The Frequency Domain and the Periodogram

Jared Fisher

Lecture 2b

Announcements

Announcements

- Homework 1 was due last night. Thanks for your hard work everybody!
- We plan to open a new/fifth lab section next week. Stay tuned!
- Project Checkpoint 1 is due next week. Let's talk about the project!

Project

- See files on bCourses
- Two datasets: stock price data, COVID-19 data. These datasets will provide two very different analysis experiences.

Recap

Full Model

$$Y_t = m_t + s_t + X_t$$

- m_t is the **deterministic** trend
- s_t is the **deterministic** seasonal effect
- X_t is as stationary process, perhaps white noise
- **Idea:** Remove trend and seasonality, so that residuals exhibit steady behavior over time, i.e. looks stationary.

“Nonparametric” Seasonality via Indicators

$$\hat{s}_i := \text{average of } X_i,\, X_{i+d},\, X_{i+2d},\, \ldots$$

Note though that we’re fitting d parameters with n observations. n must be sufficiently larger than d.

Parametric Seasonality Function

$$s_t = \sum_{k=1}^{K} \left(a_k \cos(2\pi tk/d) + b_k \sin(2\pi tk/d)\right)$$

- a is the "Amplitude"
- $f = k/d$ is the "Frequency"
- d/k is the "Period"
- No need for $K > d/2$
- See R code explanation of why

Definition: Sinusoids

We define the set of sinusoid functions as

$$\{g(t) = R\cos(2\pi ft + \Phi) \ : \ R \in R_+, \ f \in R_+, \ \Phi \in [0, 2\pi/f)\},$$

where

- R is called the *amplitude*
- f is called the *frequency*
- Φ is called the *phase*
- $1/f$ is called the *period*

Sinusoids rewritten a different way

- Estimating the phase shift Φ is nontrivial with the tools in this class, but we can rewrite the sinusoid equation to be more convenient (you will show this in lab).
- With $A = R\cos(\Phi)$ and $B = -R\sin(\Phi)$ one can rewrite sinuosoids as

$$\{g(t) = A\cos(2\pi ft) + B\sin(2\pi ft) \ : \ A, B \in R, \ f \in R_{+}\}.$$

- This is helpful as we can find the coefficients A and B with linear models, but that means we must find the appropriate frequencies f first. The frequency domain will help with this!

Textbook

Reading applicable to today's lecture: section 4.1 and page 182 of Section 4.3

First: A Brief Review/Overview of Complex Numbers

Introduction to Complex Roots via Example

- Consider this polynomial of interest: $1 - z + 0.5z^2$
- What are roots? i.e. set equal to 0 and solve for z:

$$0 = 1 - z + 0.5z^2$$

- Recall for $0 = az^2 + bz + c$, $z = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$
- Plug in values:

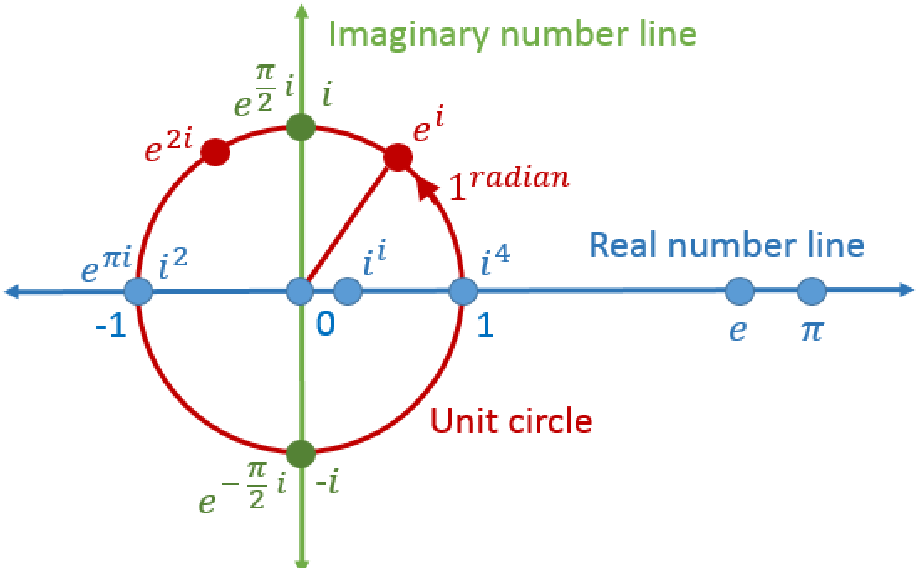
$$\frac{-(-1) \pm \sqrt{(-1)^2 - 4(0.5)(1)}}{2(0.5)} = \frac{1 \pm \sqrt{1-2}}{1} = 1 \pm \sqrt{-1}$$

