The Periodogram, part 2

Jared Fisher

Lecture 3a

Announcements

Announcements

- Project Checkpoint 1 is due tomorrow: Wednesday Feb 10
- Fifth lab section: I'll let you know as soon as I hear.
- Homework 2 is due next week, Wednesday Feb 17
- Midterm 1 is on Thursday Feb 25

Recap

Full Model

$$Y_t = m_t + s_t + X_t$$

- m_t is the trend model
- s_t is the model of seasonal effects (e.g. sinusoids)
- 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.

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 showed this in Lab 2).
- 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!

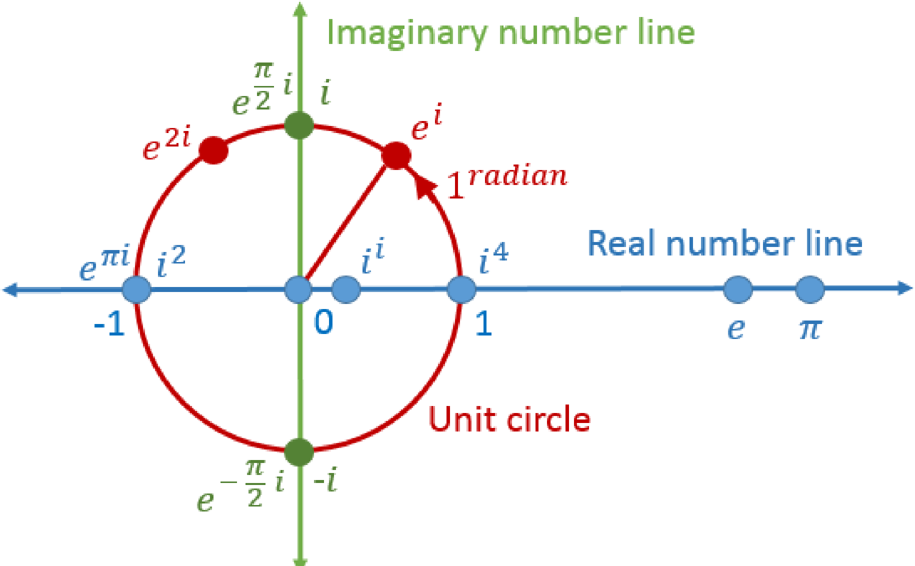
Textbook

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

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: $e^{i\theta} = cos(\theta) + i * sin(\theta)$

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

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

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

(Proof on Lecture 2b slides)

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 $|b_j|, 1 \leq j \leq n/2$.

