TSWGE Realization Generation Sine Cosine

Segment 2.3: Screen Share

In this screen cast, we will use tswge to generate realizations from a signal plus noise model with a sin and/or cos signal.

Description of the variables:

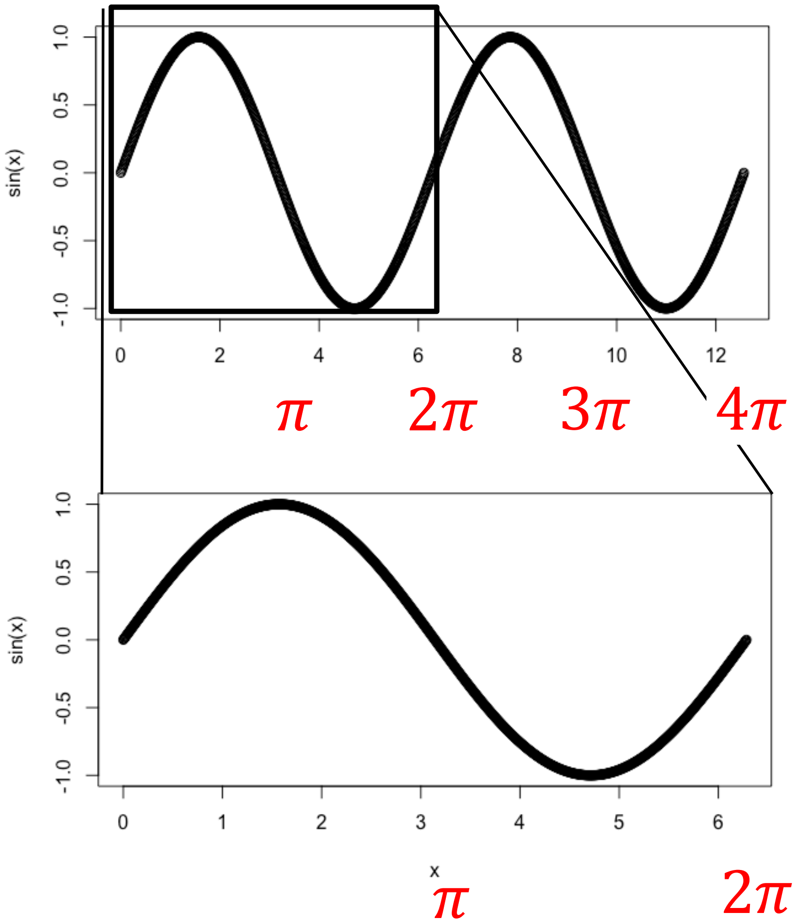
Identify the function… Xt = ….

```
#Code for generating random realizations from a NON stationary process
Xt1 = gen.sigplusnoise.wge(n = 1000,freq=c(.005,0),psi = c(runif(1,0,2*pi),0),vara = .00, coef = c(1,0),plot = TRUE)
Xt2 = gen.sigplusnoise.wge(n = 1000,freq=c(.005,0),psi = c(runif(1,0,2*pi),0),vara = .00, coef = c(1,0),plot = TRUE)
Xt3 = gen.sigplusnoise.wge(n = 1000,freq=c(.005,0),psi = c(runif(1,0,2*pi),0),vara = .00, coef = c(1,0),plot = TRUE)
par(mfrow = c(3,1))
plot(seq(0,1000,length = 1000),Xt1,main = "Realization 1 (Seed = 1):  cos(t + 0)",type = "l",xlab = "",ylab = "", cex.axis = 1.5)
plot(seq(0,1000,length = 1000),Xt2,main = "Realization 2 (Seed = 2):  cos(t + 0)",type = "l",xlab = "",ylab = "", cex.axis = 1.5)
plot(seq(0,1000,length = 1000),Xt3,main = "Realization 3 (Seed = 3):  cos(t + 0)",type = "l",xlab = "",ylab = "", cex.axis = 1.5)
```

DataScience@SMU

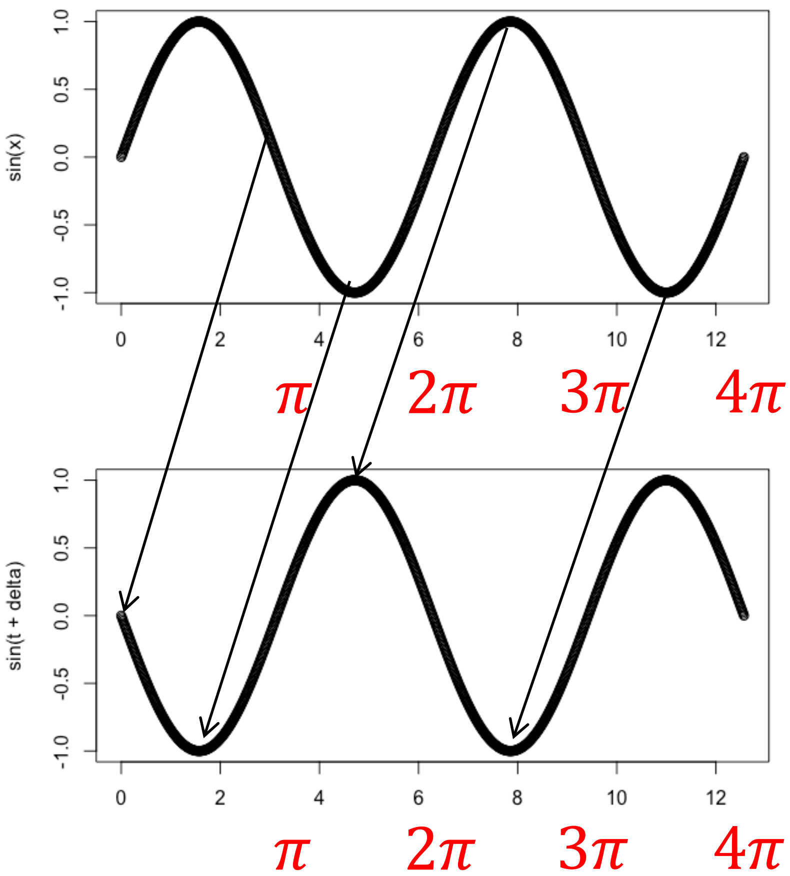
Sine and Cosine Review

Sin(t)



Period = 2π

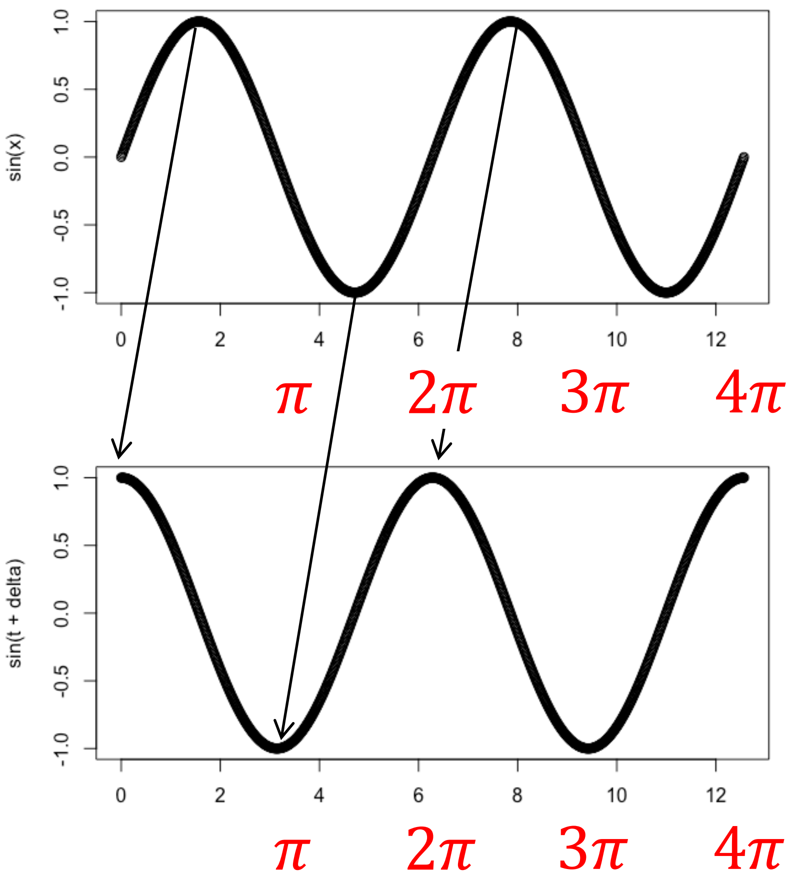
Sin(t+**Δ**): Phase Shift by **Δ**
Horizontal Shift Left by **Δ**



$\boldsymbol{\Delta} = 0$ $\qquad\sin(t)$

$\boldsymbol{\Delta} = \pi$ $\qquad\sin(t + \pi)$

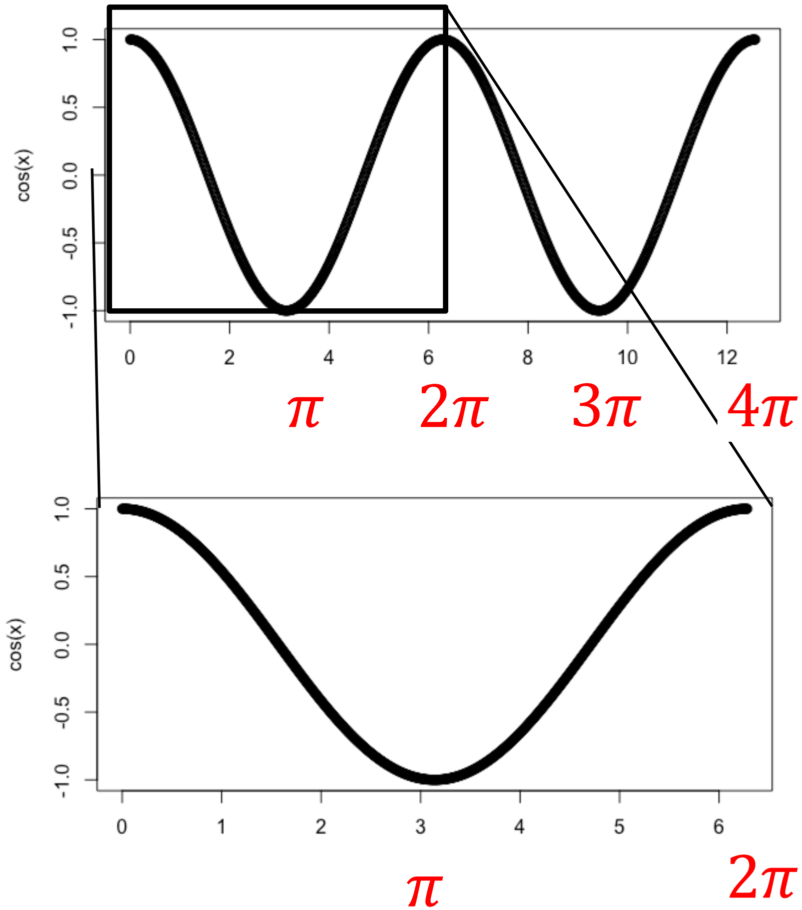
Sin(t+Δ): Phase Shift by Δ
Horizontal Shift Left by Δ



$\boldsymbol{\Delta} = 0$ $\sin(t)$

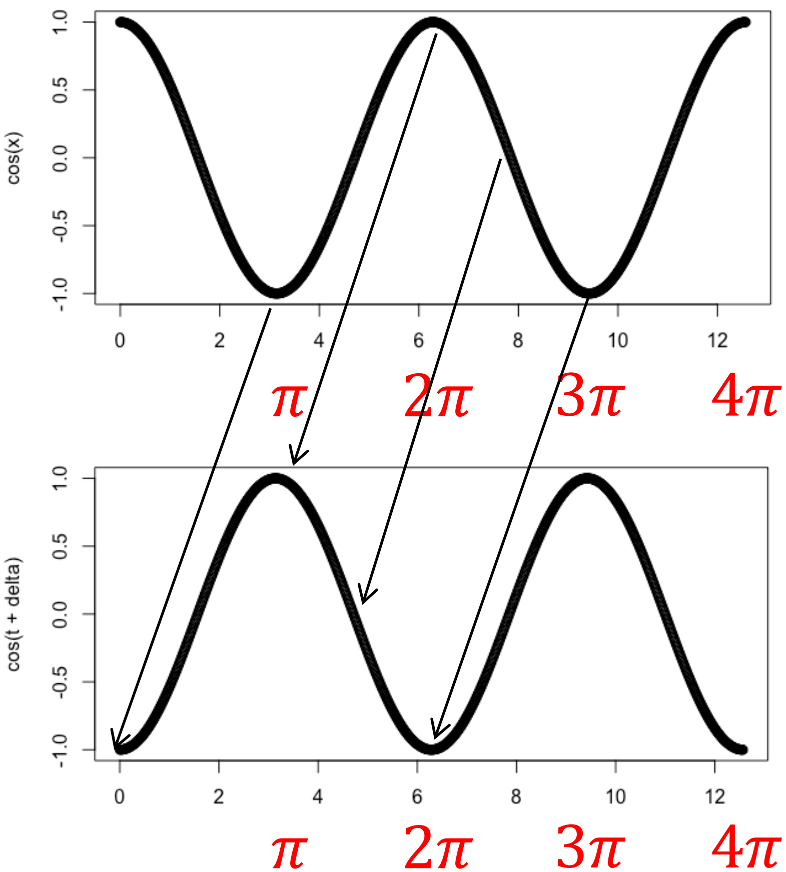
$\boldsymbol{\Delta} = 1.57$ $\sin(t + 1.57)$

Cos(t)