- Thus the roots are $1 + i$ and $1 - i$

Brief Review of Complex Numbers

- Imaginary number: $i = \sqrt{-1}$
- Complex number: $z = a + bi$, where a,b are real valued
- $\bar{z} = a - bi$ is the complex conjugate of $z = a + bi$
- Euclidean distance: $d(a + bi) = \sqrt{a^2 + b^2}$
- We often ask if roots are within the unit circle, or $\sqrt{a^2 + b^2} \leq 1$

Complex Unit Circle



Complex Polar Coordinates

- $z = a + bi$
- $r = d(a + bi) = \sqrt{a^2 + b^2}$
- $a = r * cos(\theta),\ b = r * sin(\theta)$
- Note Euler's equation: $e^{i\theta} = cos(\theta) + i * sin(\theta)$

$$\begin{aligned} z &= r * cos(\theta) + r * sin(\theta)i \\ &= r * e^{i\theta} \end{aligned}$$

Back to Example

- The roots of the example polynomial were $1+i$ and $1-i$.
- Magnitudes:

$$\sqrt{(1)^2 + (\pm 1)^2} = \sqrt{1+1} = \sqrt{2} > 1$$

- As the roots are outside the unit circle! This will be important later in the course.

Frequency Domain: How we can choose what frequency of sinusoid to use

Note

- We now define a transformation of data, which expresses the data in terms of its sinusoidal waves of different frequencies
- This will allow us to see which frequencies are prevalent in the time series

Definition: Discrete Fourier Transform

For data $x_0, \ldots, x_{n-1} \in C$ the discrete Fourier transform (DFT) is given by $b_0, \ldots, b_{n-1} \in C$, where

$$b_j = \sum_{t=0}^{n-1} x_t \exp\left(-\frac{2\pi ijt}{n}\right) \text{ for } j = 0, \ldots, n-1.$$

(In R, the DFT is calculated by the function *fft()*.)

- The frequencies j/n for $j = 0, \ldots, n-1$ as called **Fourier frequencies**.

Notes on DFT

- It always holds that $b_0 = \sum x_t$.
- When $x_0, \ldots, x_{n-1} \in R$ are real numbers (in general, can be complex), then

$$\begin{aligned} b_{n-j} &= \sum_t x_t \exp\left(-\frac{2\pi i(n-j)t}{n}\right) \\ &= \sum_t x_t \exp\left(\frac{2\pi ijt}{n}\right) \exp\left(-2\pi it\right) = \bar{b}_j. \end{aligned}$$

- For example, for $n = 11$, the DFT can be written as:

$$b_0, b_1, b_2, b_3, b_4, b_5, \bar{b}_5, \bar{b}_4, \bar{b}_3, \bar{b}_2, \bar{b}_1.$$

- For $n = 12$, it is

$$b_0, b_1, b_2, b_3, b_4, b_5, \bar{b}_6, \bar{b}_5, \bar{b}_4, \bar{b}_3, \bar{b}_2, \bar{b}_1.$$

  Note that b_6 is necessarily real because $b_6 = \bar{b}_6$.

Note on DFT

- DFT $b_0, \ldots, b_{n-1}$ is in one-to-one correspondence with the data $x_0, \ldots, x_{n-1}$, because the original data can be uniquely recovered by its DFT, as the following theorem shows.
- $\Rightarrow$ the DFT $b_0, \ldots, b_{n-1}$ and the data $x_0, \ldots, x_{n-1}$ contain equivalent information.

