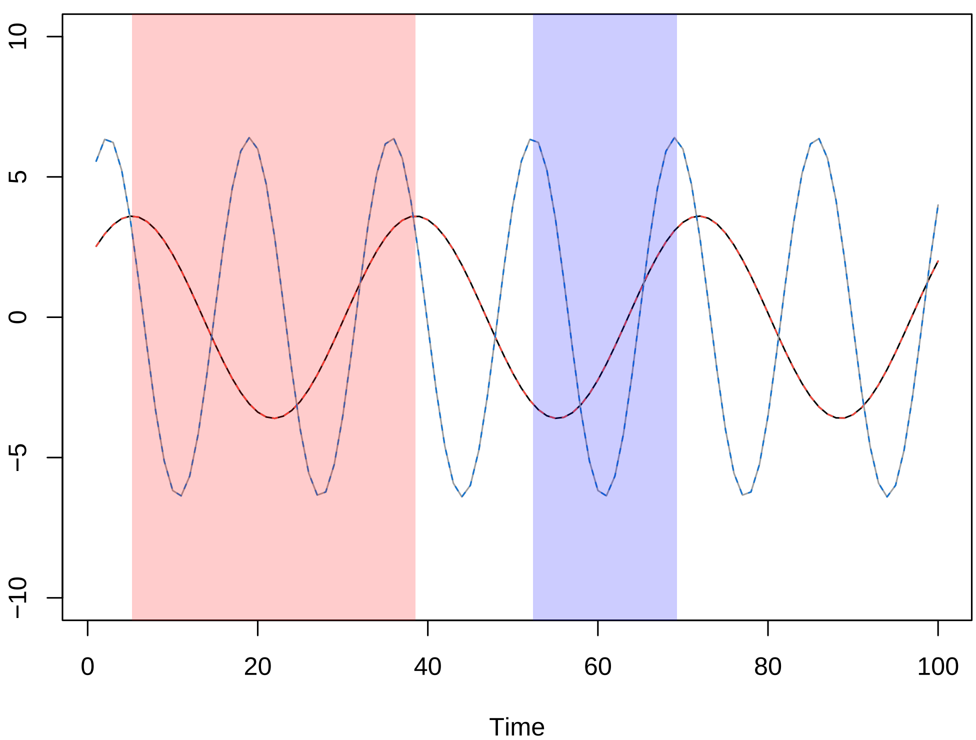

Figure 1: Two examples of cosine processes, the first (in red) having a frequency $\omega = 3/100$ and amplitude $\sqrt{2^2+3^2} \approx 3.6$, and the second (in blue) having a frequency $\omega = 6/100$ and amplitude $\sqrt{4^2+5^2} \approx 6.4$.

- The quantity A is called the *amplitude* and ϕ the *phase* of the process. The amplitude controls how high the peaks are, and the phase determine where (along the cosine cycle) the process starts at the origin $t = 0$

- We can introduce randomness into the process (1) by allowing A and ϕ to be random

- It will be useful to reparametrize. In general, recall the trigonometric identity (cosine compound angle formula):

$$\cos(a+b) = \cos(a)\cos(b) - \sin(a)\sin(b) \tag{2}$$

Thus, starting with (1), we can rewrite this as $x_t = A\cos(\phi)\cos(2\pi\omega t) - A\sin(\phi)\sin(2\pi\omega t)$. Simply letting $U_1 = A\cos(\phi)$, $U_2 = -A\sin(\phi)$, we can therefore write

$$x_t = U_1\cos(2\pi\omega t) + U_2\sin(2\pi\omega t) \tag{3}$$

with U_1, U_2 our two random variables, determining the amplitude of the cosine and sine components separately