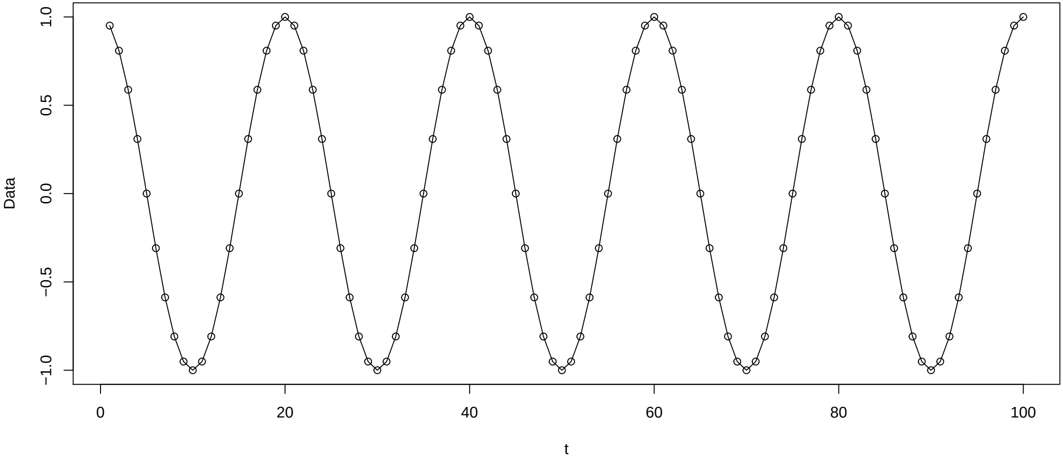
Definition: Periodogram

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

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

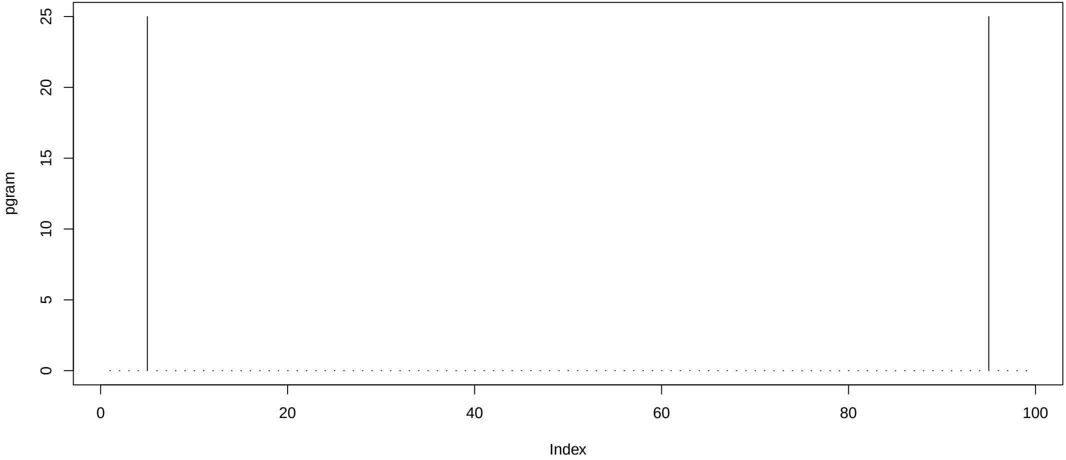
Example Data: $cos(2\pi t * 5/100)$

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