Theorem: Inverse Fourier Transform (IDFT)

For data $x_0, \dots, x_{n-1}$ and its DFT $b_0, \dots, b_{n-1}$, it holds that

$$x_t = \frac{1}{n} \sum_{j=0}^{n-1} b_j \exp\left(\frac{2\pi ijt}{n}\right) \text{ for } t = 0, \dots, n-1.$$

Proof (page 1)

- Start with the right hand side of IDFT

$$\frac{1}{n}\sum_{j=0}^{n-1} b_j \exp\left(\frac{2\pi ijt}{n}\right)$$

- Insert the DFT formula

$$= \frac{1}{n}\sum_{j=0}^{n-1}\left\{\sum_{s=0}^{n-1} x_s \exp\left(-\frac{2\pi ijs}{n}\right)\right\}\exp\left(\frac{2\pi ijt}{n}\right)$$

$$= \frac{1}{n}\sum_{s=0}^{n-1} x_s \sum_{j=0}^{n-1} \exp\left(\frac{2\pi ij(t-s)}{n}\right)$$

Proof (page 2)

- Note the inner sum equals n when $s = t$.

$$\sum_{j=0}^{n-1} \exp\left(\frac{2\pi ij(t-s)}{n}\right)$$

- Take out the j exponent

$$\sum_{j=0}^{n-1} \exp\left(\frac{2\pi i(t-s)}{n}\right)^j$$

- For $s \neq t$ we have that $\exp\left(\frac{2\pi i(t-s)}{n}\right) \neq 1$

Proof (page 3)

- Apply the finite geometric series formula to the inner sum

$$
\begin{aligned}
&= \frac{1-\exp\left(\frac{2\pi i(t-s)}{n}\right)^{n}}{1-\exp\left(\frac{2\pi i(t-s)}{n}\right)} \\
&= \frac{1-\exp\left(2\pi i(t-s)\right)}{1-\exp\left(\frac{2\pi i(t-s)}{n}\right)} \\
&= \frac{1-1}{1-\exp\left(\frac{2\pi i(t-s)}{n}\right)} \\
&= 0.
\end{aligned}
$$

as $exp(ai\pi) = (-1)^a$ for integer a, and a is always even.

Aside: Sinusoids from DFT

To see why the DFT expresses the data in terms of its sinusoidal wave components, note that for $x = (x_0, \dots, x_{n-1})$ one can write

$$x = \frac{1}{n} \sum_{j=0}^{n-1} b_j u^j.$$

with vectors

$$u^j = (1, \exp(2\pi i j/n), \exp(2\pi i 2j/n), \dots, \exp(2\pi i (n-1) j/n))$$

for $j = 0, \dots, n-1$. That is, the sinusoid with frequency j/n evaluated at the time points $t = 0, 1, \dots, (n-1)$.

- the vectors u^j are an orthogonal basis: $(u^l)^T u^k = 0$ for $l \neq k$.

Real vs Complex

- Note that the DFT $b_0, \ldots, b_{n-1}$ of real valued data $x_0, \ldots, x_{n-1}$ can be complex valued.
- To visualize the DFT, we plot its absolute value.
- Note that b_0 is always just the sum of the data, which does not capture much information.
- Further because $b_{n-j} = \bar{b}_j$, it is enough to look at $|b_j|, 1 \le j \le n/2$.

Real vs Complex

- Note that the DFT $b_0, \ldots, b_{n-1}$ of real valued data $x_0, \ldots, x_{n-1}$ can be complex valued.
- To visualize the DFT, one rather plots its absolute value.
- Note that b_0 is always just the sum of the data, which does not capture much information.
- Further because $b_{n-j} = \bar{b}_j$, it is enough to look at $|b_j|, 1 \leq j \leq n/2$.

Periodogram

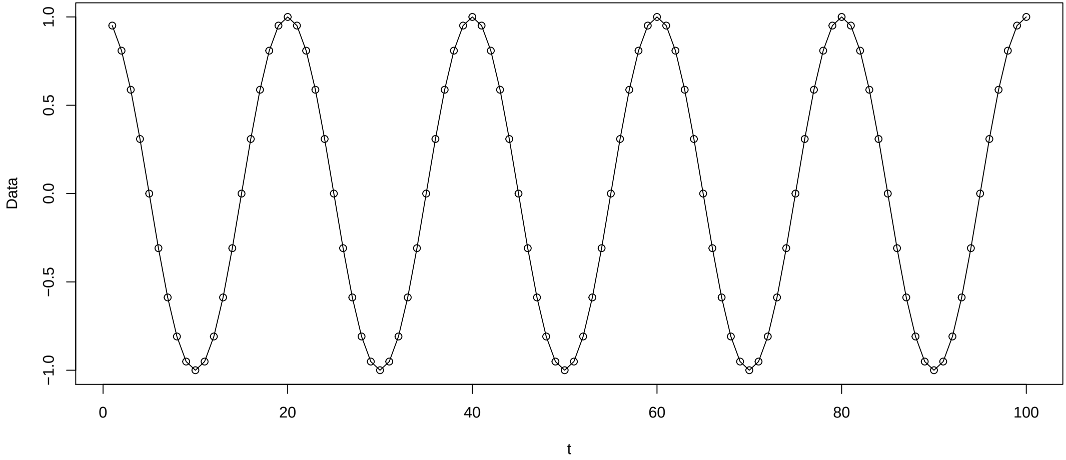
Definition: Periodogram

For real values data $x_0, \ldots, x_{n-1}$ with DFT $b_0, \ldots, b_{n-1}$ the **periodogram** is defined as