- Note that another way of writing the relationship between A, ϕ and U_1, U_2 is (why?):

$$A = \sqrt{U_1^2 + U_2^2}, \quad \phi = \tan^{-1}(-U_2/U_1)$$

- An interesting fact (that you can try to verify as a challenge):

$$U_1, U_2 \sim N(0,1), \text{ independently} \iff A \sim \chi^2_2, \ \phi \sim \text{Unif}(-\pi, \pi), \text{ independently}$$

1.1 Stationarity

- If U_1, U_2 are uncorrelated, each with mean zero and variance σ^2, then the periodic process x_t, $t = 1, 2, 3, \ldots$ defined in (3) is stationary
- To check this: simply compute the mean function

$$\mu_t = \mathbb{E}(x_t) = 0$$

which is constant in time; and the auto-covariance function

$$\begin{aligned}
\gamma(s,t) &= \text{Cov}(x_s, x_t) \\
&= \text{Cov}\Big(U_1\cos(2\pi\omega s) + U_2\sin(2\pi\omega s),\ U_1\cos(2\pi\omega t) + U_2\sin(2\pi\omega t)\Big) \\
&= \text{Cov}\Big(U_1\cos(2\pi\omega s),\ U_1\cos(2\pi\omega t)\Big) + \text{Cov}\Big(U_2\sin(2\pi\omega s),\ U_1\cos(2\pi\omega t)\Big) \\
&\quad + \text{Cov}\Big(U_1\cos(2\pi\omega s),\ U_2\sin(2\pi\omega t)\Big) + \text{Cov}\Big(U_2\sin(2\pi\omega s),\ U_2\sin(2\pi\omega t)\Big) \\
&= \sigma^2\cos(2\pi\omega s)\cos(2\pi\omega t) + 0 + 0 + \sigma^2\sin(2\pi\omega s)\sin(2\pi\omega t) \\
&= \sigma^2\cos(2\pi\omega(s-t))
\end{aligned}$$

which only depends on the lag $s - t$ (where in the last line we used the identity (2) once again)

1.2 General mixtures

- As a generalization of (3), we can also mix together a total of p periodic processes, defining

$$x_t = \sum_{i=1}^{p} \Big(U_{j1}\cos(2\pi\omega_j t) + U_{j2}\sin(2\pi\omega_j t)\Big) \tag{4}$$

for U_{j1}, U_{j2}, $j = 1, \ldots, p$ all uncorrelated random variables with mean zero, where U_{j1}, U_{j2} have variance σ_j^2

- As a generalization of the above calculation, you'll show on your homework that the process x_t, $t = 1, 2, 3, \ldots$ defined in (4) is stationary, with auto-covariance function

$$\gamma(h) = \sum_{j=1}^{p} \sigma_j^2 \cos(2\pi\omega_j h)$$

- Figure 2 displays a couple of mixture processes of the form (4) (with $p = 2$ and $p = 3$). Note the regular repeating nature of the mixture processes. One might wonder how we can decompose a such a mixture into its frequency components (periodic processes, each of the form (3)). This is, in fact, one of the main objectives in spectral analysis