Period = 2π

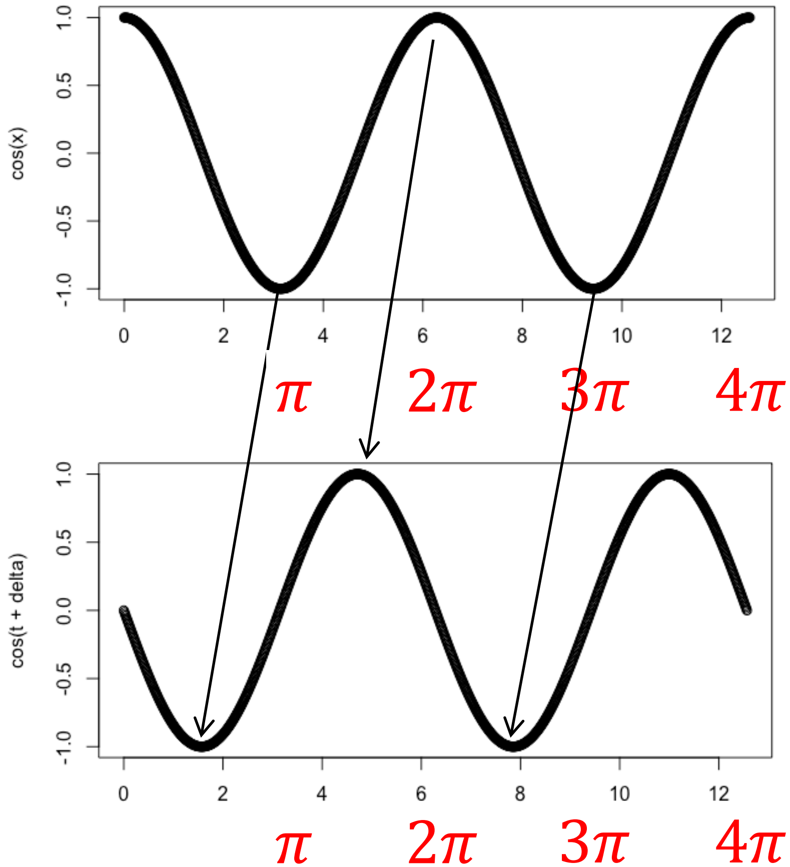
Cos(t+Δ): Phase Shift by Δ
Horizontal Shift Left by Δ



$\boldsymbol{\Delta} = 0 \qquad \cos(t)$

$\boldsymbol{\Delta} = \pi \qquad \cos(t + \pi)$

Cos(t+Δ): Phase Shift by Δ
Horizontal Shift Left by Δ



$\boldsymbol{\Delta} = 0 \qquad \cos(t)$

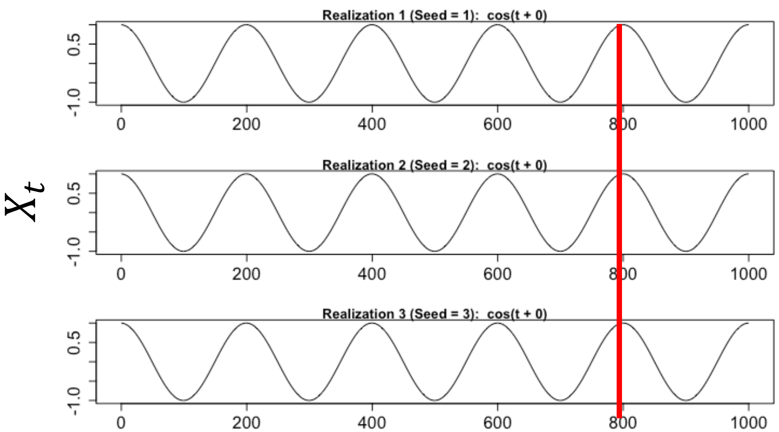
$\boldsymbol{\Delta} = 1.57 \quad \cos(t + 1.57)$

DataScience@SMU

Stationary and Nonstationary Time Series

Below are 3 realizations from 2 different time series. Which provides the most evidence of a stationary time series?

$$X_t = \cos(t + \Delta)$$
$$\Delta = 0$$



$\Delta = 0$

$\Delta = 0$

$\Delta = 0$

$Time$

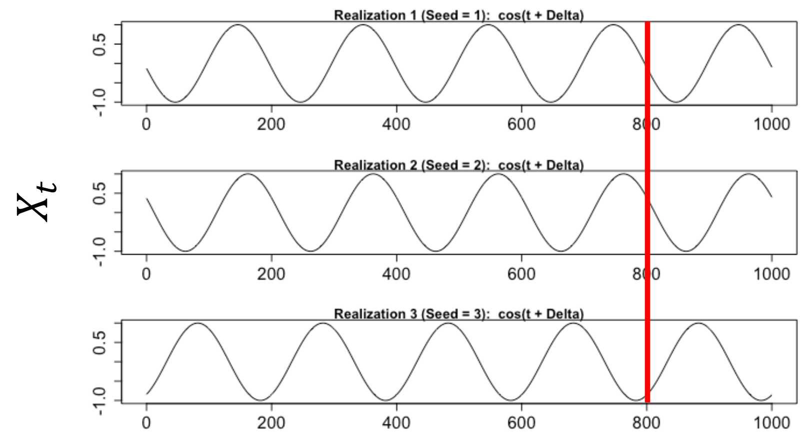
There appears to be evidence against stationarity.

$Condition\ 1$:

$The\ mean\ of\ X_t\ \left(E[X_t] or\ \mu_{X_t}\right)$

$appears\ to\ depend\ on\ t$.

$$X_t = \cos(t + \Delta)$$
$$\Delta \sim \text{Unif}(0, 2\pi)$$



$\Delta = 1.5$

$\Delta = .7$

$\Delta = 3.7$

$Time$

The evidence appears to be consistent with a stationary series.

$Condition\ 1$:

$The\ mean\ of\ X_t\ \left(E[X_t] or\ \mu_{X_t}\right)$

$appears\ to\ be\ independent\ of\ t$.

DataScience@SMU

Frequency/Period Review

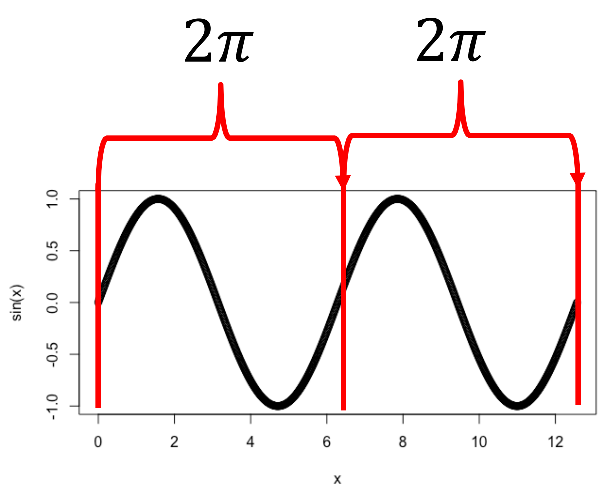
Spectral Analysis:

Analysis of Periodicities in the Data

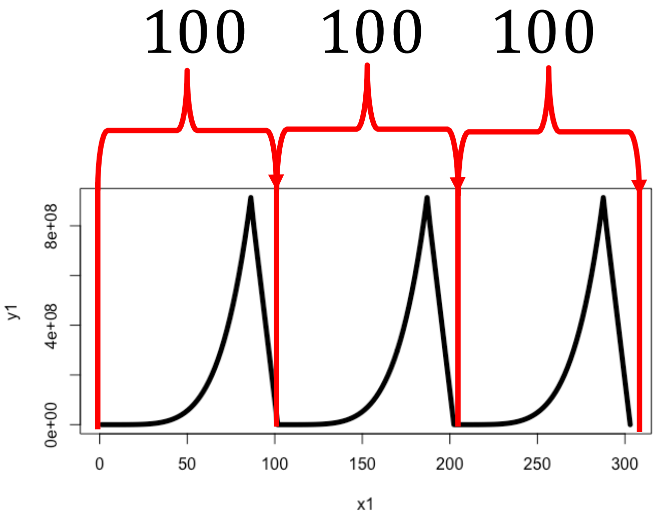
Technical details

Periodic functions

- $f(x)$ is periodic function with period p if p is the smallest value such that $f(x) = f(x + kp)$ for all x and integer k.



Period = 2π



Period = 100

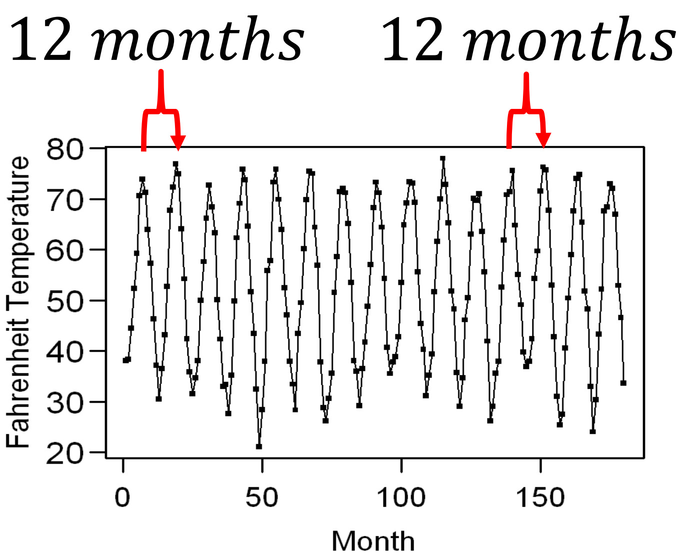
Spectral Analysis:

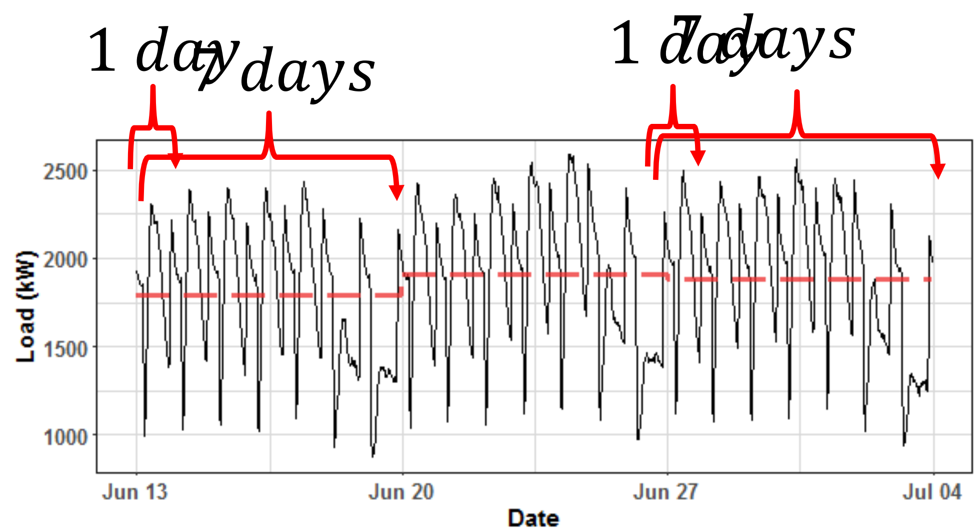
Analysis of Periodicities in the Data

Technical details

Psuedo-periodic data

- Data are pseudo-periodic with period p if p is the smallest value such that a cycle appears to repeat itself.





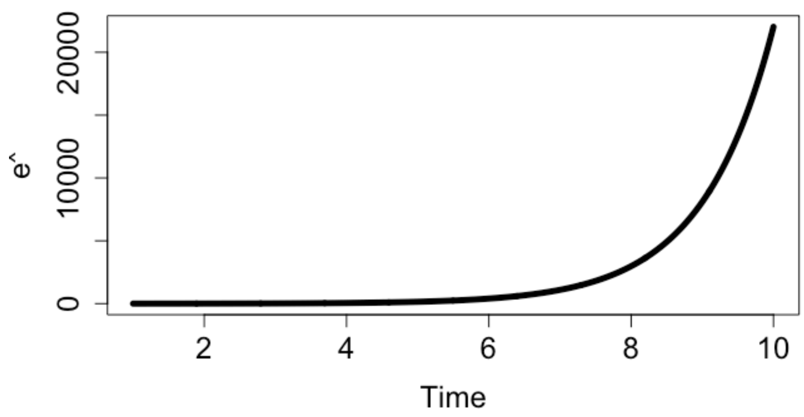
Spectral Analysis:

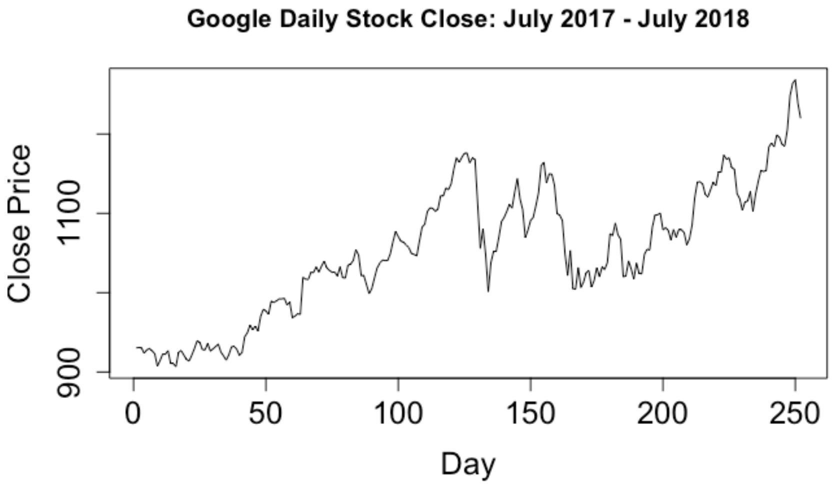
Analysis of Periodicities in the Data

Technical details

Aperiodic functions/data

- $f(x)$ is non-periodic (*aperiodic*) if no such p exists.