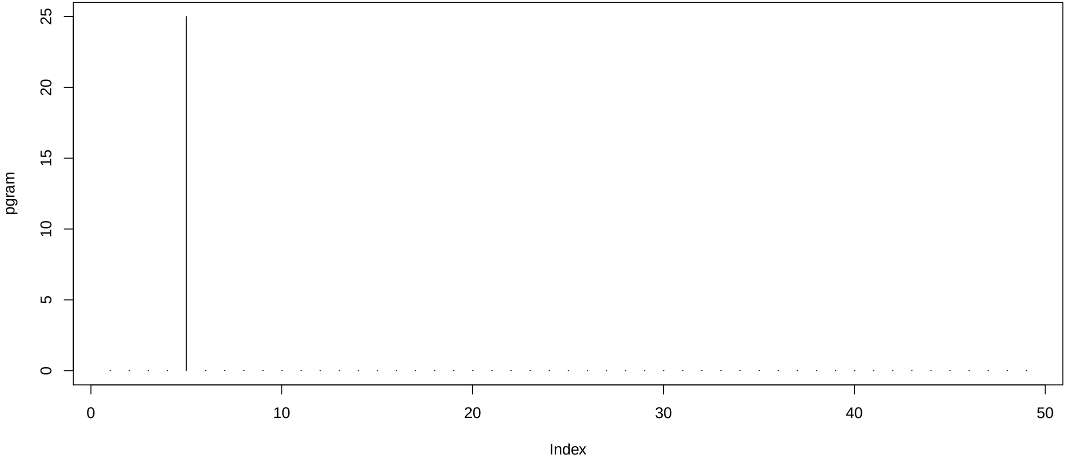
Example: $cos(2\pi t * 5/100)$

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

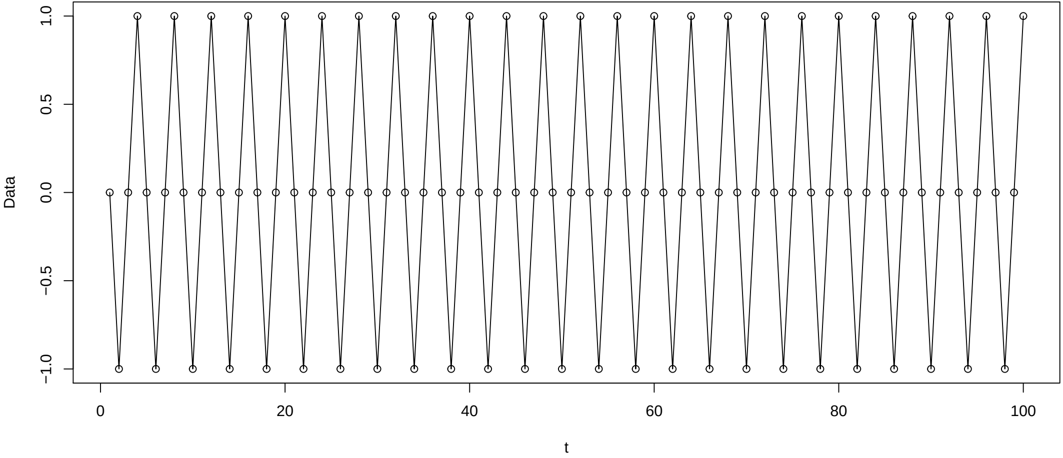


Example Periodogram: $cos(2\pi t * 5/100)$

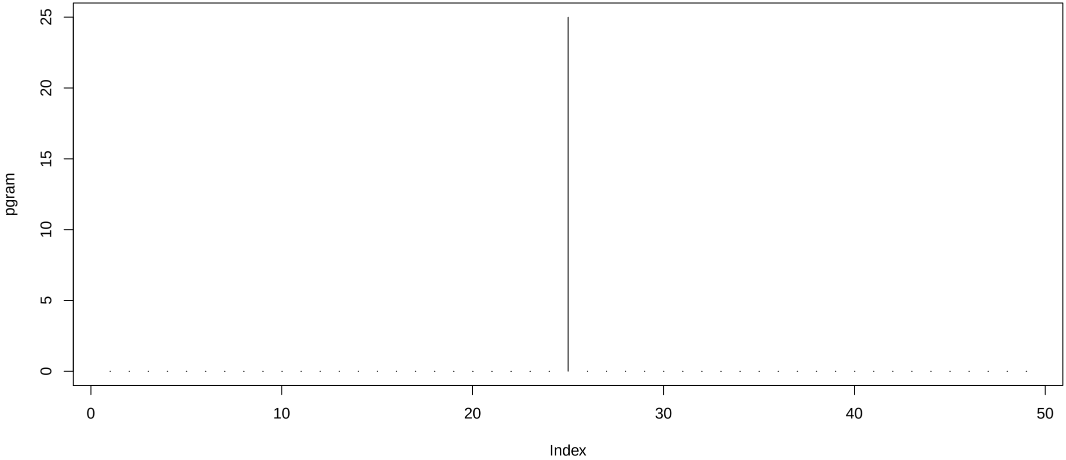
```
pgram = abs(fft(cos2)[2:50])^2/n
plot(pgram, type = "h")
```