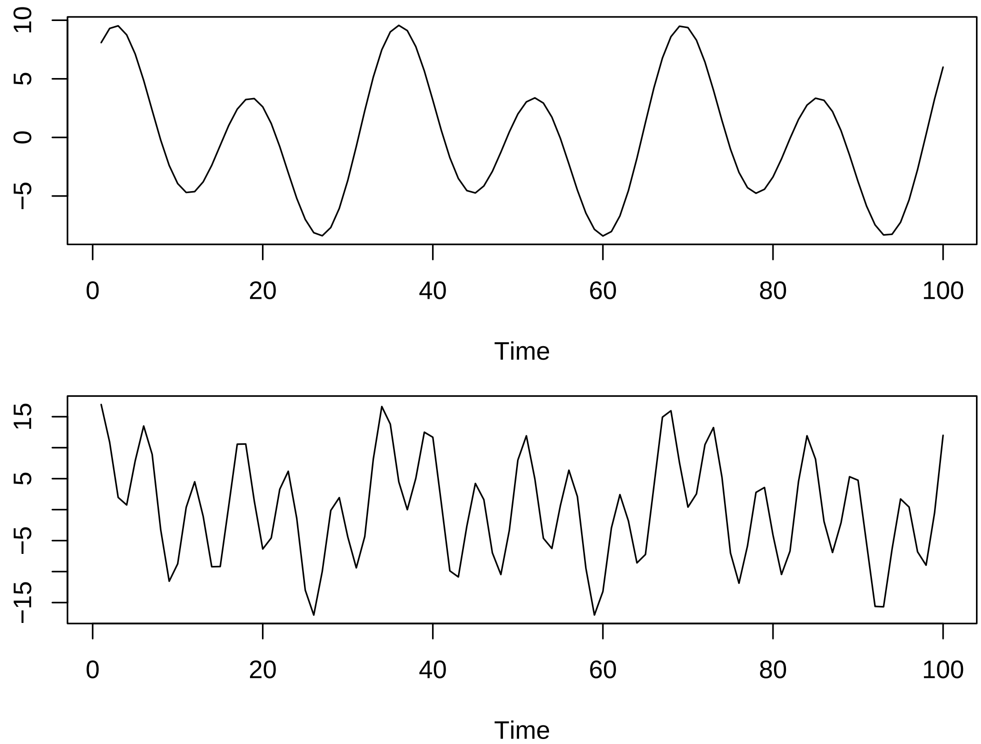

Figure 2: Mixture of periodic processes of different frequencies (and amplitudes).

- And the answer, as we'll see next, is given by something you're already quite familiar with ... regression!

2 Fourier decomposition

- Given a time series x_t, $t = 1, \dots, n$, consider *seeking* a decomposition like (4). We could do this by regressing this time series onto cosine and sine features of different frequencies,

$$c_{tj} = \cos(2\pi j/n \cdot t), \quad t = 1, \dots, n$$
$$s_{tj} = \sin(2\pi j/n \cdot t), \quad t = 1, \dots, n$$

(which we call these "basis functions" in the context of this particular regression problem), and so the regression model is

$$x_t \approx a_0 + \sum_{j=1}^{p} (a_j c_{tj} + b_j s_{tj}), \quad t = 1, \dots, n \tag{5}$$

Thus the regression coefficients a_j, b_j, $j = 1, \dots, p$ represent the amplitudes

- How large should p in the above regression model? That is, how many cosine and sine basis functions do we need? An amazing fact (at least, it will probably seem amazing if you've never seen Fourier decomposition before): *for any time series x_t, $t = 1, \dots, n$, we only need to set $p = (n-1)/2$, and then the representation in* (5) *will be exact!*

- That is, there are coefficients $\hat{a}_j, \hat{b}_j$, $j = 1, \ldots, p$ that will give us an equalities in (5), for all t
- (This assumes that n is odd; if n is even, then we need to add an additional component $a_{n/2}\cos(\pi t) = a_{n/2}(-1)^t$, and the same claim holds: the representation is exact)
- To find the coefficients $\hat{a}_j, \hat{b}_j$, $j = 1, \ldots, p$, we can simply perform regresion (least squares). We let $x \in \mathbb{R}^n$ denote our time series represented as a vector, which serves as the response vector in our regression problem, and we assemble our cosine and sine basis functions into a feature matrix

$$Z = \begin{bmatrix} \frac{1}{\sqrt{2n}} & \cos(2\pi\frac{1}{n}\cdot 1) & \sin(2\pi\frac{1}{n}\cdot 1) & \cos(2\pi\frac{2}{n}\cdot 1) & \ldots & \sin(2\pi\frac{n-1}{2n}\cdot 1) \\ \frac{1}{\sqrt{2n}} & \cos(2\pi\frac{1}{n}\cdot 2) & \sin(2\pi\frac{1}{n}\cdot 2) & \cos(2\pi\frac{2}{n}\cdot 2) & \ldots & \sin(2\pi\frac{n-1}{2n}\cdot 2) \\ \vdots & & & & & \\ \frac{1}{\sqrt{2n}} & \cos(2\pi\frac{1}{n}\cdot n) & \sin(2\pi\frac{1}{n}\cdot n) & \cos(2\pi\frac{2}{n}\cdot n) & \ldots & \sin(2\pi\frac{n-1}{2n}\cdot n) \end{bmatrix} \in \mathbb{R}^{n\times n}$$