$$I(j/n) = \frac{|b_j|^2}{n} \quad \text{for } j = 1, \ldots, \lfloor n/2 \rfloor$$

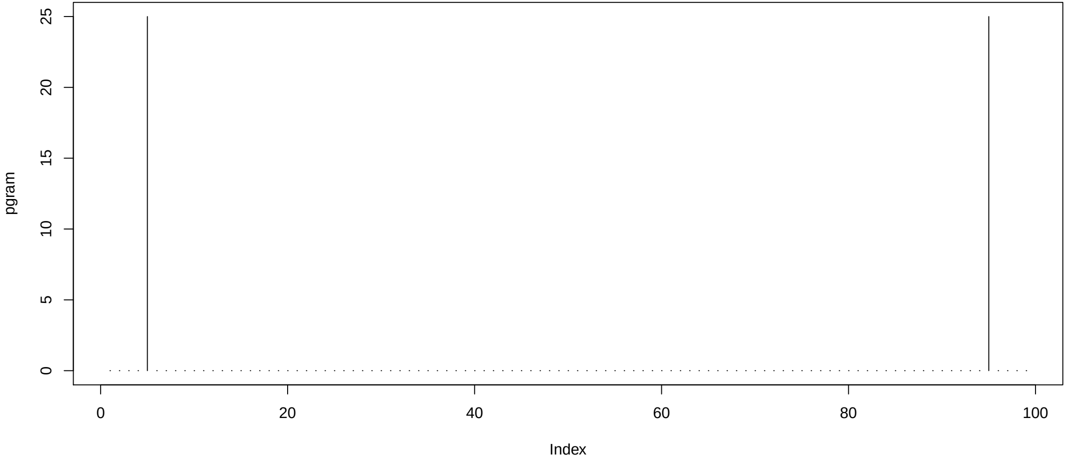
Example Data: $cos(2\pi t * 5/100)$

```
n=100; t = 1:n; cos2 = cos(2*pi*t*(5/n))
plot(t, cos2, ylab = "Data", type = "o")
```



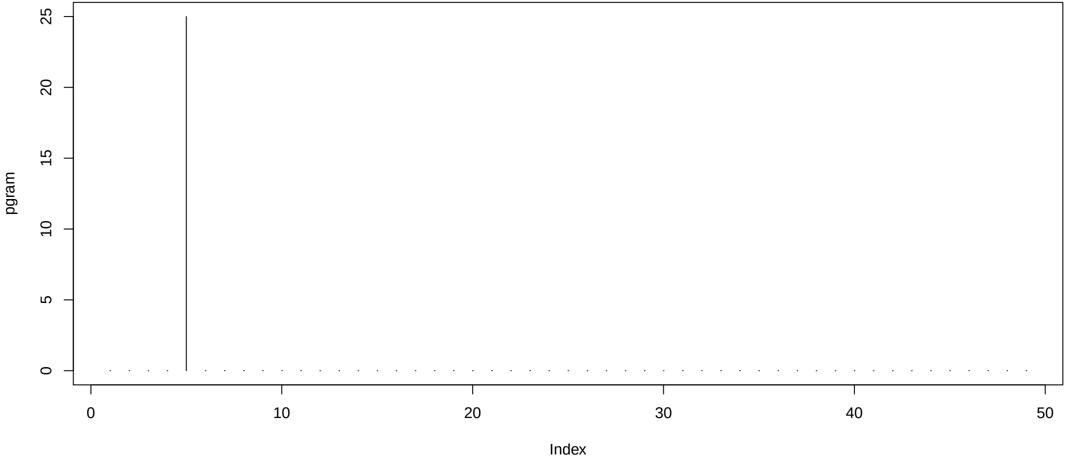
Example: $cos(2\pi t * 5/100)$

```
pgram = abs(fft(cos2)[2:100])^2/n
plot(pgram, type = "h")
```

