Unit 20: Discrete Fourier Transform and Periodogram

Jeffrey Woo

Department of Statistics, University of Virginia

Spring 2016

Readings for Unit 20

Textbook chapter 4.4.

Last Unit

1. Autocovariance generating function for ARMA processes.
2. 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.

1. Discrete Fourier Transform
2. Periodogram
3. Sampling Distribution of Periodogram
4. Worked Example

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 ______________________.

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 frequency ω_j while the Fourier transform does for all frequencies $-1/2 \leq \omega_j \leq 1/2$.

Discrete Fourier Transform

Using Euler's formula $e^{i\alpha} = cos(\alpha) + isin(\alpha)$, we could also write the DFT (1) as

$$d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} x_t cos(2\pi\omega_j t) - in^{-1/2} \sum_{t=1}^{n} x_t sin(2\pi\omega_j t). \quad (2)$$

Discrete Fourier Transform

The real part of (2) is identified as the cosine transform $d_c(\omega_j)$ and the imaginary part as the sine transform $d_s(\omega_j)$. We could write the DFT (1) as

$$d(\omega_j) = d_c(\omega_j) - i d_s(\omega_j).$$

Periodogram

The periodogram is defined to be

$$I(\omega_j) = |d(\omega_j)|^2. \tag{3}$$

i.e. the squared modulus of the DFT (1). Recall that $|a + ib|^2 = a^2 + b^2$. Therefore,

$$I(\omega_j) = (d_c(\omega_j))^2 + (-d_s(\omega_j))^2 = d_c^2(\omega_j) + d_s^2(\omega_j)$$

A Few Notes about Complex Numbers

$$\begin{aligned}
\overline{a+ib} &= a-ib \quad \text{(complex conjugate,} \quad \bar{a}=a \text{ for real number } a.) \\
|a+ib|^2 &= a^2+b^2=(a+ib)\overline{(a+ib)}. \quad (|z| \text{ is the modulus of } z.) \\
\overline{z_1 z_2} &= \overline{z_1}\,\overline{z_2}, \quad \text{for two complex numbers } z_1, z_2.
\end{aligned}$$

Periodogram as Sample Version of Spectral Density

Inverse Fourier Transform

Another useful property is the inverse Fourier transform

$$\begin{aligned} x_t &= n^{-1/2} \sum_{j=0}^{n-1} d(\omega_j) e^{2\pi i \omega_j t} \\ &= \\ &= \\ &= \\ &= \end{aligned}$$

where $a_0 = \bar{x} = (x_1 + \cdots + x_n)/n$, and we have assumed for simplicity that n is odd 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 ______________________ 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 $\omega_j, 1 \leq j \leq m$. 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 ________________ of the periodogram.

Sampling Distribution of Periodogram

It turns out that if $\{x_t\}$ is causal, then

$$\frac{2I(\omega_{j:n})}{f(\omega_j)} \xrightarrow{d} \text{i.i.d. } \chi^2_2. \tag{4}$$

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

$$\frac{2I(\omega_{j:n})}{\chi_2^2(1-\alpha/2)} \leq f(\omega) \leq \frac{2I(\omega_{j:n})}{\chi_2^2(\alpha/2)}. \tag{5}$$

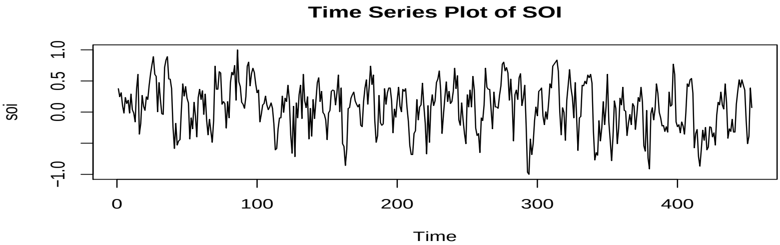
Comments

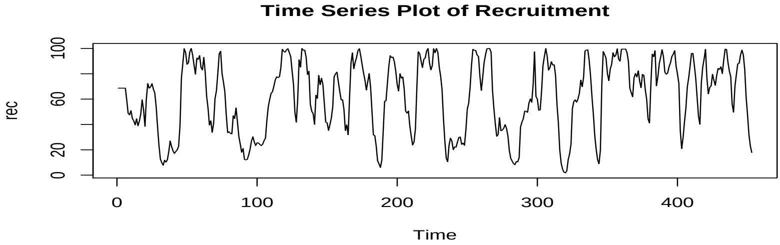
- χ^2_2 distribution has mean 2, thus the expected value of $I(\omega_j)$ is approximately $f(\omega_j)$, i.e. the periodogram is approximately ________.
- The variance of $I(\omega_j)$ is approximately $f^2(\omega_j)$. For example, for Gaussian white noise, the variance of the periodogram is σ_w^4 which ____________________.

Worked Example

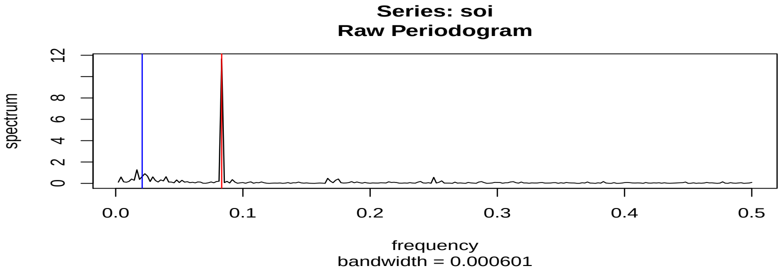
In this example, we will look at the Southern Oscillation Index ("soi.dat") and recruitment ("recruit.dat") datasets, which contain monthly data on the changes in air pressure and estimated number of new fish in the central Pacific Ocean from 1950 to 1987. The central Pacific Ocean warms approximately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

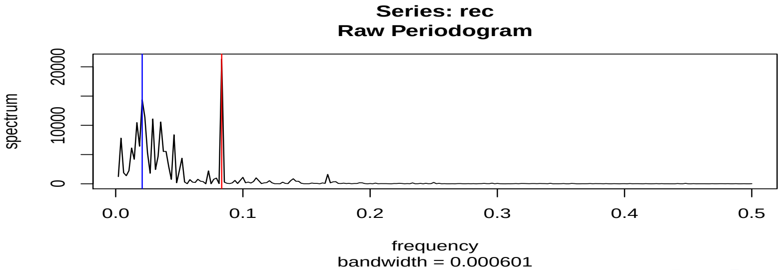
Worked Example





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 ________________.

Worked Example

From R, the value of the periodogram at $\omega = \frac{1}{12}$ is $I(\frac{1}{12}) = 11.6668$ for the SOI data. 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

$$\left(\frac{2(11.6668)}{7.3778}, \frac{2(11.6668)}{0.0506} \right) = (3.16, 460.81).$$

At $\omega = \frac{1}{48}$, $I(\frac{1}{48}) = 0.6448$, therefore an approximate 95% confidence interval for the spectrum $f(\frac{1}{48})$ is

$$\left(\frac{2(0.6448)}{7.3778}, \frac{2(0.6448)}{0.0506} \right) = (0.17, 25.47).$$

Worked Example

Question: Any comment(s) about the intervals?