- We can then perform regression of x on Z in order to estimate the coefficients:

$$(Z^\mathsf{T} Z)^{-1} Z^\mathsf{T} x$$

- However, something is very special about out matrix Z: it satisfies $Z^\mathsf{T} Z = (n/2)\cdot I$, where I the $n\times n$ identity matrix. In other words, *its columns are uncorrelated, and have squared ℓ_2 norm equal to $n/2$*. This is a very special property of the cosine and sine basis functions (and it is the foundation of the discrete Fourier transform, to be discussed shortly)
- Thus, writing z_j, $j = 1, \ldots, n$ as the columns of Z, we have

$$(Z^\mathsf{T} Z)^{-1} Z^\mathsf{T} x = \begin{bmatrix} \frac{2}{n} z_1^\mathsf{T} x \\ \frac{2}{n} z_2^\mathsf{T} x \\ \vdots \\ \frac{2}{n} z_n^\mathsf{T} x \end{bmatrix},$$

  so the multiple regression coefficients of x on Z are simply the marginal regression coefficients
- In other words, the coefficients are simply $\hat{a}_0 = \bar{x}$, and

$$\begin{aligned} \hat{a}_j &= \frac{2}{n} c_j^\mathsf{T} x = \frac{2}{n}\sum_{t=1}^n x_t \cos(2\pi j/n\cdot t) \\ \hat{b}_j &= \frac{2}{n} s_j^\mathsf{T} x = \frac{2}{n}\sum_{t=1}^n x_t \sin(2\pi j/n\cdot t) \end{aligned} \tag{6}$$

- And to be perfectly clear, this gives us the exact decomposition

$$x_t = \bar{x} + \sum_{j=1}^{(n-1)/2} \Big(\hat{a}_j \cos(2\pi j/n\cdot t) + \hat{b}_j \sin(2\pi j/n\cdot t)\Big), \quad t = 1, \ldots, n \tag{7}$$

  The reason: because Z is orthogonal, it has n linearly independent columns in n dimensions, so we can exactly represent any vector as a linear combination of its columns—which means that the fitted values from regressing x on Z will be exactly x
- (Side note on computation: at first glance, in order to compute each $\hat{a}_j$ or $\hat{b}_j$, we require $O(n)$ operations, and so computing all of them should take $O(n^2)$ time ... but in fact the *entire set of* coefficients $\hat{a}_j, \hat{b}_j$, $j = 1, \ldots, (n-1)/2$ can be computed in $O(n\log n)$ time using what is known as the fast Fourier transform, which we'll return to below)

2.1 Periodogram

- Given a series x_t, $t = 1, \ldots, n$, we can define an object from the coefficients (6) in the decomposition (7) that is called the *periodogram*, denoted P_x. This takes values at frequencies j/n, for $j = 1, \ldots, (n-1)/2$, and is defined by

$$P_x(j/n) = \frac{n}{4}\big(\hat{a}_j^2 + \hat{b}_j^2\big) \tag{8}$$

  When the underlying series is clear from the context, we will drop the subscript and simply write P

- Large values of the periodogram indicate which frequencies are predominant in the given series. This is illustrated in Figure 3, which displays the periodogorams for the two series in Figure 2. Another nice real data example (from a 1923 texbook on numerical analysis!) is given in Figure 4