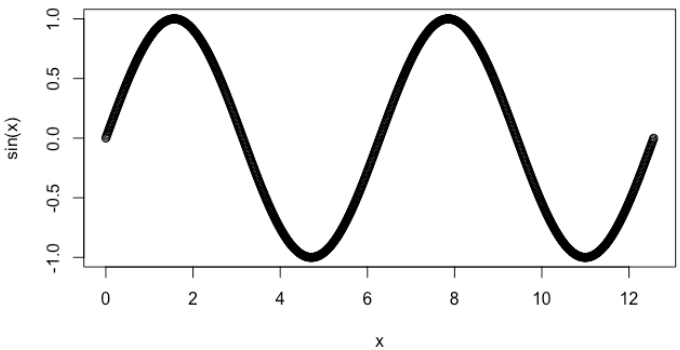
Spectral Analysis:

Analysis of Periodicities in the Data

Technical details

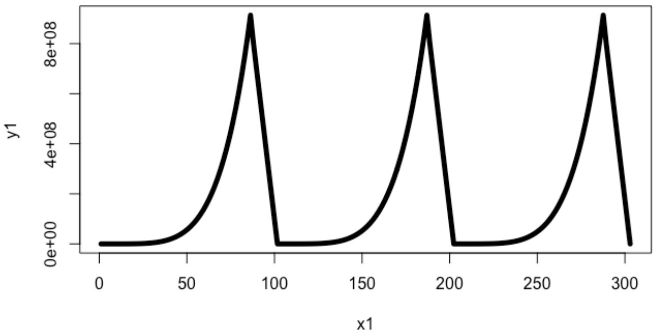
Frequency = number of cycles per unit (period = cycle)

= 1/period



Period = 2π

Frequency = $\frac{1}{2\pi} = .159$



Period = 100

Frequency = $\frac{1}{100} = .01$

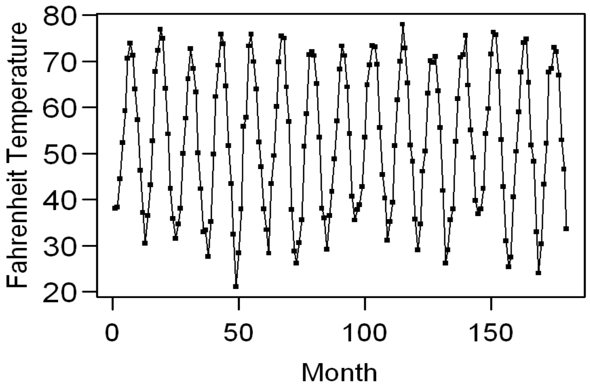
Spectral Analysis:

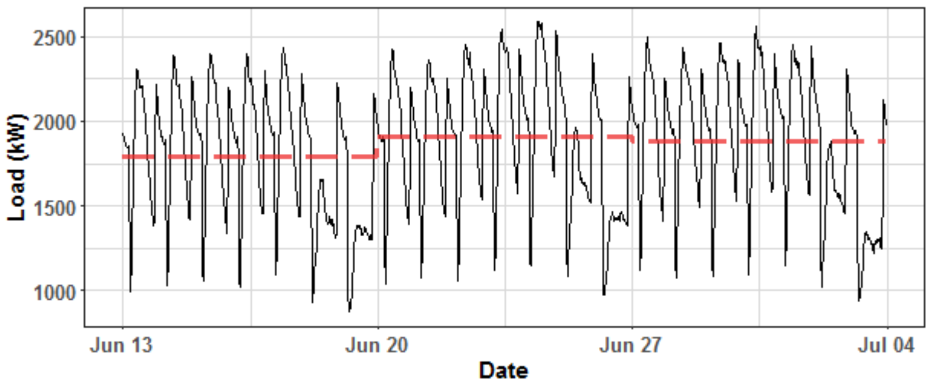
Analysis of Periodicities in the Data

Technical details

Frequency = number of cycles per unit (period = cycle)

= 1/period





Period = 12 months

Frequency = $\frac{1}{12} = .083$

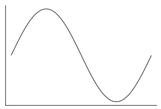
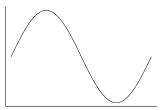
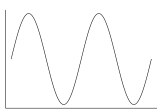
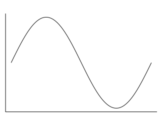
Period = 24 hours

Frequency = $\frac{1}{24} = .0146$

Period = 7 days = 168 hours

Frequency = $\frac{1}{168} = .006$

Frequency Terminology

		Period	Frequency
$\sin t$		2π	$1/(2\pi)$
$\sin 2\pi t$		1	1
$\sin 2\pi 2t$		$1/2$	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\sin 2\pi f t$		$1/f$	f

Period and Frequency of Sin and Cos

Consider:

$$\sin(Bt + C)$$

Period = $\frac{2\pi}{B}$

Frequency = $\frac{B}{2\pi}$

Example:

$$\sin(1t)$$

Period = $\frac{2\pi}{1} = 2\pi$

Frequency = $\frac{1}{2\pi} = .159$

Example:

$$\sin(2\pi(.19)t)$$

Period = $\frac{2\pi}{2\pi(.19)} = \frac{1}{.19} = 5.26$

Frequency = $\frac{2\pi(.19)}{2\pi} = .19$

* Similar result for cos

Segment 2: 2.2 Instructor Concept Check $\sin(Bt + C)$

Consider the function and corresponding realization:

$$\sin(2\pi(.17)t)$$

What is the *B*?

What is the period?

What is the frequency?

Segment 2: 2.2 Instructor Concept Check $\sin(Bt + C)$

Consider the function and corresponding realization:

$$\sin(2\pi(.17)t)$$

What is the *B*? $2\pi(.17)$

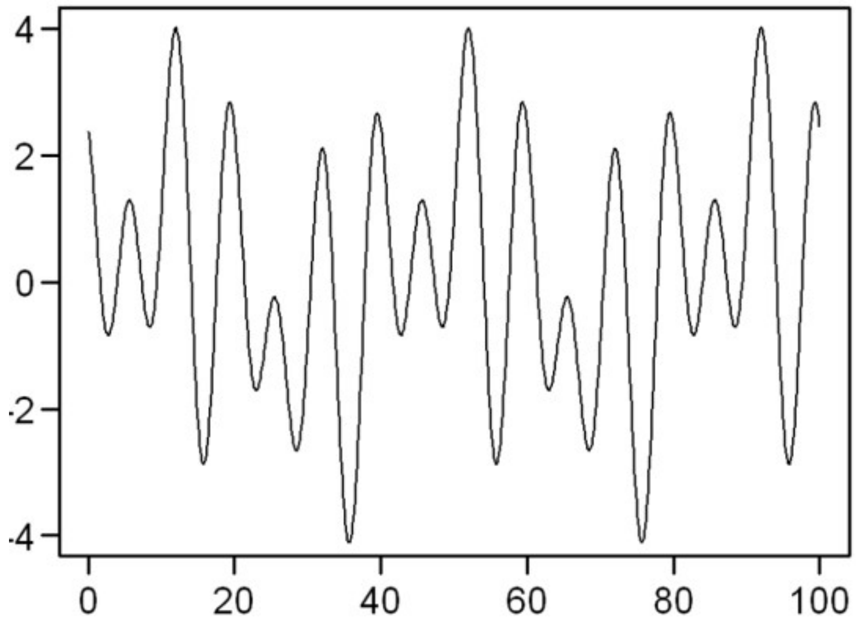
What is the frequency? $\frac{B}{2\pi} = \frac{2\pi(.17)}{2\pi} = .17$

What is the period? $\frac{2\pi}{B} = \frac{2\pi}{2\pi(.17)} = \frac{1}{(.17)} = \frac{1}{Frequency} = 5.88$

DataScience@SMU

Screencast for Composite Sine Function with *R*

A Series Containing Several Cycles

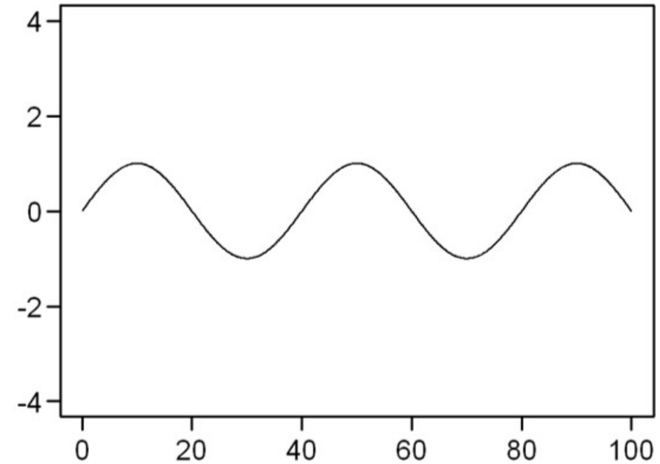


Question:

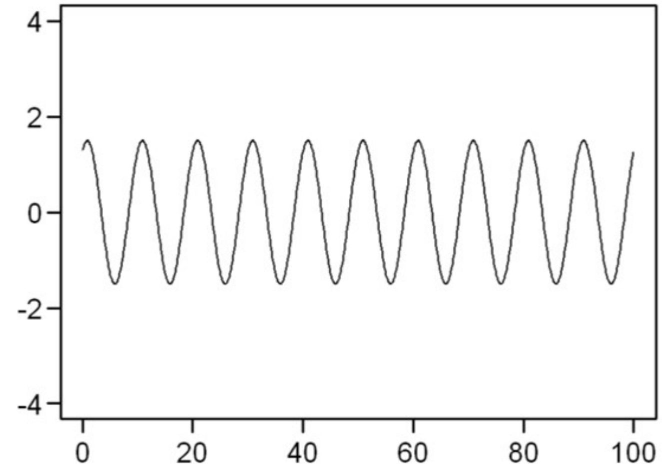
What cycles are present in the data ?

Actual Data Are a Sum of:

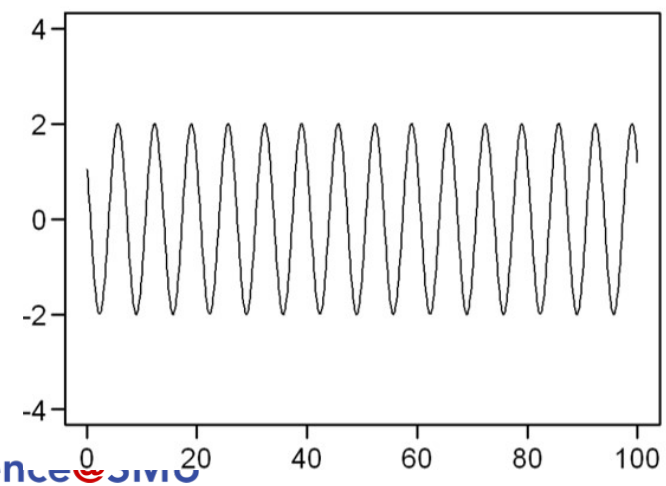
Cycle length 40 (freq. = 1/40)



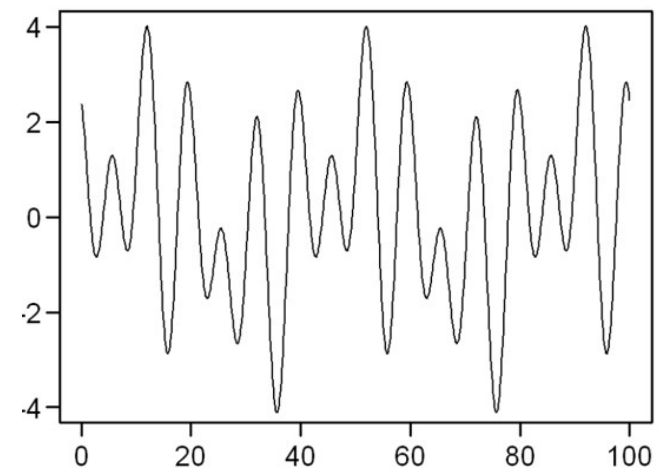
Cycle length 10 (freq. = 1/10)



Cycle length 6.67 (freq. = .15)



Sum



The Actual Signal Is the Sum

$$X(t) = \sin(2\ \ (.025)t) + 1.5\sin(2\ \ (.1)t + 1) + 2\sin(2\ \ (.15)t + 2.5)$$

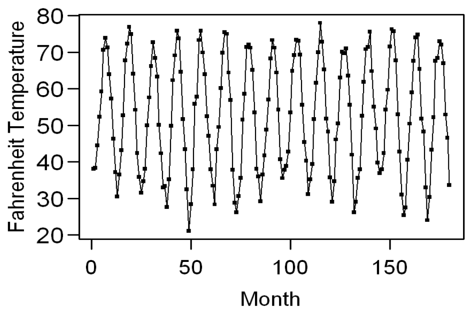
```
t = seq(1,100,length = 100)
y1 = sin(2*pi*.025*t)
y2 = sin(2*pi*.1*t+1)
y3 = sin(2*pi*.15*t+2.5)
ysum = y1+y2+y3
plot(t,ysum,type = "l")
```

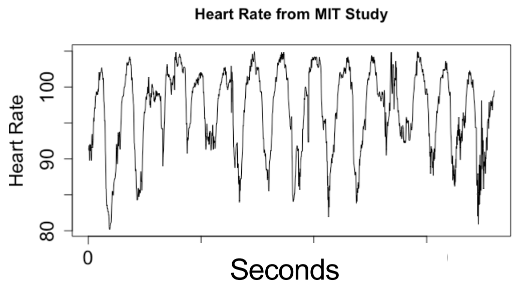
DataScience@SMU

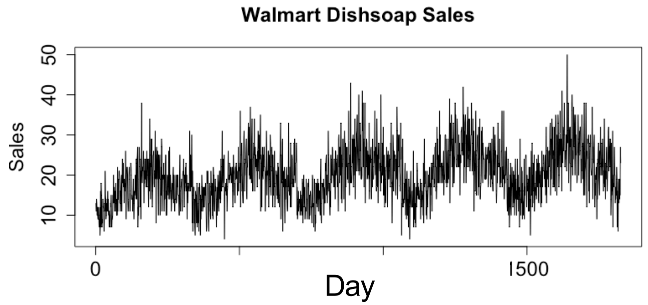
Fundamental Idea

Fundamental Ideas

- Many realizations are made up of various frequency components, while others do not have a frequency component.
- If we can discover the *frequency content* of a set of data (or lack thereof), then we can better understand the process generating the data.
  - That is, better understand the physical process
  - *Examples*: weather, heart beat, retail sales, and so on







DataScience@SMU

DataScience@SMU

Representing Functions as Sines and Cosines

Fourier Expansion

Spectral Analysis

The decomposition of functions (usually depending on time) into a sum of sine and cosine terms

Fourier Series

For a broad class of functions:

$$f(x) = \lim_{n \to \infty} \left\{ \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx \right\}$$

