Unit 20: The Discrete Fourier Transform and The Periodogram

Taylor R. Brown

Department of Statistics, University of Virginia

Fall 2020

Readings for Unit 20

Textbook chapter 4.3.

Last Unit

1. Rational spectrum representation for ARMA processes.

This Unit

1. Discrete Fourier Transform
2. Periodogram

Motivation

So far, we've looked at the spectral density, which is a population quantity. We'll next consider the sample version: the periodogram.

Discrete Fourier Transform

Given data $x_1, x_2, \ldots, x_n$, we define the discrete Fourier transform (DFT) as

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} x_t e^{-2\pi i \omega_j t} \tag{1}$$

where $\omega_j = j/n, j = 0, 1, \ldots, n-1$, and the frequencies $\omega_j = j/n$ are called the Fourier or **fundamental frequencies**.

Discrete Fourier Transform

Another way to write it:

$$
\begin{aligned}
d(\omega_j) &= n^{-1/2} \sum_{t=1}^{n} x_t e^{-2\pi i \omega_j t} \\
&= n^{-1/2} \sum_{t=1}^{n} x_t [cos(-2\pi\omega_j t) + isin(-2\pi\omega_j t)] \\
&= n^{-1/2} \sum_{t=1}^{n} x_t cos(2\pi\omega_j t) - in^{-1/2} \sum_{t=1}^{n} x_t sin(2\pi\omega_j t) \\
&= d_c(\omega_j) - id_s(\omega_j)
\end{aligned}
$$

Discrete Fourier Transform

The difference between the DFT (1) and the Fourier transform in Unit 18 is that the DFT does the computation for discrete frequencies $\omega_j = j/n$ while the Fourier transform does for all frequencies $-1/2 \leq \omega_j \leq 1/2$.

A few more moves

In

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} x_t e^{-2\pi i \omega_j t}$$

note that $d(0) = \sqrt{n}\bar{x}$, and when $j \neq 0$

$$\sum_{t=1}^{n} e^{-2\pi i \omega_j t} = \sum_{t=1}^{n} e^{\frac{-2\pi ijt}{n}} = \sum_{t=1}^{n} \left(e^{\frac{-2\pi ij}{n}}\right)^t$$

$$= \left(e^{\frac{-2\pi ij}{n}}\right) \frac{1 - \left(e^{\frac{-2\pi ij}{n}}\right)^n}{1 - e^{\frac{-2\pi ij}{n}}} = 0$$

So, when $j \neq 0$, we can write the DFT as

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} (x_t - \bar{x}) e^{-2\pi i \omega_j t}$$

Periodogram

Recall we want to esitmate $f(\omega) = \sum_{h \in \mathbb{Z}} \gamma(h) e^{-2\pi i \omega h}$

The periodogram is defined to be

$$I(\omega_j) = |d(\omega_j)|^2. \tag{2}$$

Why is this a good sample estimate of $f(w)$?

Periodogram

For $j = 0$, we have $I(0) = n\bar{x}^2$. But, for $j \neq 0$

$$
\begin{aligned}
I(\omega_j) = |d(\omega_j)|^2 &= \left| n^{-1/2} \sum_{t=1}^{n} (x_t - \bar{x}) e^{-2\pi i \omega_j t} \right|^2 \\
&= \left(n^{-1/2} \sum_{t=1}^{n} (x_t - \bar{x}) e^{-2\pi i \omega_j t} \right) \left(n^{-1/2} \sum_{s=1}^{n} (x_s - \bar{x}) e^{2\pi i \omega_j s} \right) \\
&= n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} (x_t - \bar{x})(x_s - \bar{x}) e^{-2\pi i \omega_j [t-s]} \\
&= n^{-1} \sum_{h=-(n-1)}^{(n-1)} \sum_{t=1}^{n-|h|} (x_{t+|h|} - \bar{x})(x_t - \bar{x}) e^{-2\pi i \omega_j h} \\
&= \sum_{h=-(n-1)}^{(n-1)} \hat{\gamma}(h) e^{-2\pi i \omega_j h}
\end{aligned}
$$

A few things we need

If $j = 0$ or $j = n/2$

$$\sum_{t=1}^{n} \cos^2(2\pi\omega_j t) = n$$

$$\sum_{t=1}^{n} \sin^2(2\pi\omega_j t) = 0$$

Inverse Fourier Transform

The inverse Fourier transform can be used to decompose $\{x_t\}$:

$$
\begin{aligned}
x_t &= n^{-1/2} \sum_{j=0}^{n-1} d(\omega_j) e^{2\pi i \omega_j t} \\
&= \bar{x} + n^{-1/2} \sum_{j=1}^{n-1} d(\omega_j) e^{2\pi i \omega_j t} \\
&= \bar{x} + n^{-1/2} \sum_{j=1}^{n-1} d_c(\omega_j) [\cos(2\pi\omega_j t) + i \sin(2\pi\omega_j t)] \\
&\quad - i n^{-1/2} \sum_{j=1}^{n-1} d_s(\omega_j) [\cos(2\pi\omega_j t) + i \sin(2\pi\omega_j t)] \\
&= \bar{x} + n^{-1/2} \sum_{j=1}^{n-1} d_c(\omega_j) \cos(2\pi\omega_j t) + n^{-1/2} \sum_{j=1}^{n-1} d_s(\omega_j) \sin(2\pi\omega_j t)
\end{aligned}
$$