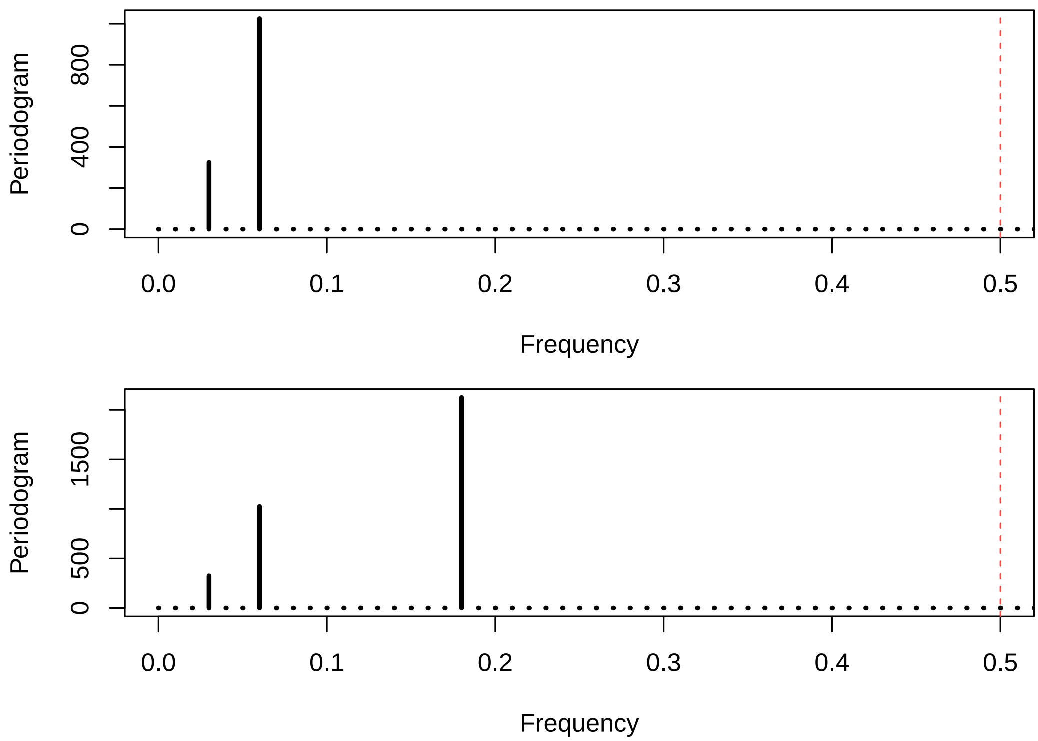

Figure 3: Periodograms for the mixture processes in Figure 2.

- If we think back to the mixture process (4) as a model for our data, then the periodogram gives us a breakdown of which frequencies are the largest *sources of variance*: recall $\sigma_j^2 = \mathbb{E}(U_{j1}^2 + U_{j2}^2)/2$ is the variance at frequency ω_j

2.2 Discrete Fourier transform

- The coefficients (6) used in the decomposition (7) can be computed efficiently by recognizing their connection to what is called the *discrete Fourier transform* (DFT)

- The DFT of a series x_t, $t = 1, \ldots, n$, is denoted d_x and defined as

$$d_x(j/n) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t \exp(-2\pi i j/n \cdot t), \quad j = 0, 1, \ldots, n-1 \tag{9}$$

  where i is the imaginary unit, which satisfies $i^2 = -1$. Thus the DFT is complex-value. As before, when the underlying series is clear from the context, we will drop the subscript and simply write d