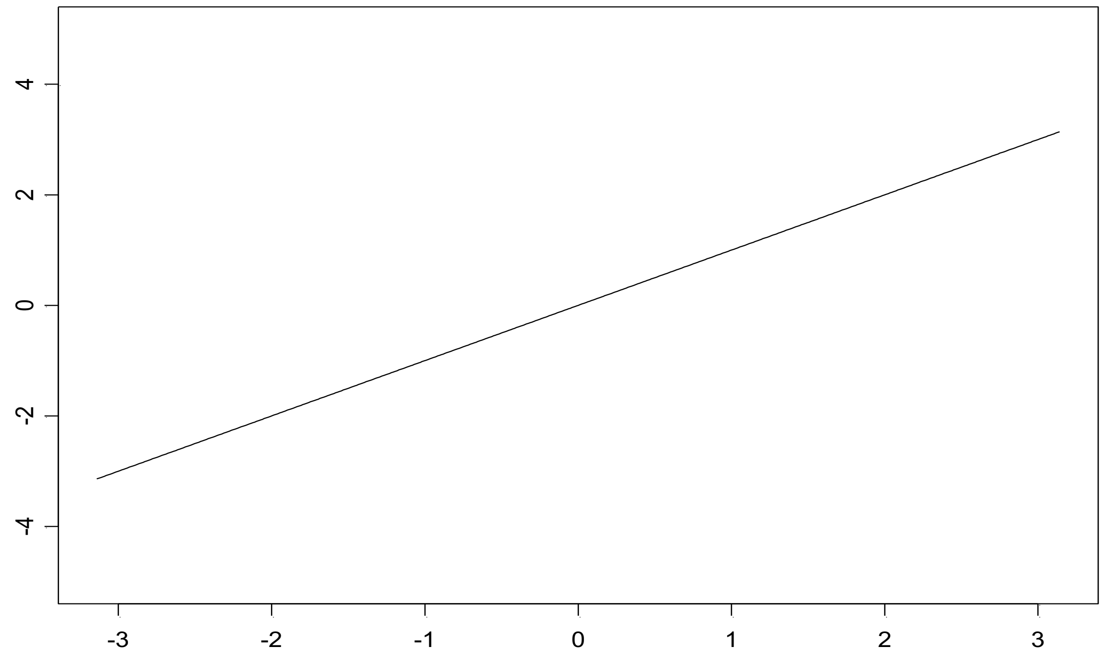
In the case where f is defined over $[-\pi, \pi]$, the coefficients are given by:

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos kt \, dt, \qquad k = 0,1, \ldots$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin kt \, dt, \qquad k = 0,1, \ldots$$

Example: Fourier Expansion for $f(x) = x$

$\boldsymbol{f(x) = x}$



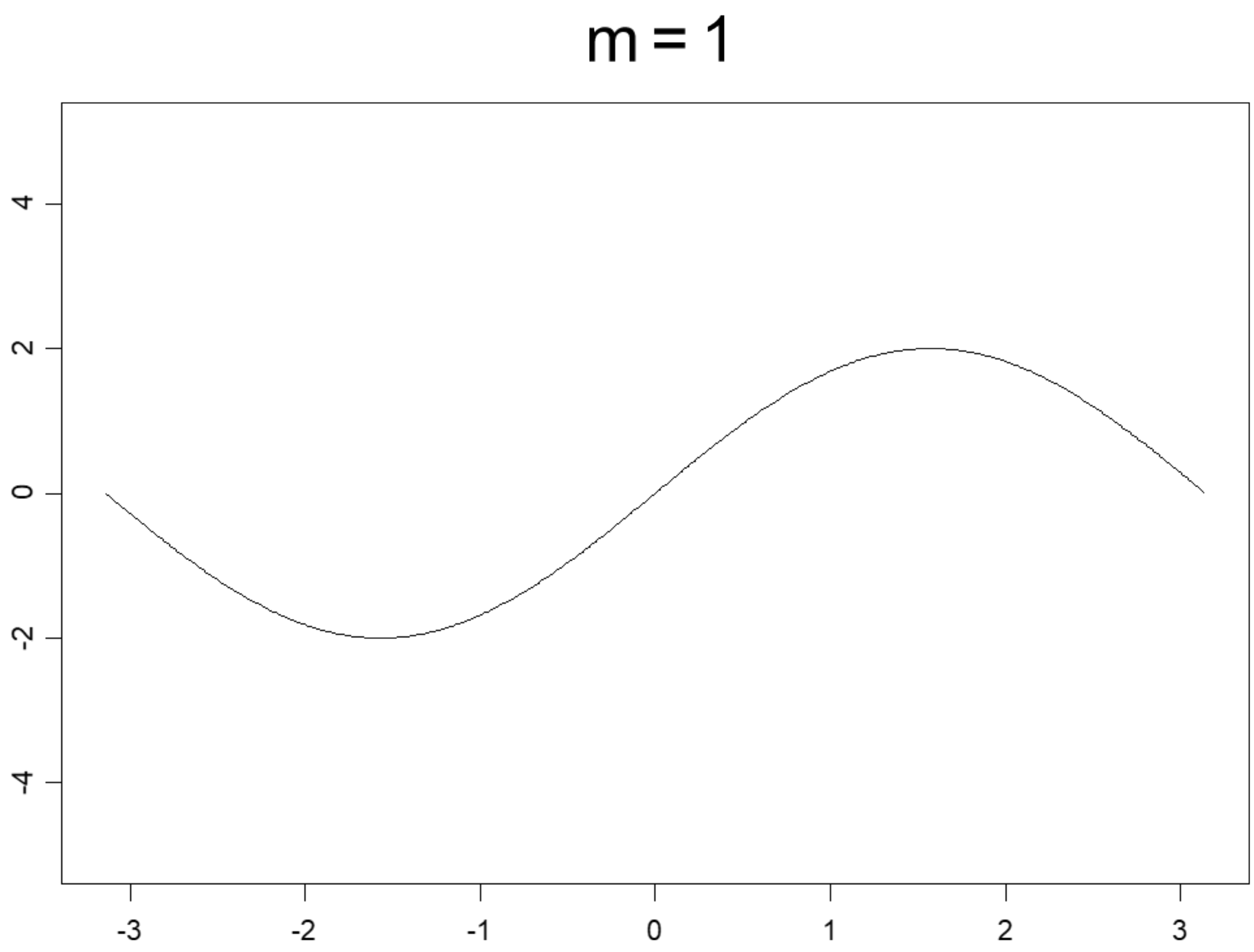
$$x = 2 \quad {}_{n=1} \quad \frac{(\quad 1)^n}{n} \sin(nx), \qquad x \quad (\qquad , \quad)$$

Example: Fourier Expansion for $f(x) = x$

$$x = \quad 2 \quad \underset{n=1}{} \frac{(\quad 1)^n}{n} \sin(nx), \qquad x \quad (\qquad , \quad)$$

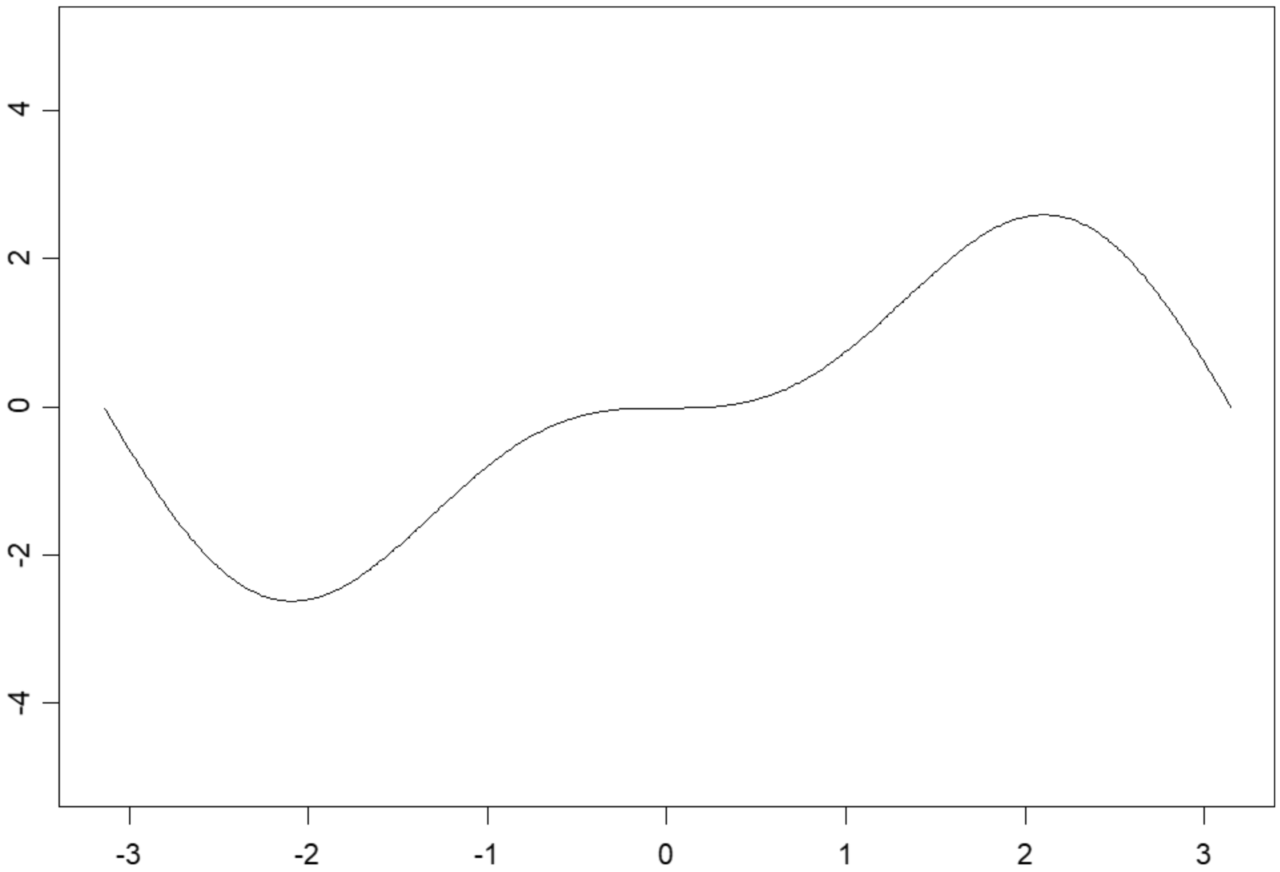
Fourier approximation – sum **m** terms

$$x \qquad 2 \quad \overset{\mathbf{m}}{\underset{n=1}{}} \frac{(\quad 1)^n}{n} \sin(nx), \qquad x \quad (\qquad , \quad)$$

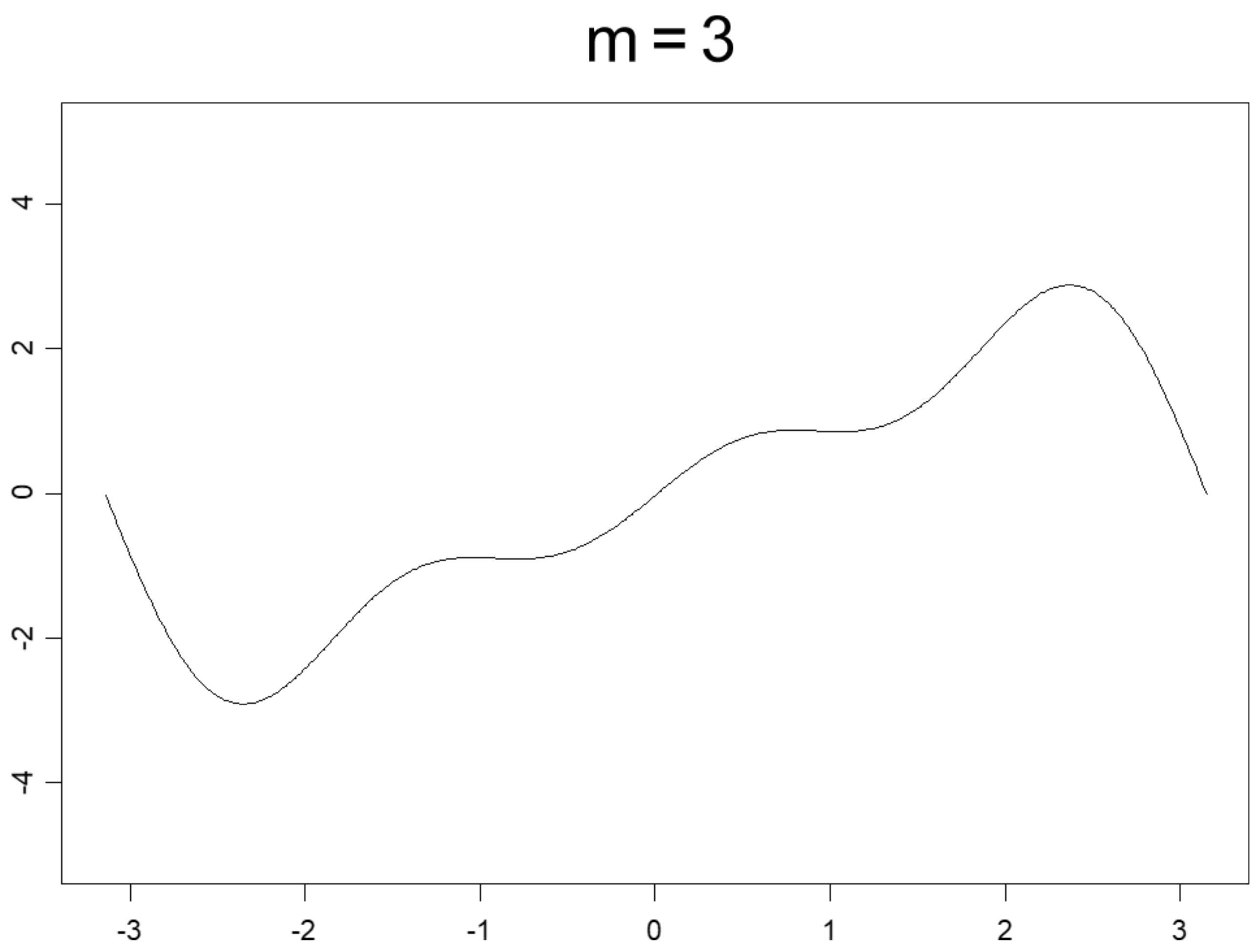


$$x \qquad 2 \, {}^{1}_{n=1} \frac{(\quad 1)^{n}}{n} \sin(nx) = 2\sin(x), \qquad x \quad (\qquad , \quad)$$