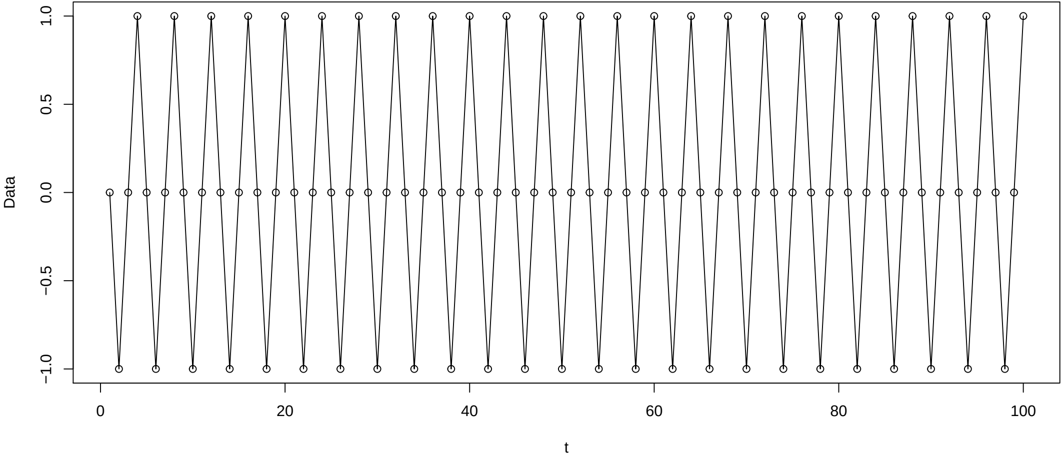


Example Periodogram: $cos(2\pi t * 5/100)$

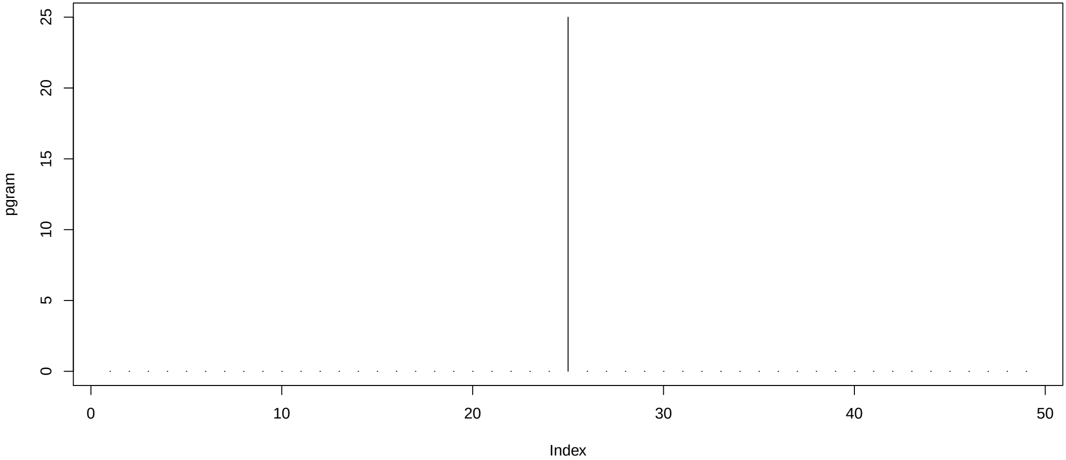
```r
pgram = abs(fft(cos2)[2:50])^2/n
plot(pgram, type = "h")
```