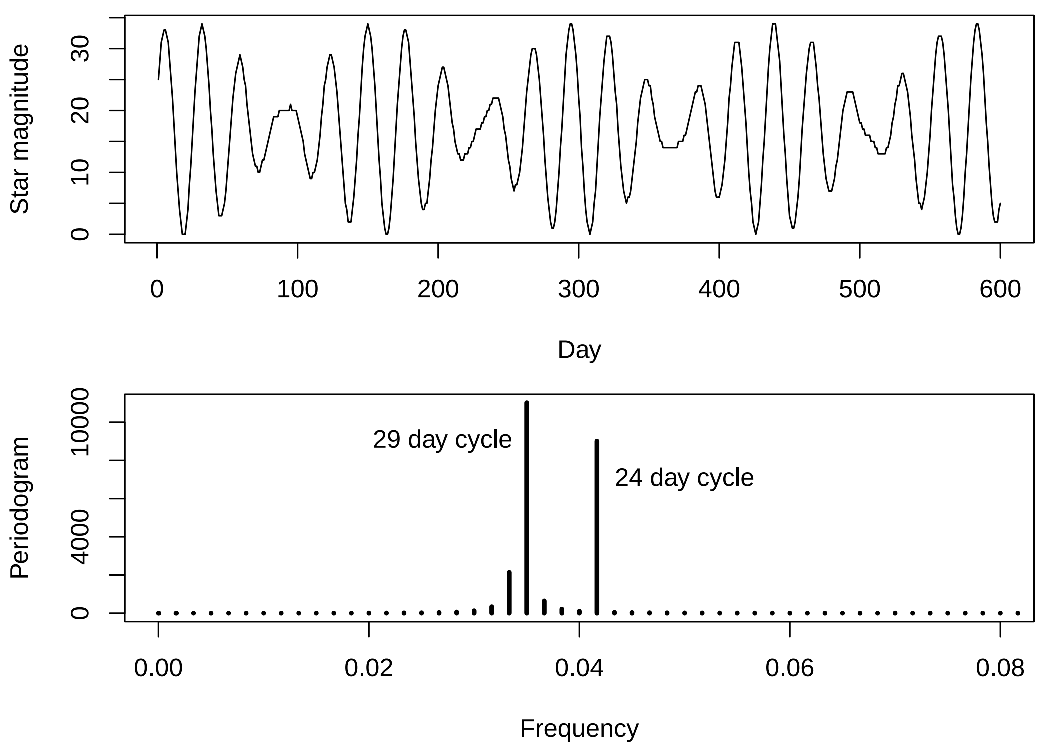

Figure 4: Periodogram for star magnitude data over 600 consecutive days (originally from "The Calculus of Observations" by Whittaker and Robinson, adapted by SS). Note the large values of the periodogram at 0.35 and 0.41, which correspond to $1/0.35 \approx 29$ and $1/0.41 \approx 24$ day cycles. For more on the interpretation, see Example 4.3 of SS.

- Recalling Euler's formula, $e^{i\theta} = \cos(\theta) + i\sin(\theta)$, we also have (using the fact that cosine is an even function and sine is odd):

$$d(j/n) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t \cos(2\pi j/n \cdot t) - \frac{i}{\sqrt{n}} \sum_{t=1}^{n} x_t \sin(2\pi j/n \cdot t), \quad j = 0, 1, \ldots, n-1$$

- Thus, from the DFT, we can compute each cosine and sine coefficient in (6) by

$$\hat{a}_j = \frac{2}{\sqrt{n}} \mathrm{Re}\{d(j/n)\} \quad \text{and} \quad \hat{b}_j = -\frac{2}{\sqrt{n}} \mathrm{Im}\{d(j/n)\}$$

  where for a complex number $z = a + bi$, we use $\mathrm{Re}\{z\} = a$ and $\mathrm{Im}\{z\} = b$ to denote its real and imaginary parts

- Note the following interesting connection to the periodogram. Since the the modulus of each entry of the DFT satisfies (by definition) $|d(j/n)|^2 = \mathrm{Re}\{d(j/n)\}^2 + \mathrm{Im}\{d(j/n)\}^2$, the periodogram in (8) is

$$\begin{aligned} P(j/n) &= \frac{n}{4}\left(\hat{a}_j^2 + \hat{b}_j^2\right) \\ &= \frac{n}{4}\left(\frac{4}{n}\mathrm{Re}\{d(j/n)\}^2 + \frac{4}{n}\mathrm{Im}\{d(j/n)\}^2\right) \\ &= |d(j/n)|^2 \end{aligned}$$

- In other words, the periodogram is simply the squared modulus of the DFT!