m = 2

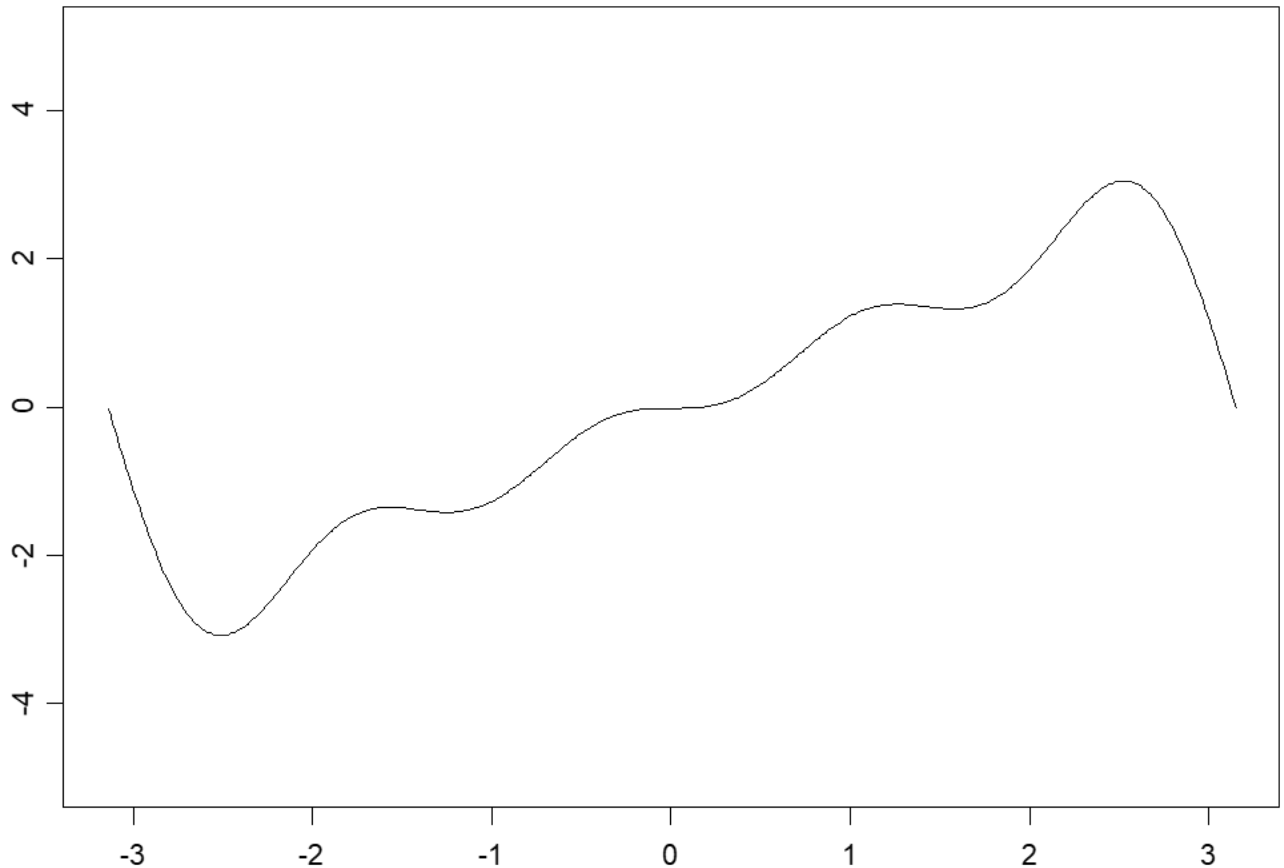


$$x \quad 2 \;{}_{n=1}^{2} \frac{(\quad 1)^{n}}{n}\sin(nx) = \quad 2 \quad \sin(x) \quad \frac{1}{2}\sin(2x) \quad = 2\sin(x)+\sin(2x), \quad x \quad (\quad , \quad)$$

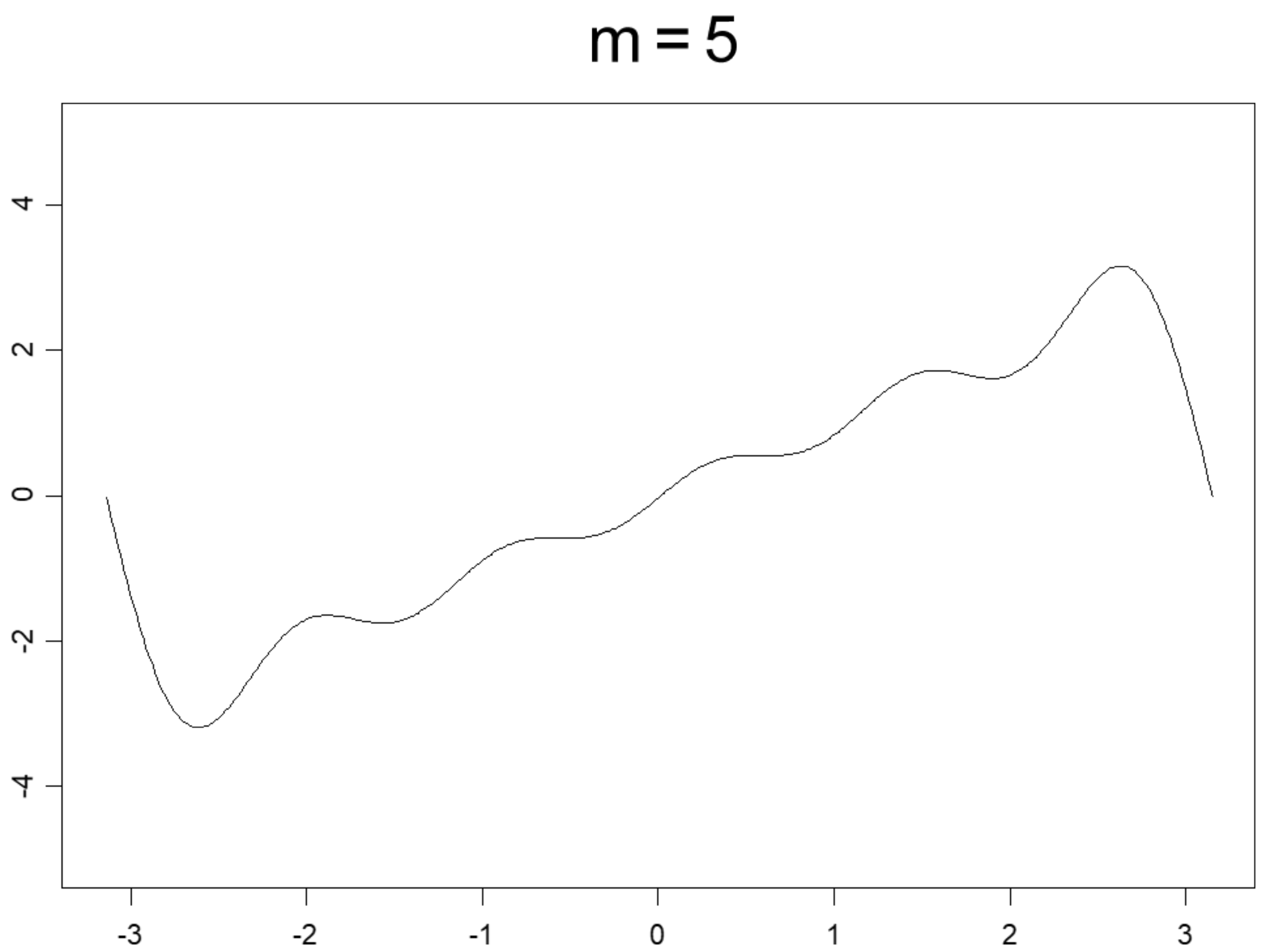


$$x \qquad 2 \;{}_{n=1}^{3} \frac{(\quad 1)^{n}}{n} \sin(nx) = 2\sin(x) + \sin(2x) + \frac{2}{3}\sin(3x), \qquad x \quad (\quad , \quad)$$

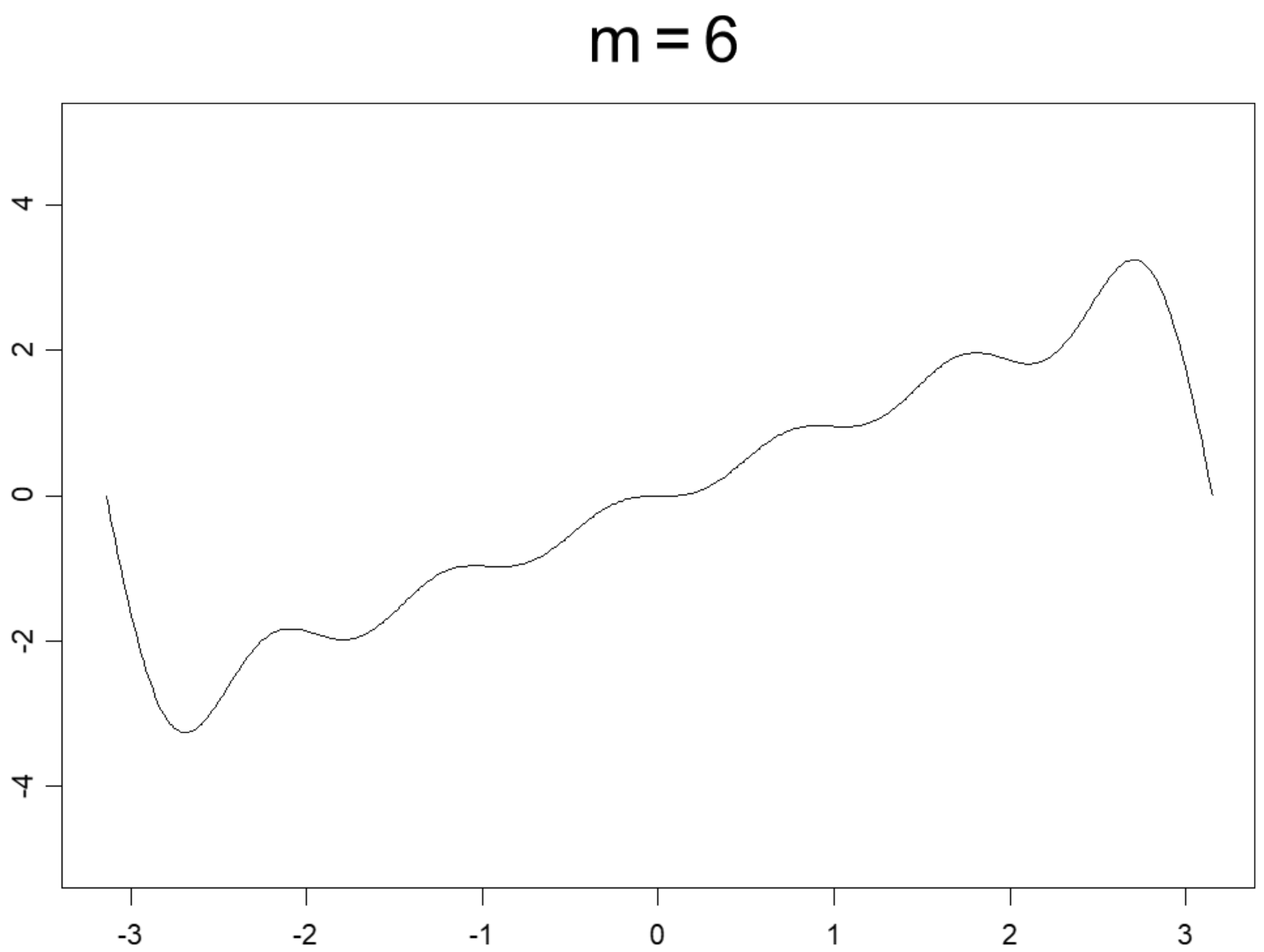
m = 4



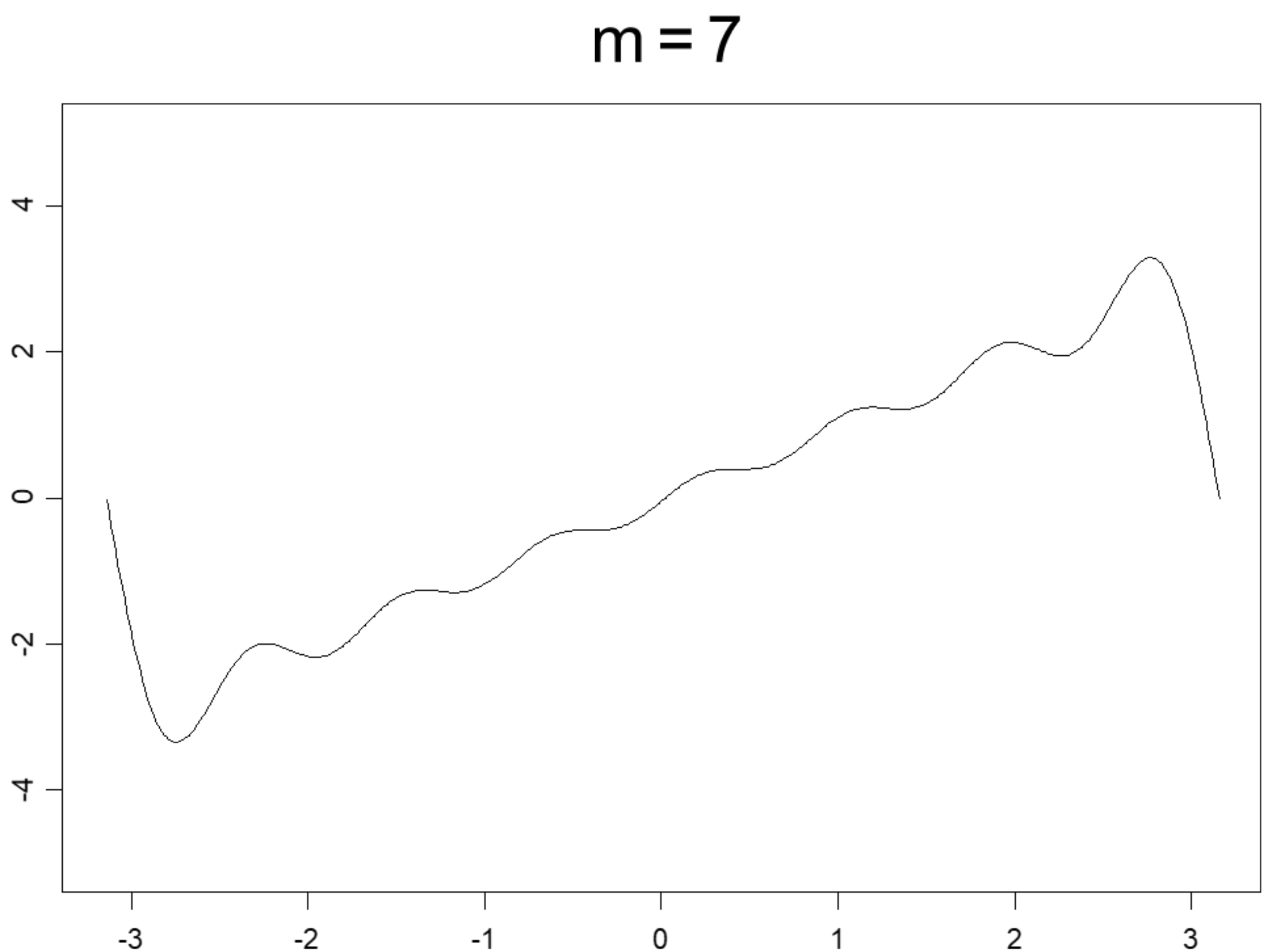
$$x \qquad 2 \;{}_{n=1}^{4} \frac{(\;\;1)^{n}}{n}\sin(nx), \qquad x \quad (\quad , \quad)$$



$$x \qquad 2 \;{}^{5}_{n=1} \frac{(\quad 1)^{n}}{n}\sin(nx), \qquad x \quad (\quad , \quad)$$