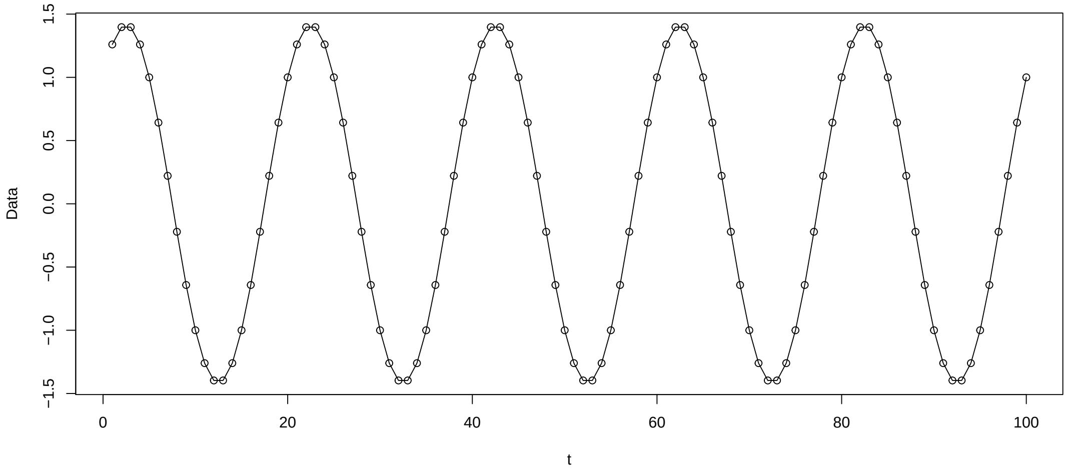
Example Data: $cos(2\pi t * 25/100)$



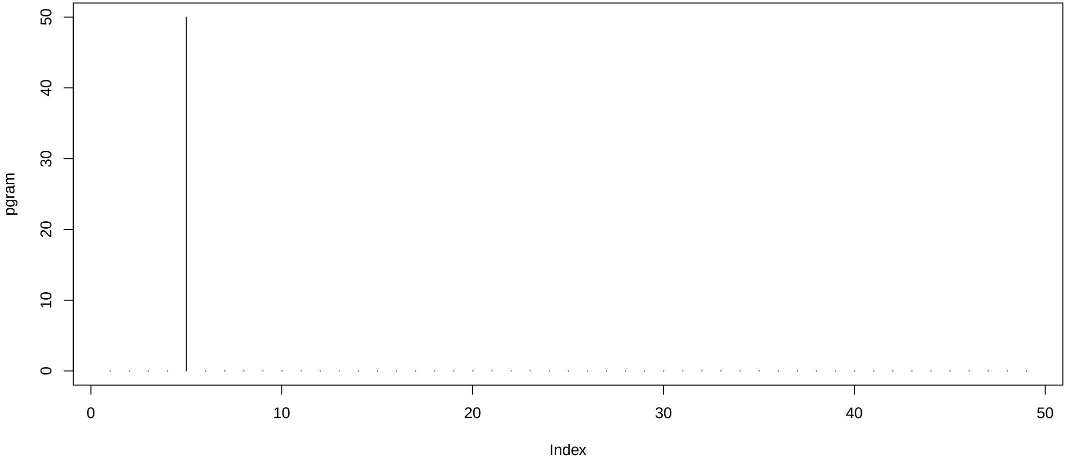
Example Periodogram: $cos(2\pi t * 25/100)$



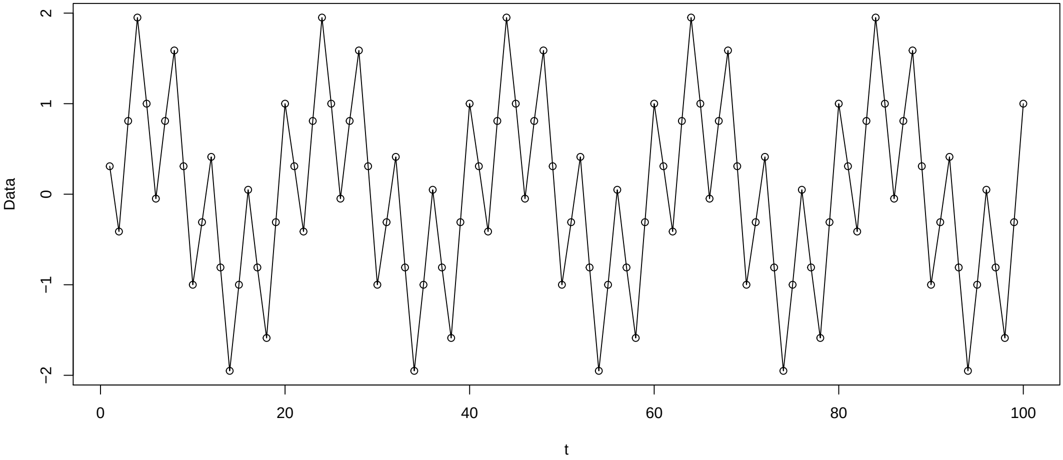
Example Data: $cos(2\pi t * 5/100) + sin(2\pi t * 5/100)$