- Side note: the entire DFT can be computed rapidly using an algorithm called the fast Fourier transform (FFT), which takes $O(n \log n)$ operations (most efficient in practice when n is a highly composite integer, such as a power of 2). The modern generic FFT algorithm is credited to Cooley and Tukey in the 1960s, but similar ideas were arround much earlier
- Side side note: different software implementations scale the FFT/DFT differently, so you have to be careful to consult the documentation. For example, the `fft()` function in R computes it without the leading factor of $n^{-1/2}$, and with an additional factor of $\exp(2\pi i j/n)$, but this doesn't matter since we're only using it in our examples to compute the squared modulus, i.e., the periodogram, across frequencies

3 Spectral density

- Now we turn to a general fact about stationary processes in time series, which is called the *spectral representation* of the auto-covariance function
- In particular, if x_t, $t = 1, 2, 3, \ldots$ is stationary with autovariance function $\gamma(h) = \mathrm{Cov}(x_{t+h}, x_t)$, then there exists a unique increasing function F, called the *spectral distribution function* corresponding to the process, such that

$$\begin{aligned} F(\omega) &= 0 \quad && \text{for } \omega \le -1/2, \\ F(\omega) &= \gamma(0) \quad && \text{for } \omega \ge 1/2, \end{aligned}$$

  and for any $h = 0, \pm 1, \pm 2, \ldots,$

$$\gamma(h) = \int_{-1/2}^{1/2} \exp(2\pi i \omega h)\, dF(\omega)$$

- Above, we can think of F as being analogous to a cumulative distribution function (CDF), and thus the integral as being analogous as an expectation defined with respect the distribution that governs F. The only difference is that the total mass here of need not be 1, but is instead $\gamma(0)$
- We will generally ignore this distinction and call F a distribution anyway, and in the case F is differentiable, we will denote its derivative by f and call this the *spectral density*
- When the spectral density exists,[1] note that we have the representation

$$\gamma(h) = \int_{-1/2}^{1/2} \exp(2\pi i \omega h)\, f(\omega)\, d\omega, \quad h = 0, \pm 1, \pm 2, \ldots$$

- There is also an inverse relationship: we can represent f in terms of γ, via

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) \exp(-2\pi i \omega h), \quad \omega \in [-1/2, 1/2]$$

  In other words, the spectral density and autovariance function are *Fourier transform* pairs
- The important high-level perspective to remember here: *the auto-covariance function and spectral density contain the same information about a time series, but express it in different ways.* The auto-covariance function expresses the variation broken down by *lags*, whereas the spectral density expresses variation broken down by *frequencies*—or by *cycles* (remembering that the inverse of a frequency is a cycle)
- Next we compute the spectral density in a number of our favorite example stationary time series models

[1] It exists when the autovariance function is absolutely summable, which means that it satisfies $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$. See Appendix C of SS for details.

3.1 White noise

- Recall our most basic stationary process which is white noise: x_t, $t = 1, 2, 3, \ldots$ are uncorrelated random variables, with zero mean, and constant variance σ^2. In other words,

  $\gamma(0) = \sigma^2$ and $\gamma(Precisely, So by definition (this one is kind of vacuous), we have mean function \mu_t = 0$ and variance function $\sigma_t^2 = \sigma^2$, which are constant functions (do not vary in time)

  $x_t w_t$, $t = 1, 2, 3, \ldots$, where

3.2 Moving average

3.3 Autoregressive model

3.4 Sample spectral density

4 Linear filtering

4.1 Lagged regression

5 ST* decomposition