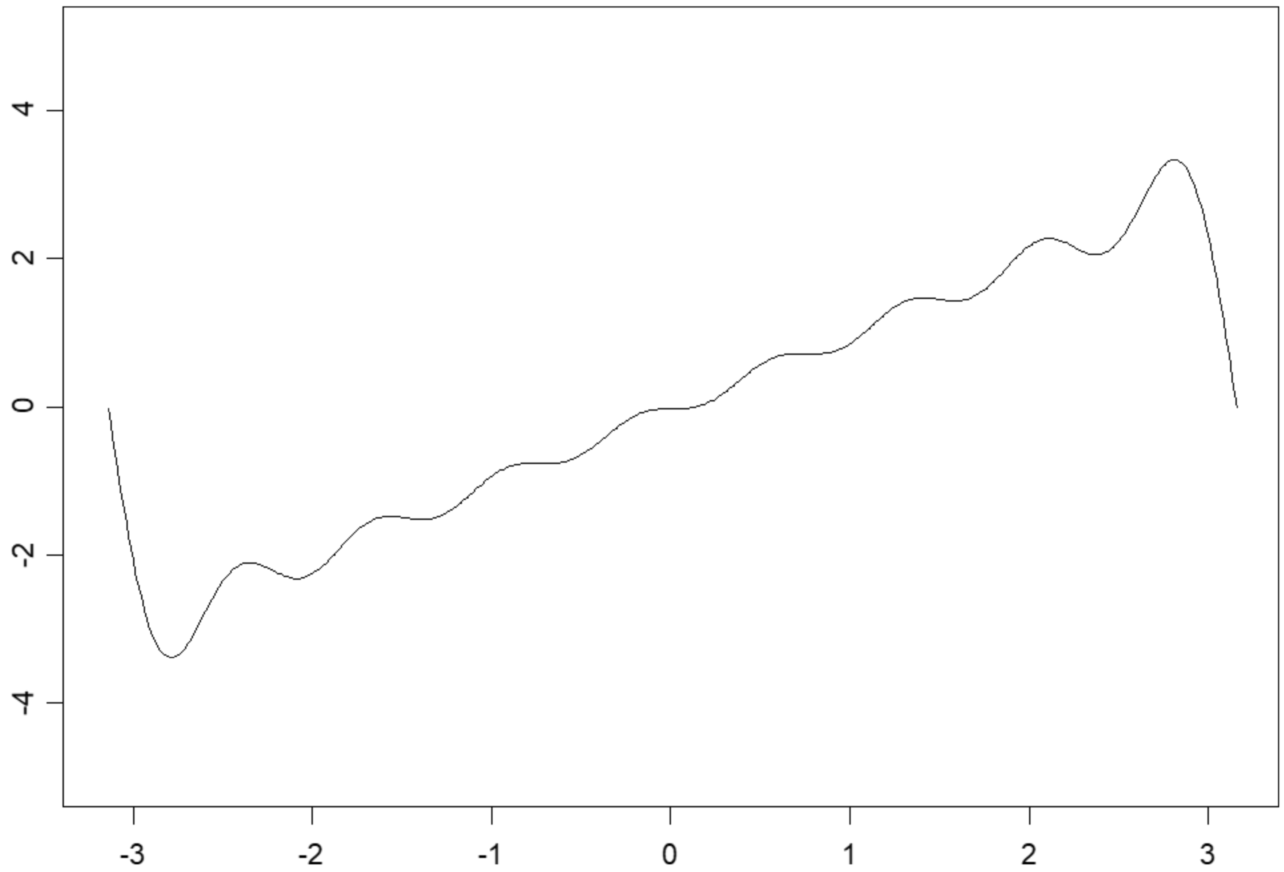


$$x \quad 2 {}_{n=1}^{6} \frac{(\quad 1)^{n}}{n} \sin(nx), \qquad x \quad (\quad , \quad)$$



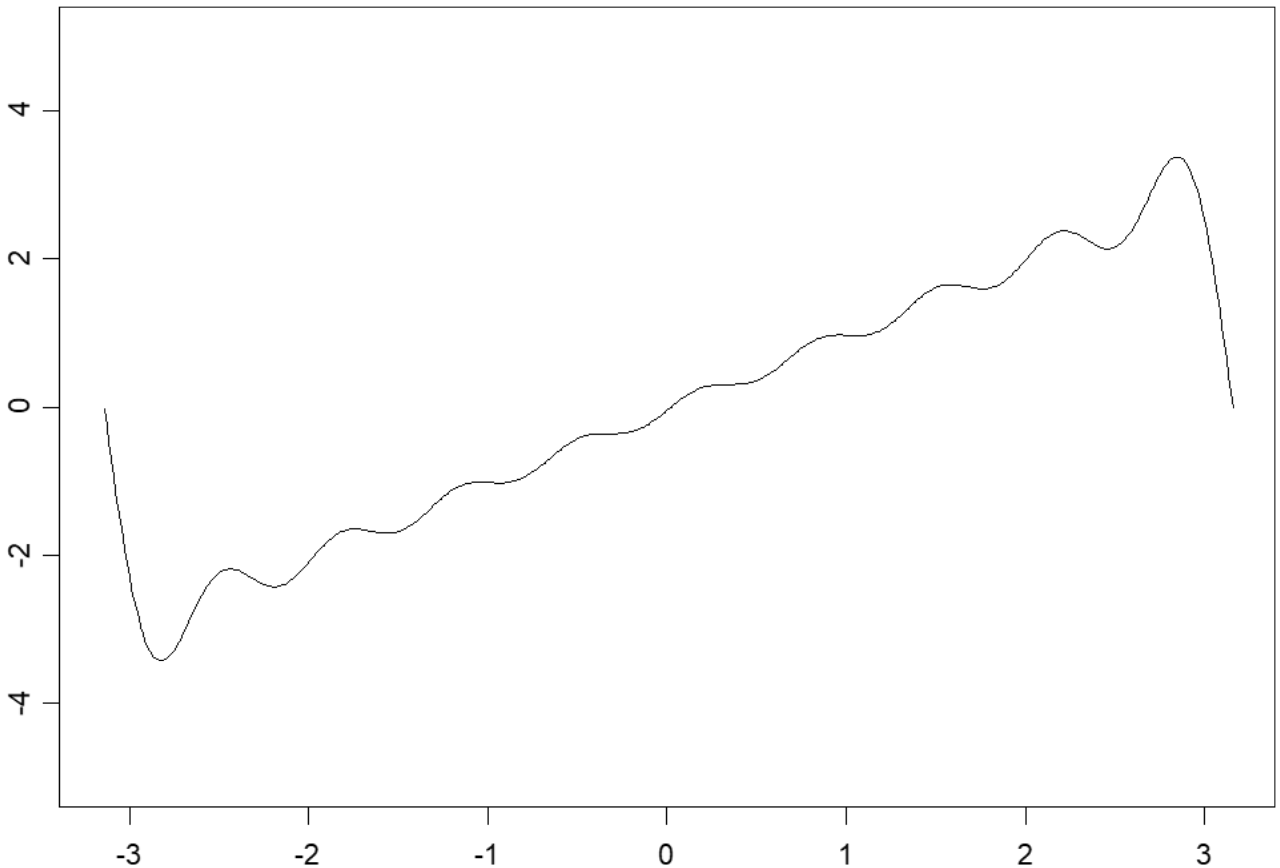
$$x \approx -2\sum_{n=1}^{7}\frac{(-1)^n}{n}\sin(nx), \quad x \in (-\pi, \pi)$$

m = 8



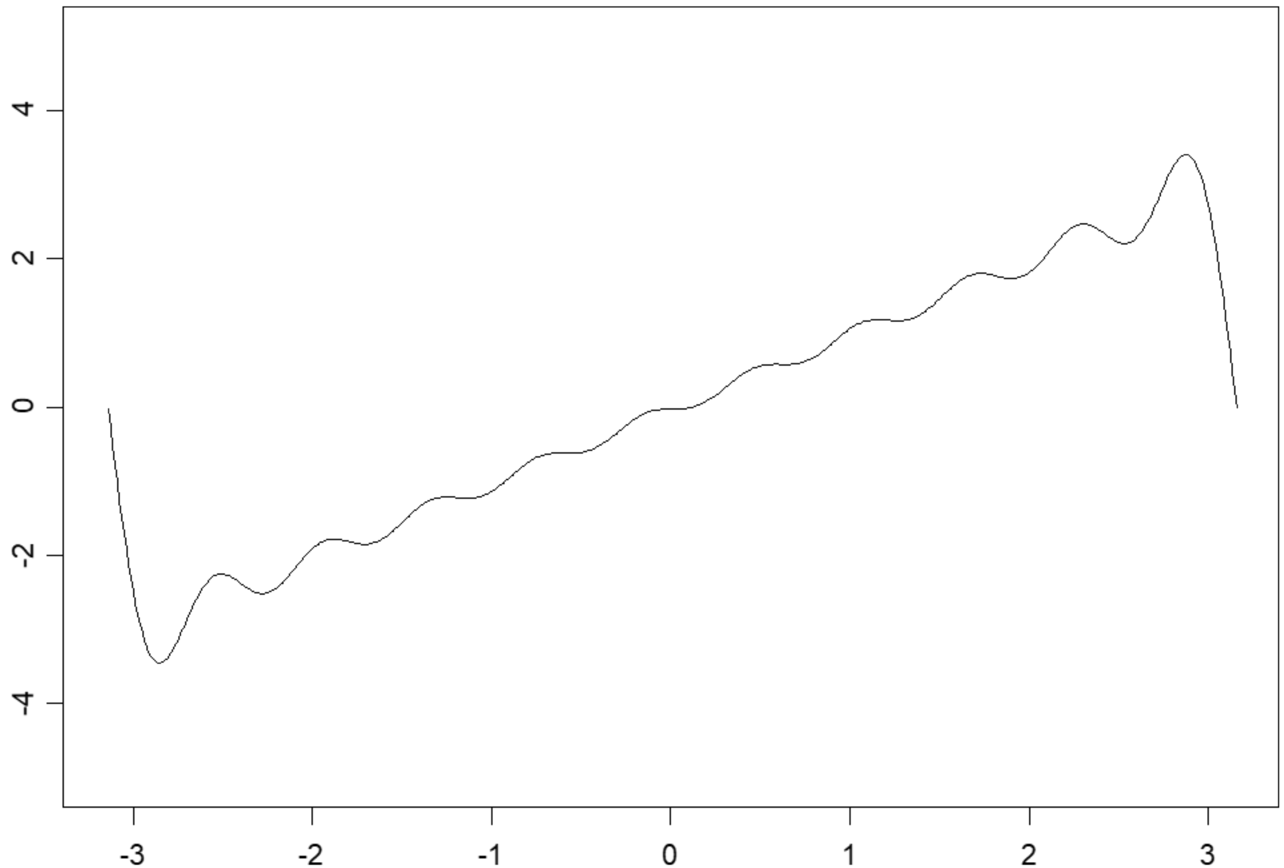
$$x \quad 2 \;\; {}_{n=1}^{\;\;8} \; \frac{(\quad 1)^{n}}{n} \sin(nx), \qquad x \quad (\quad , \quad)$$

m = 9



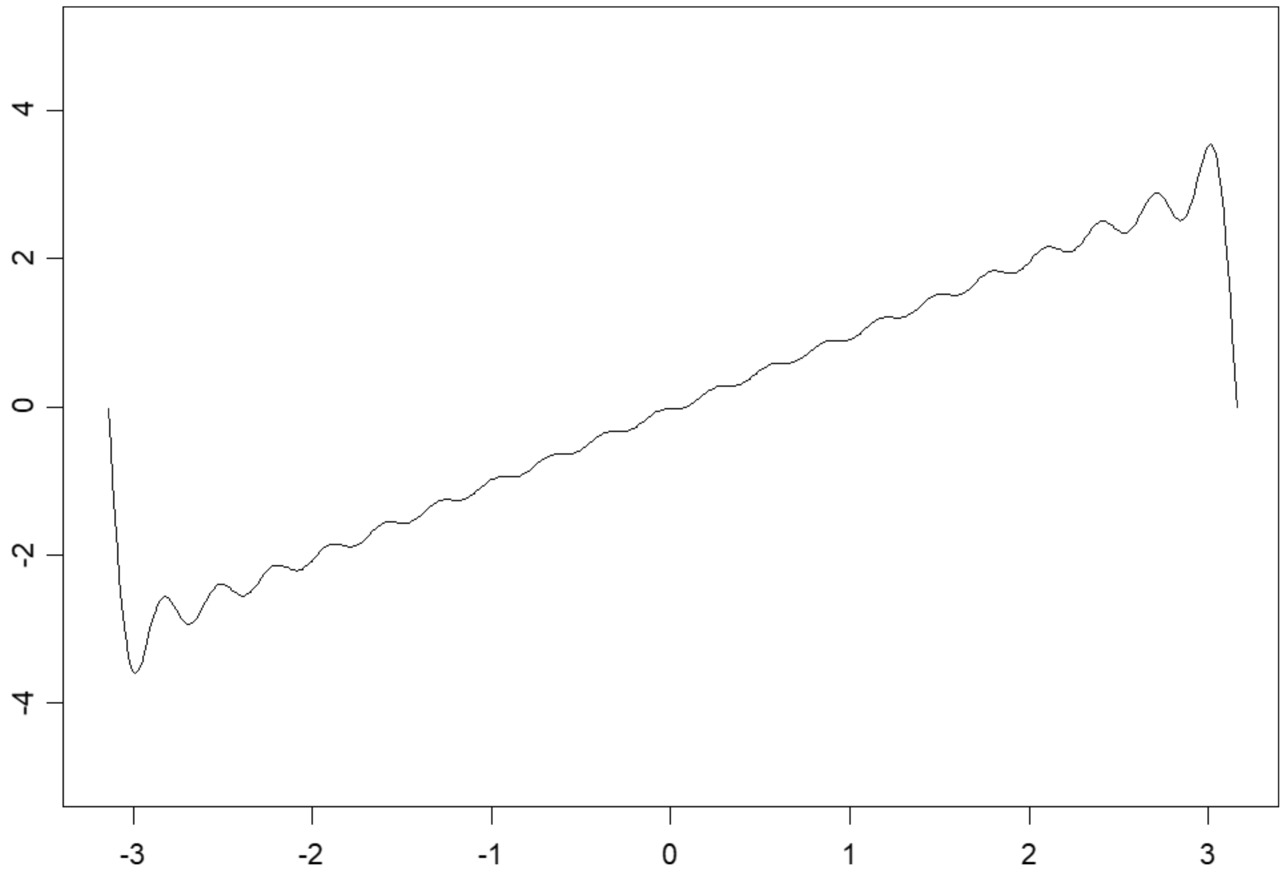
$$x \quad 2 \, {}^{9}_{n=1} \frac{(\quad 1)^{n}}{n} \sin(nx), \qquad x \quad (\quad , \quad)$$