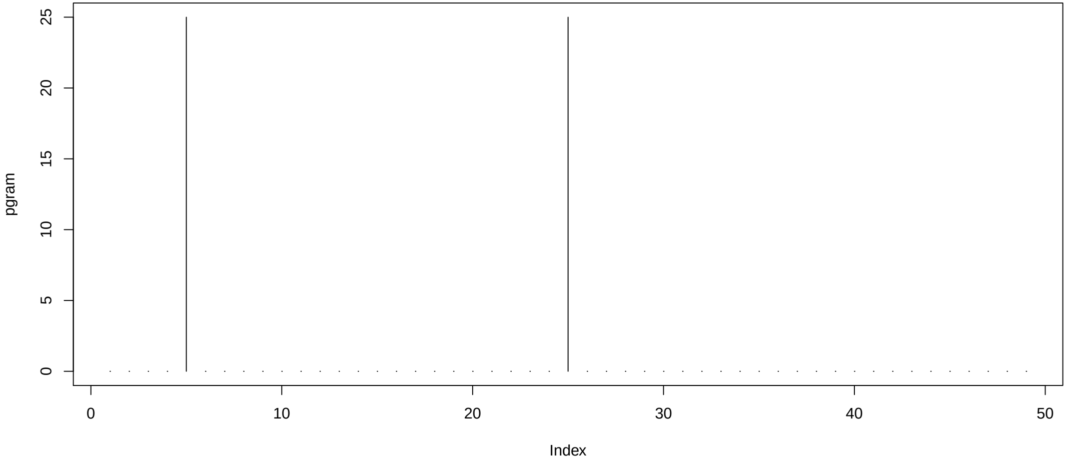
Example Periodogram: $cos(2\pi t * 5/100) + sin(2\pi t * 5/100)$



Example Data: $cos(2\pi t * 25/100) + sin(2\pi t * 5/100)$



Example Periodogram: $cos(2\pi t * 25/100) + sin(2\pi t * 5/100)$



Notes on Periodogram

Recall b_j gives the jth coefficient of the data $x = (x_0, \ldots, x_{n-1})$ in the basis $u^0, \ldots, u^{n-1}$, which corresponds to the sinusoids of Fourier frequency j/n, thus:

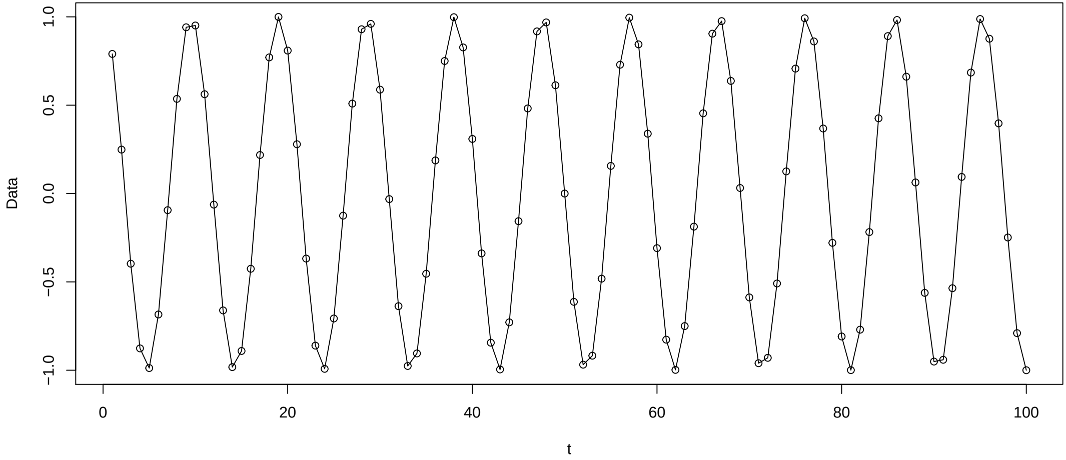
1. If the periodogram shows a single spike for $I(j/n)$ we are sure that the data is a single sinusoid with Fourier frequency j/n.
2. If it shows two spikes, say at $I(j_1/n)$ and $I(j_2/n)$, then the data are a linear combination of two sinusoids at Fourier frequencies j_1/n and j_2/n with the strengths of these sinusoids depending on the size of the spikes.

Notes on Periodogram

3. Multiple spikes indicate that the data is made up of many sinusoids at Fourier frequencies.
4. Sometimes one can see multiple spikes in the DFT even when the structure of the data is not very complicated. A typical example is *leakage* due to the presence of a sinusoid at a non-Fourier frequency.