Inverse Fourier Transform (continued)

If n is odd, then $\cos(2\pi\omega_j n) = \cos(2\pi\omega_{n-j} n)$, so

$$
\begin{aligned}
x_t &= \bar{x} + n^{-1/2} \sum_{j=1}^{n-1} d_c(\omega_j) \cos(2\pi\omega_j t) + n^{-1/2} \sum_{j=1}^{n-1} d_s(\omega_j) \sin(2\pi\omega_j t) \\
&= a_0 + 2n^{-1/2} \sum_{j=1}^{m} d_c(\omega_j) \cos(2\pi\omega_j t) + 2n^{-1/2} \sum_{j=1}^{m} d_s(\omega_j) \sin(2\pi\omega_j t) \\
&= a_0 + \sum_{j=1}^{m} a_j \cos(2\pi\omega_j t) + \sum_{j=1}^{m} b_j \sin(2\pi\omega_j t)
\end{aligned}
$$

where $a_0 = \bar{x} = (x_1 + \cdots + x_n)/n$, and $m = (n-1)/2$.

Interpreting the Periodogram

We can think of the inverse Fourier transform as a regression of x_t on sines and cosines with the coefficients equal to $2/\sqrt{n}$ times the sine part and the cosine part of the Fourier transforms respectively. Therefore, $d_c(\omega_j)$ and $d_s(\omega_j)$ measure the contribution the frequency ω_j has in **explaining the variation** in the time series. The bigger $d_c(\omega_j)$ and $d_s(\omega_j)$, the greater the contribution from the frequency ω_j.

Interpreting the Periodogram

Additionally, one can further show that

$$\sum_{t=1}^{n}(x_t - \bar{x})^2 = 2\sum_{j=1}^{m}[d_c^2(\omega_j) + d_s^2(\omega_j)] = 2\sum_{j=1}^{m} I(\omega_j).$$

Interpreting the Periodogram

The sum of squares can be decomposed into 2 times the sum of the periodograms over frequencies $\omega_j, 1 \le j \le m$. In other words, the variation in the series x_t is distributed over frequencies ω_j, where the amount of variation explained by frequency ω_j is $2I(\omega_j)$.

Interpreting the Periodogram

Thus, we can interpret the periodogram as the amount of variation at a certain frequency. This is how we also interpret the spectral density. The periodogram is the sample version of the spectral density, which is a population quantity.

Sampling Distribution of Periodogram

Let $\omega_{j:n}$ denote a frequency of the form j_n/n, where $\{j_n\}$ is a sequence of integers so that $j_n/n \to \omega$ as $n \to \infty$. It turns out that

$$E[I(\omega_{j:n})] \to f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h)e^{-2\pi i\omega h}.$$

The spectral density is the **long run average** of the periodogram.

Sampling Distribution of Periodogram

It turns out that if $\{x_t\}$ is causal and $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, then

$$\frac{2I(\omega_{j:n})}{f(\omega_j)} \xrightarrow{d} \text{i.i.d. } \chi_2^2. \tag{3}$$

provided $f(\omega_j) > 0$ for $j = 1, \cdots, m$ for any collection of m distinct frequencies ω_j with $\omega_{j:n} \to \omega_j$.

Confidence Intervals

From (3), an approximate $100(1-\alpha)\%$ confidence interval for the spectral density takes the form

$$\frac{2I(\omega_{j:n})}{\chi_2^2(1-\alpha/2)} \leq f(\omega) \leq \frac{2I(\omega_{j:n})}{\chi_2^2(\alpha/2)}. \tag{4}$$

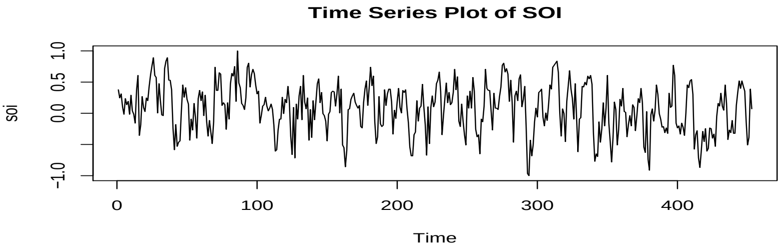
Comments

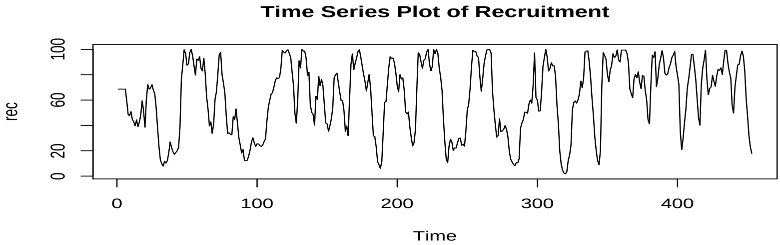
- χ^2_2 distribution has mean 2, thus the expected value of $I(\omega_j)$ is approximately $f(\omega_j)$, i.e. the periodogram is approximately **unbiased**.
- The variance of $I(\omega_j)$ is approximately $f^2(\omega_j)$. For example, for Gaussian white noise, the variance of the periodogram is σ_w^4 which **does not decrease with** n.