m = 10



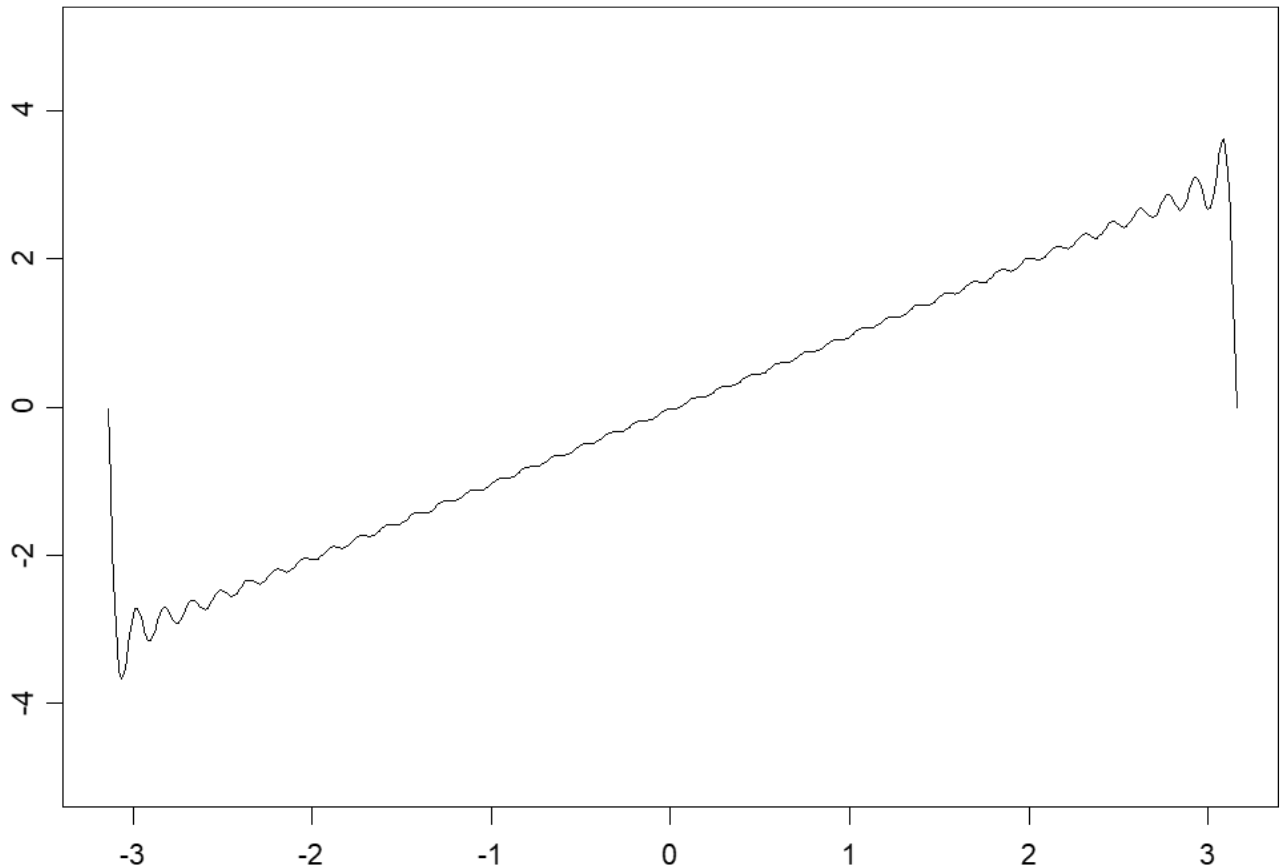
$$x \quad 2 \overset{10}{\underset{n=1}{\ }} \frac{(\quad 1)^{n}}{n} \sin(nx), \qquad x \quad (\quad , \quad)$$

m = 20

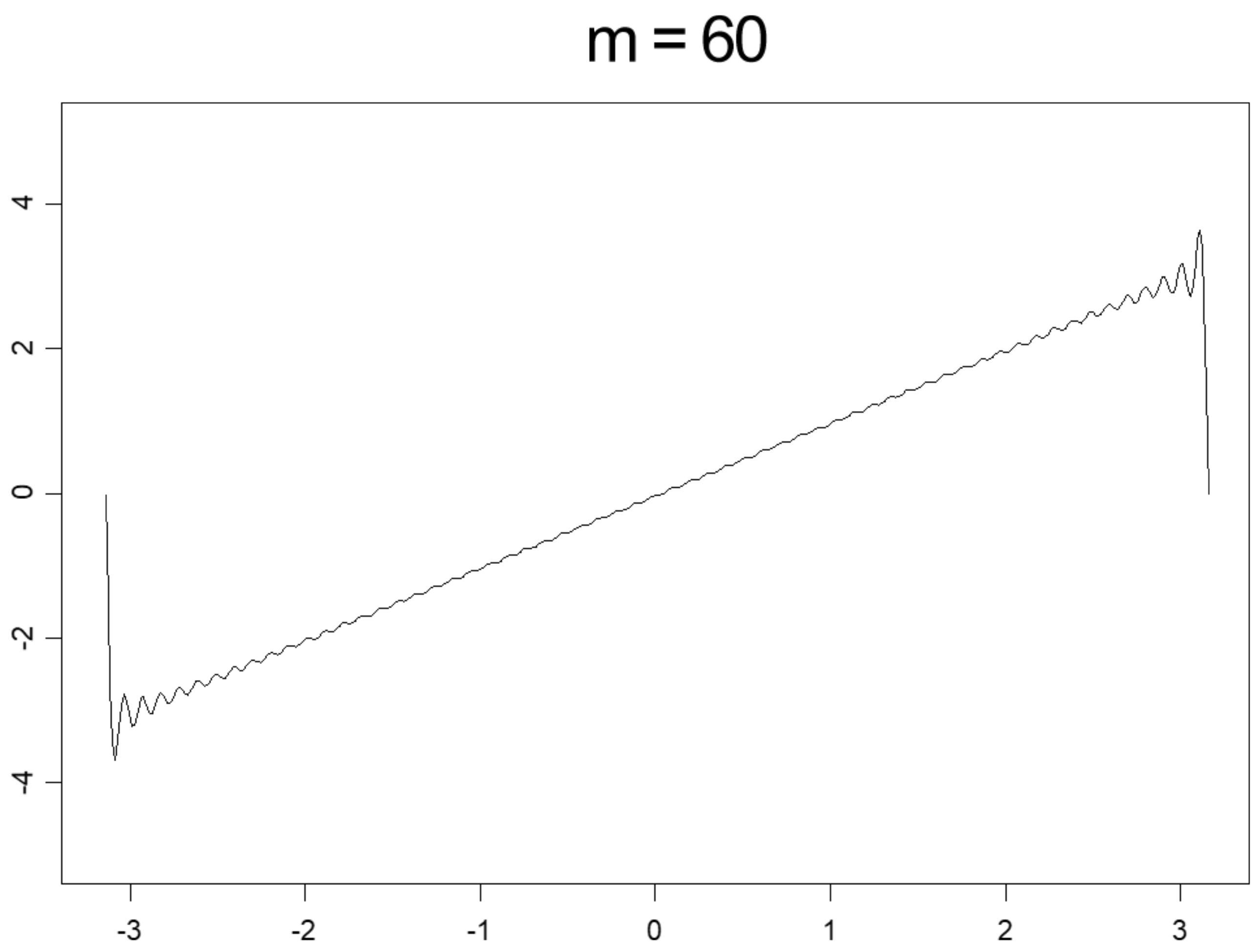


$$x \approx 2\sum_{n=1}^{20}\frac{(-1)^{n}}{n}\sin(nx), \quad x \in (-\pi, \pi)$$

m = 40

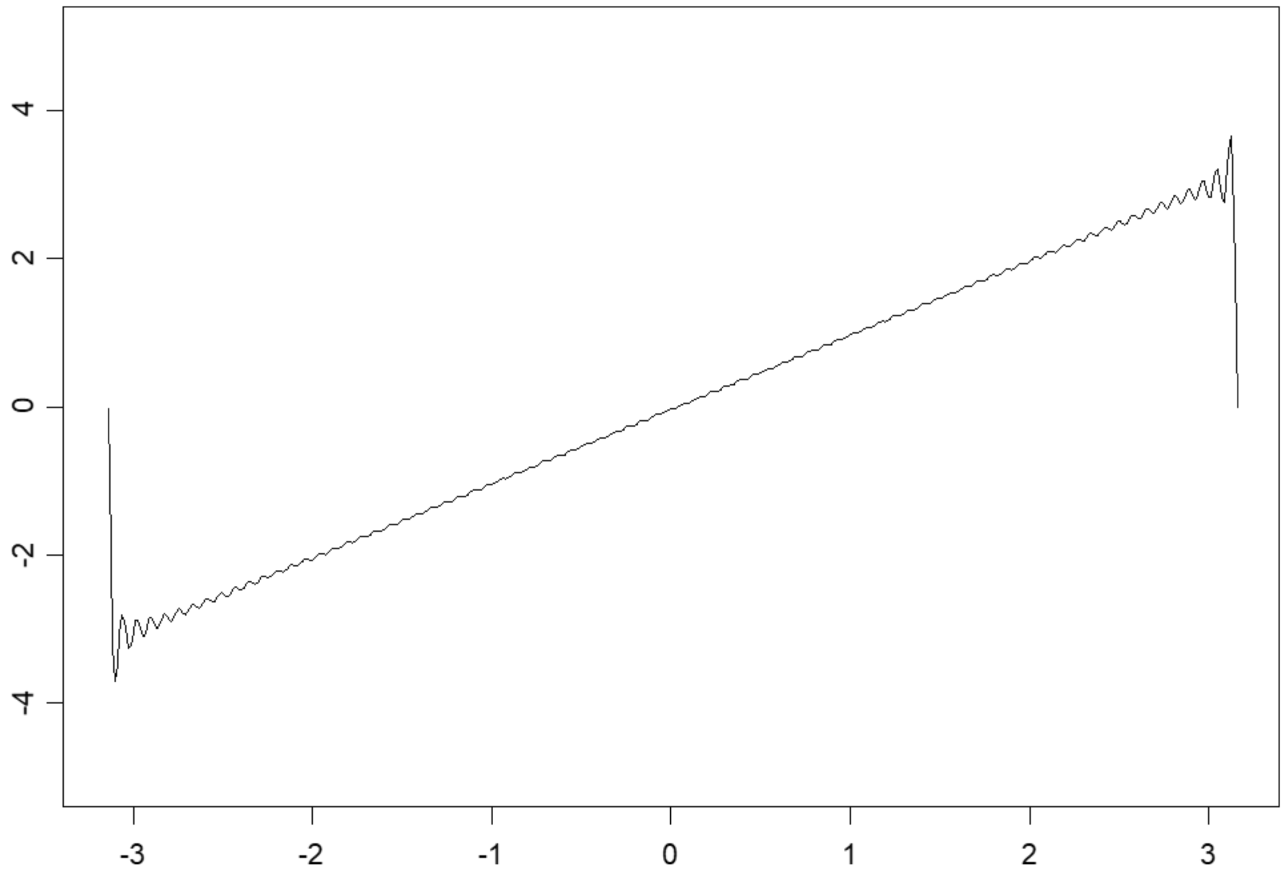


$$x \quad 2 \overset{40}{\underset{n=1}{}} \frac{(\quad 1)^{n}}{n} \sin(nx), \qquad x \quad (\quad , \quad)$$