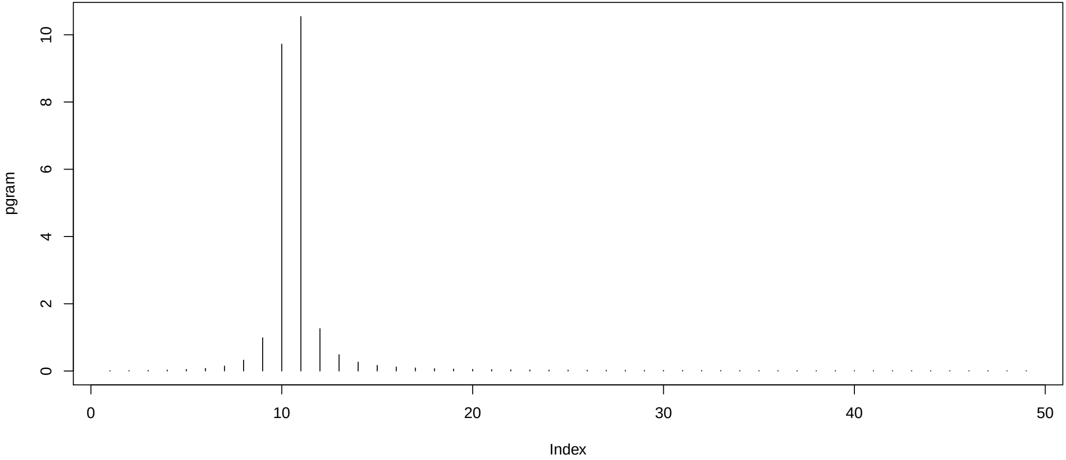
Example Data: $cos(2\pi t * 10.5/100)$

```
t = 1:100; cos2 = cos(2*pi*t*(10.5/100))
plot(t, cos2, ylab = "Data", type = "o")
```



Example Periodogram: $cos(2\pi t * 10.5/100)$

```r
pgram = abs(fft(cos2)[2:50])^2/n
plot(pgram, type = "h")
```



Theorem Intro

The following theorem shows an important relation between periodogram $I(j/n)$ and the sample ACVF $\hat{\gamma}(h)$ of some data $x_0, \ldots, x_{n-1}$.

Theorem: Connection between periodogram and $\hat{\gamma}$

For some data $x_0, \ldots, x_{n-1}$ let $\hat{\gamma}(h)$ for $h = 0, \ldots, n-1$ be its sample ACVF. Then

$$I(j/n) = \sum_{h=-(n-1)}^{n-1} \hat{\gamma}(h) \exp\left(-\frac{2\pi ijh}{n}\right) \text{ for } j = 1, \ldots, \lfloor n/2 \rfloor.$$

Proof (page 1; skipping in class)

First, by the formula for the sum of a geometric series, observe that

$$\sum_{t=0}^{n-1} \exp\left(-\frac{2\pi ijt}{n}\right) = 0 \text{ for } j = 1, \ldots, \lfloor n/2 \rfloor.$$

In other words, if the data is constant i.e., $x_0 = \cdots = x_{n-1}$, then b_0 equals nx_0 and b_j equals 0 for all other j. Because of this, we can write:

$$b_j = \sum_{t=0}^{n-1} (x_t - \bar{x}) \exp\left(-\frac{2\pi ijt}{n}\right) \text{ for } j = 1, \ldots, \lfloor n/2 \rfloor.$$

Proof (page 2; skipping in class)

Therefore, for $j = 1, \ldots, \lfloor n/2 \rfloor$, we write

$$
\begin{aligned}
|b_j|^2 = b_j \bar{b}_j &= \sum_{t=0}^{n-1} \sum_{s=0}^{n-1} (x_t - \bar{x})(x_s - \bar{x}) \exp\left(-\frac{2\pi ijt}{n}\right) \exp\left(\frac{2\pi ijs}{n}\right) \\
&= \sum_{t=0}^{n-1} \sum_{s=0}^{n-1} (x_t - \bar{x})(x_s - \bar{x}) \exp\left(-\frac{2\pi ij(t-s)}{n}\right) \\
&= \sum_{h=-(n-1)}^{n-1} \sum_{t,s:t-s=h} (x_t - \bar{x})(x_{t-h} - \bar{x}) \exp\left(-\frac{2\pi ijh}{n}\right) \\
&= n \sum_{h=-(n-1)}^{n-1} \hat{\gamma}(h) \exp\left(-\frac{2\pi ijh}{n}\right).
\end{aligned}
$$