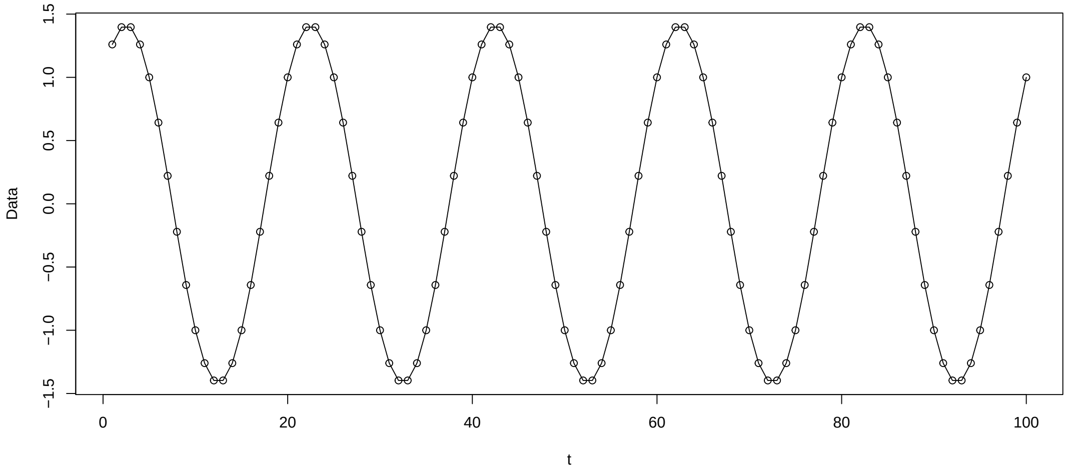
Example Data: $cos(2\pi t * 25/100)$



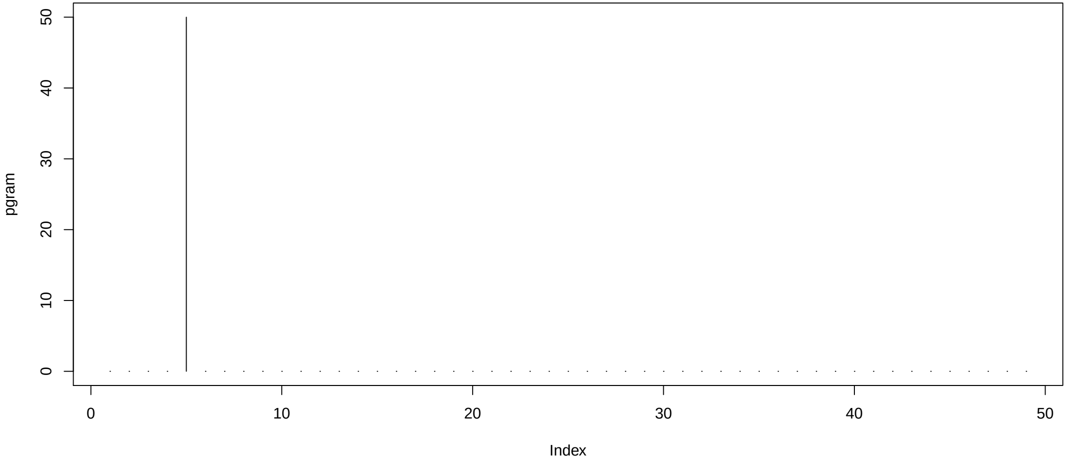
Example Periodogram: $cos(2\pi t * 25/100)$



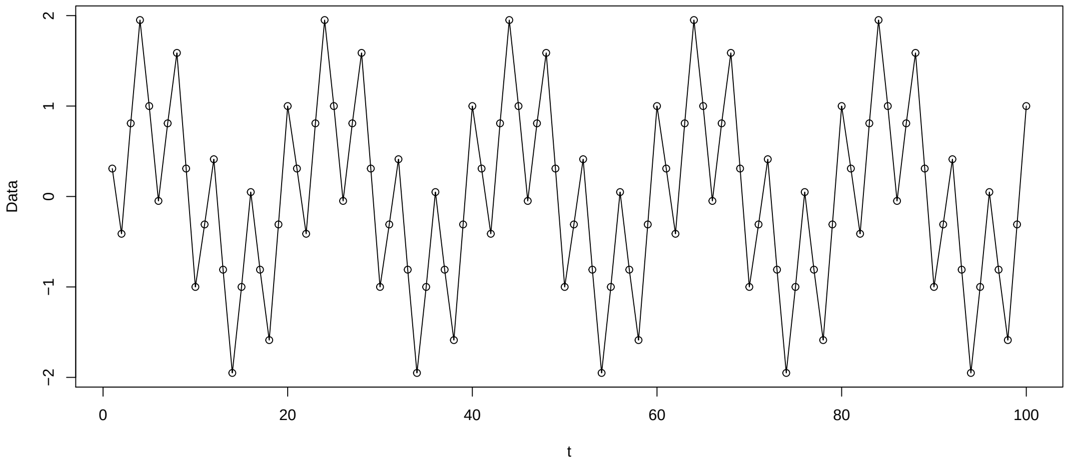
Example Data: $cos(2\pi t * 5/100) + sin(2\pi t * 5/100)$



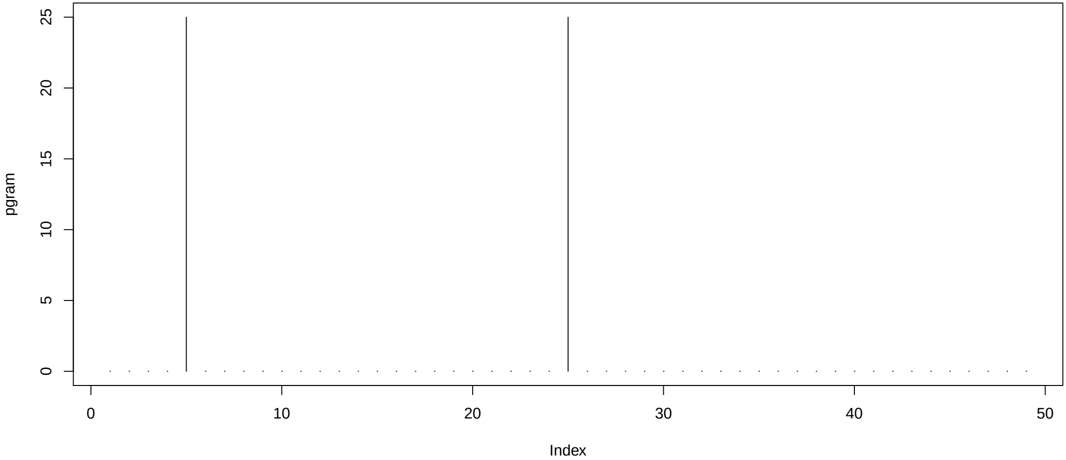
Example Periodogram: $cos(2\pi t * 5/100) + sin(2\pi t * 5/100)$