Worked Example

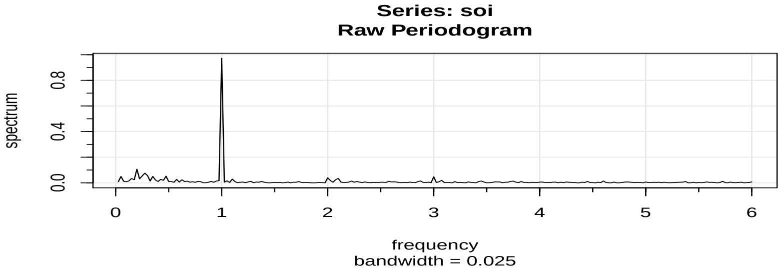
In this example, we will look at the Southern Oscillation Index and recruitment datasets, which contain monthly data on the changes in air pressure and estimated number of new fish in the central Pacific Ocean from 1950 to 1987. The central Pacific Ocean warms approximately every three to seven years due to El Niño.

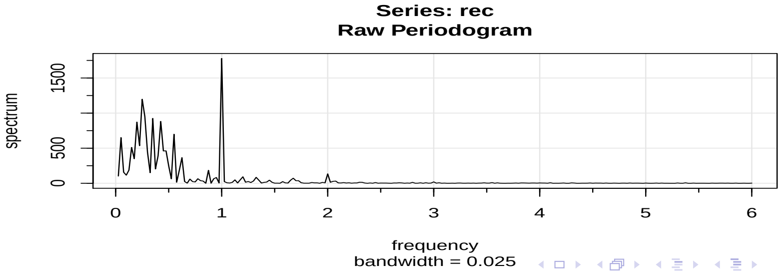
Worked Example



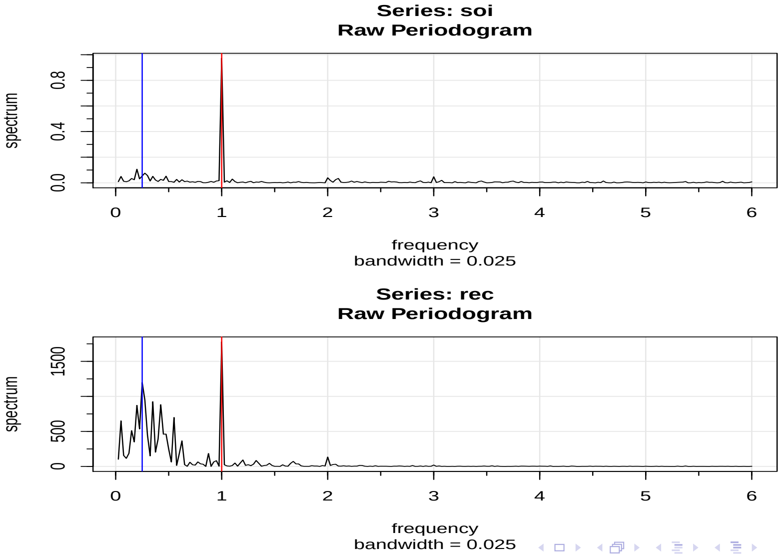


Worked Example





Worked Example



Worked Example

From the periodograms:

- obvious peak at $\omega = 1/12$ for yearly cycle.
- some peaks at around $\omega = 1/48$ for El Nino cycle. The wide band of activity suggests that this cycle is **not very regular**.

Note: the horizontal axis of the periodogram produced using the `mvspec()` function from the `astsa` package is in multiples of $\frac{1}{12}$. In Unit 6, the `spec.pgram()` function was used instead and the horizontal axis is the value of the frequency.

Worked Example

From the SOI data, the value of the periodogram at $\omega = \frac{1}{12}$ is $I(\frac{1}{12}) = 0.9722$. Since $\chi_2^2(0.025) = 0.0506$ and $\chi_2^2(0.975) = 7.3778$, an approximate 95% confidence interval for the spectrum $f(\frac{1}{12})$ is

$$\left(\frac{2(0.9722)}{7.3778}, \frac{2(0.9722)}{0.0506}\right) = (0.2636, 38.4011).$$

At $\omega = \frac{1}{48}$, $I(\frac{1}{48}) = 0.0537$, therefore an approximate 95% confidence interval for the spectrum $f(\frac{1}{48})$ is

$$\left(\frac{2(0.0537)}{7.3778}, \frac{2(0.0537)}{0.0506}\right) = (0.0146, 2.1222).$$

Worked Example

Question: Use R to compute $I(\frac{1}{12})$ and $I(\frac{1}{48})$ for the recruit data, as well as derive their corresponding approximate 95% confidence intervals.