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:

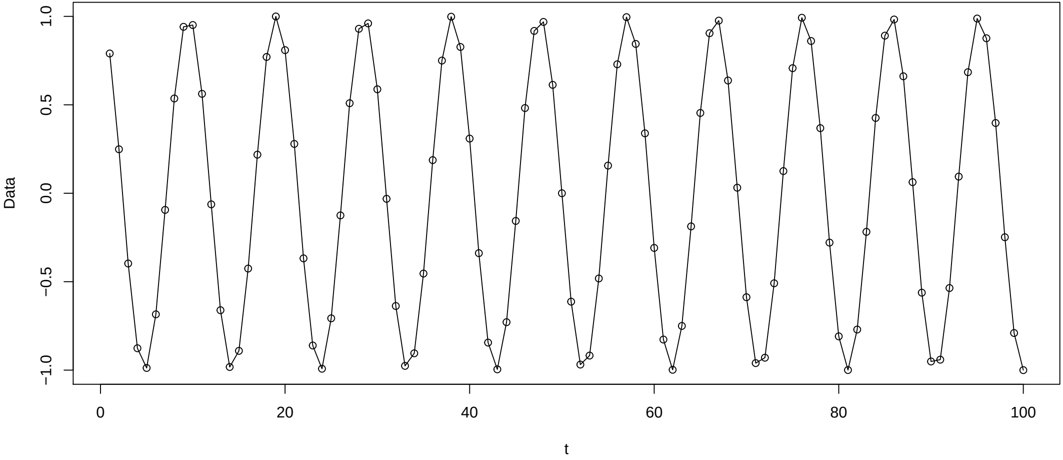
1. If the periodogram shows a single spike for $I(j/n)$ we are sure that the data is a single sinusoid with Fourier frequency j/n.
2. If it shows two spikes, say at $I(j_1/n)$ and $I(j_2/n)$, then the data are a linear combination of two sinusoids at Fourier frequencies j_1/n and j_2/n with the strengths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

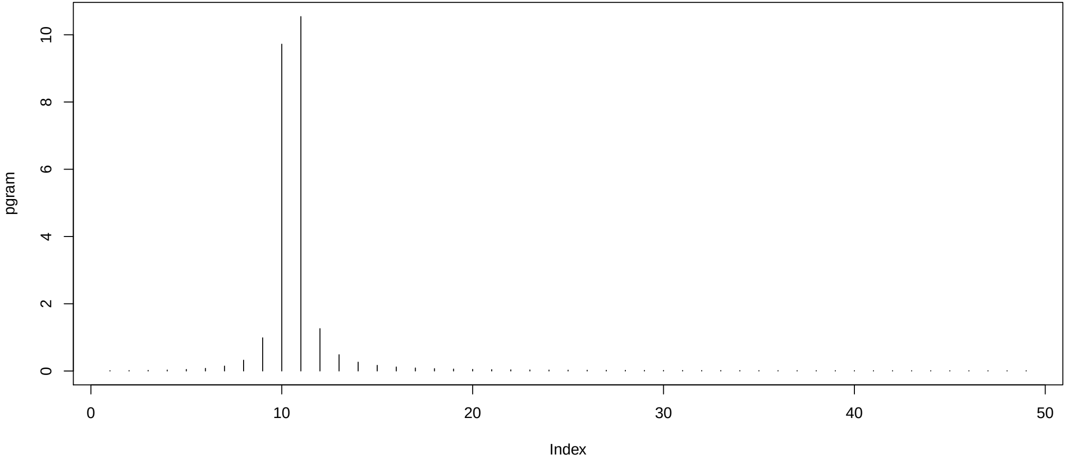
Example Data: $cos(2\pi t * 10.5/100)$

```
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 on Lecture 2b slides)

Next

The remainder of today's material will be in R!