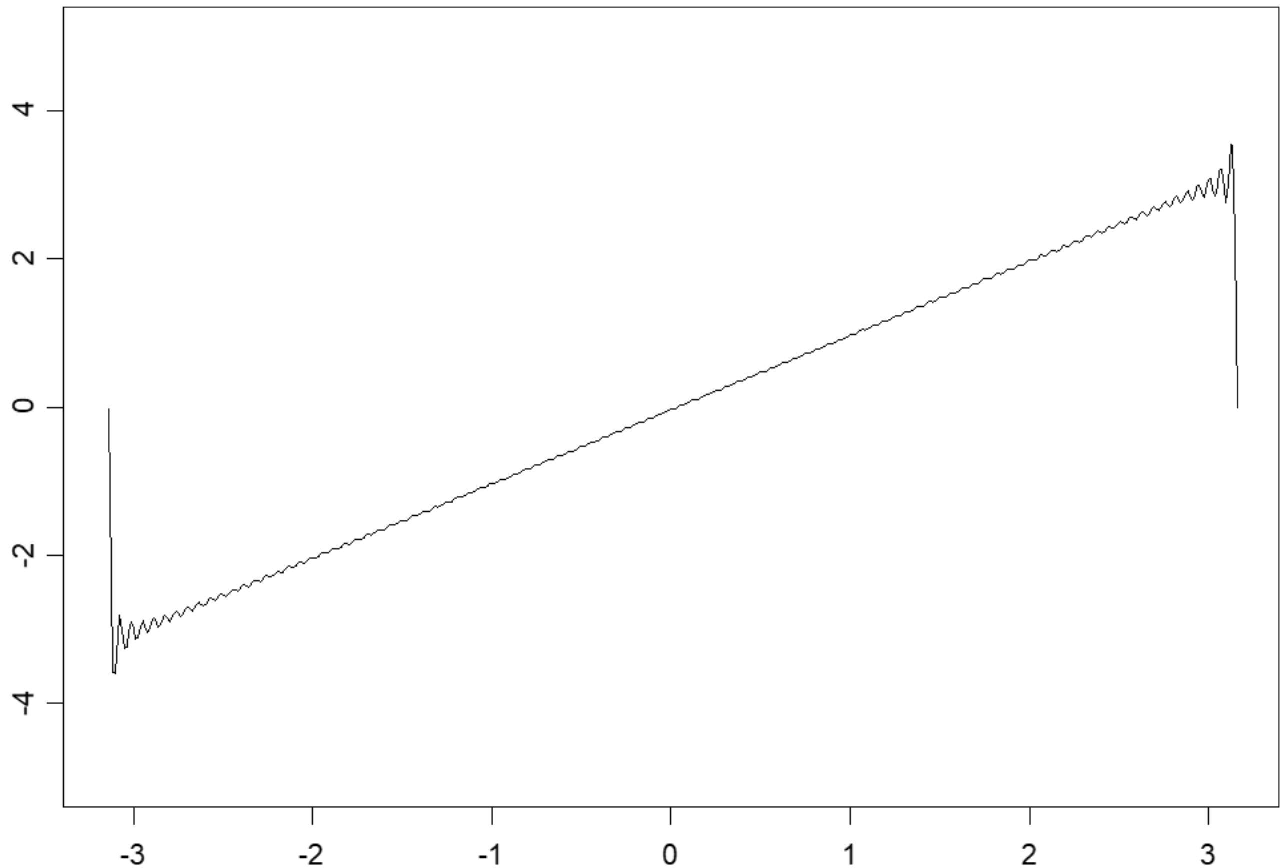
$$x \quad 2\sum_{n=1}^{60}\frac{(\quad 1)^{n}}{n}\sin(nx), \quad x \quad (\quad , \quad)$$

m = 80

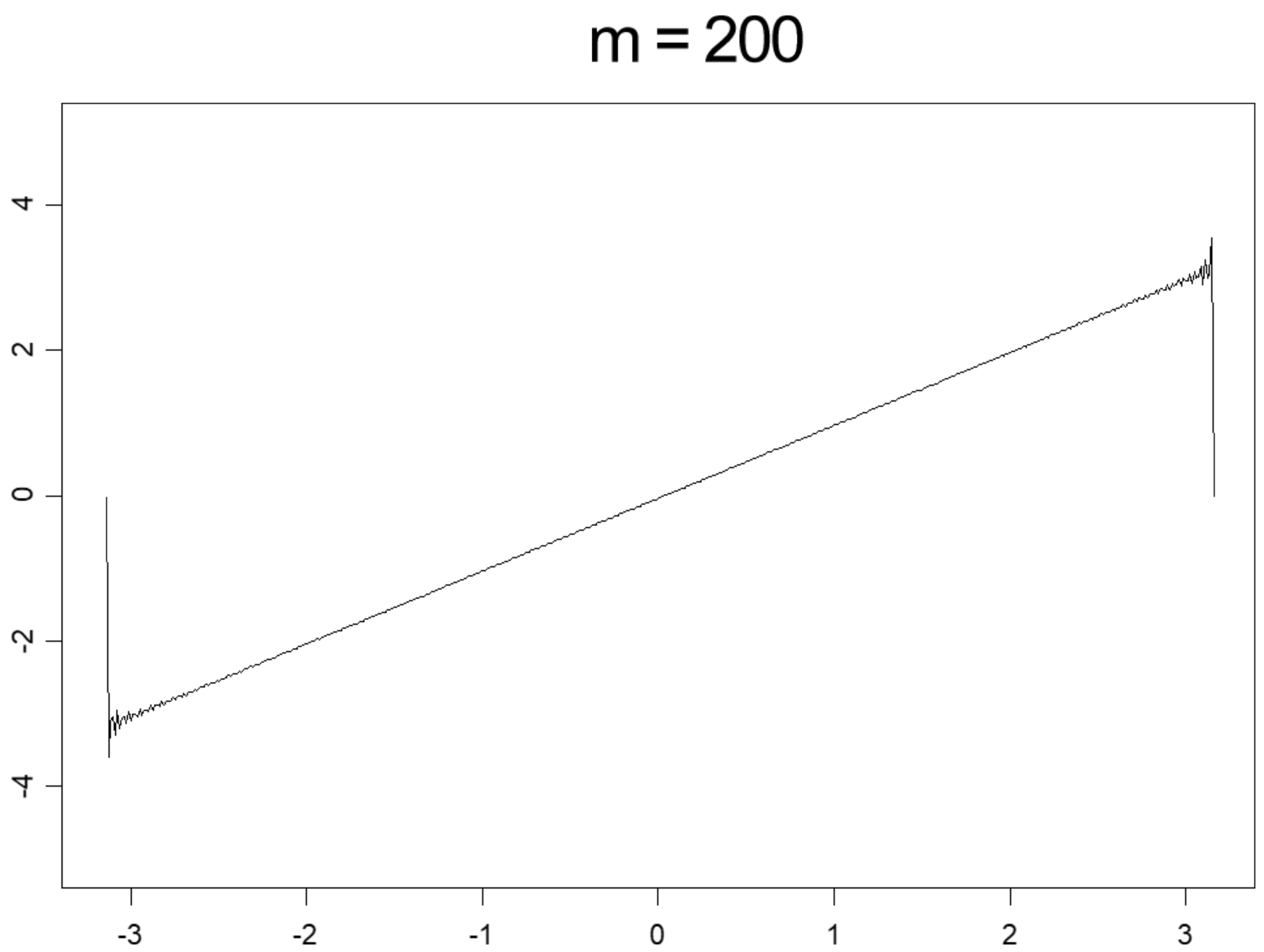


$$x \qquad 2 \sum_{n=1}^{80} \frac{(\quad 1)^n}{n} \sin(nx), \qquad x \quad (\quad , \quad)$$

m = 100

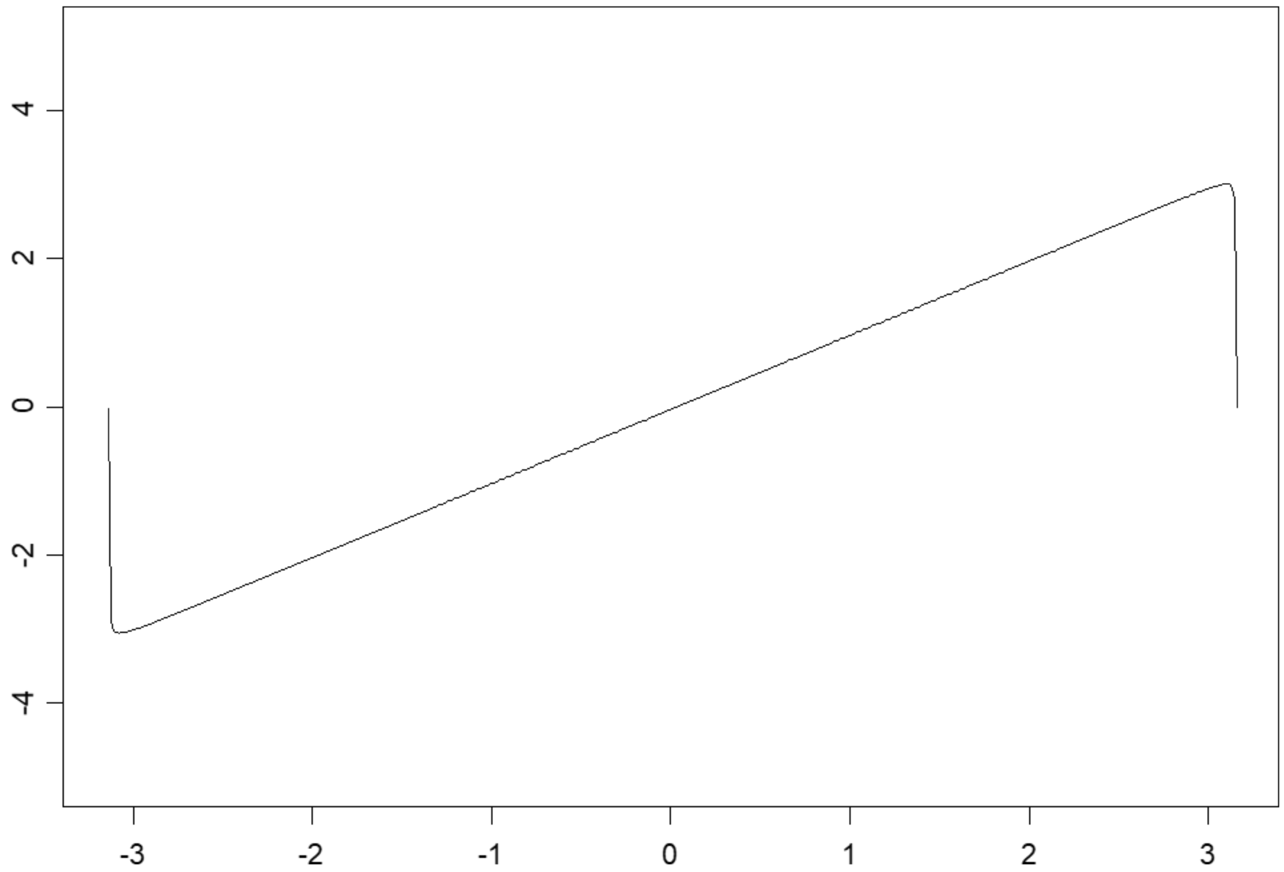


$$x \quad 2 \sum_{n=1}^{100} \frac{(\quad 1)^{n}}{n} \sin(nx), \quad x \quad (\quad , \quad)$$



$$x \approx 2\sum_{n=1}^{200}\frac{(-1)^{n}}{n}\sin(nx), \quad x \in (-\pi, \pi)$$

m = 500

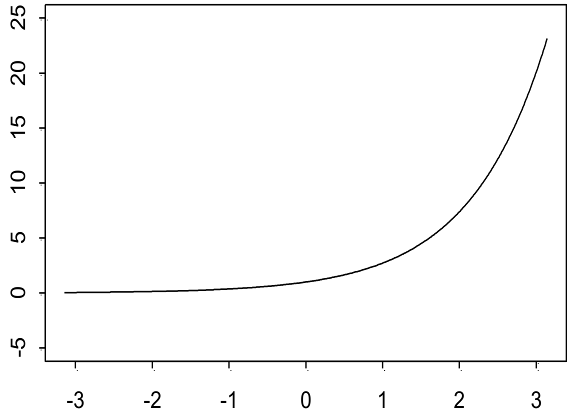


$$x \quad 2 \;{}_{n=1}^{500} \frac{(\quad 1)^{n}}{n} \sin(nx), \qquad x \quad (\quad , \quad)$$

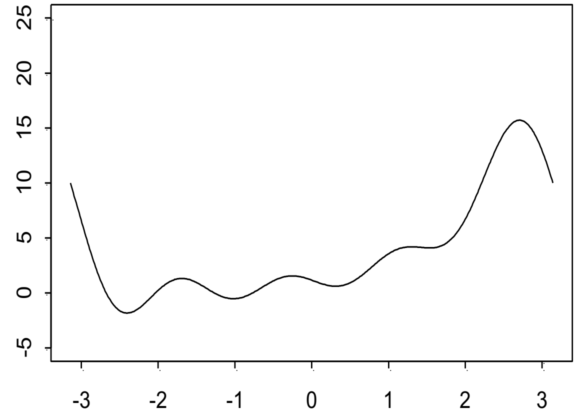
Example: Fourier expansion for $f(x) = e^x$

$$e^x = \frac{\sinh \pi}{\pi} + \frac{2\sinh \pi}{\pi} \sum_{k=1}^{\infty} \frac{(-1)^k}{1+k^2} [\cos kx - k \sin kx]$$

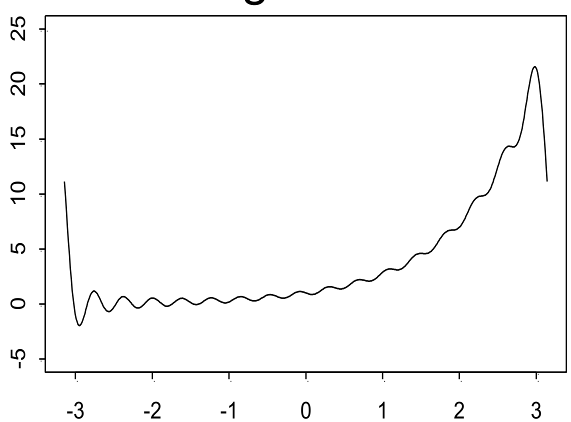
$$f(x) = e^x$$



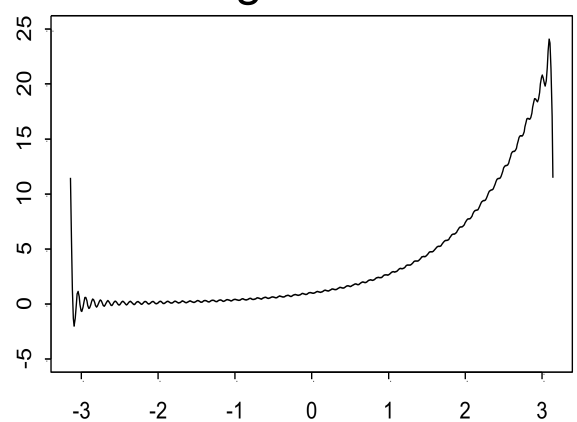
Summing first 4 terms



Summing first 16 terms



Summing first 64 terms



What does this show?

"Non-cyclic type" functions can be expressed as a linear combination of sines and cosines.

DataScience@SMU

Introduction

The Spectral Density

It would be nice to have a tool to identify the frequency content of a time series.

The good news is that, if we have a stationary time series, such a tool exists.

The most important such tool is the spectrum/spectral density.

DataScience@SMU

The Spectral Density:
Properties and Characteristics

Spectrum of a Discrete Stationary Time Series

$$X_t\,,\quad t = 0, \pm 1, \pm 2, \ldots$$

Spectrum

$$P(f) = \sigma_X^2 + 2\sum_{k=1}^{\infty} \gamma_k cos2\pi fk$$

Divide through by σ_X^2

Spectral Density

$$S(f) = 1 + 2\sum_{k=1}^{\infty} \rho_k cos2\pi fk$$

We will focus on Spectral Density

Properties of the Spectral Density

1. $S(f) \geq 0$
2. $S(f) = S(-f)$
3. $S(f) = 1 + 2\sum_{k=1}^{\infty} \rho_k cos2\pi f k$
4. $\int_{-.5}^{.5} S(f) e^{2\pi i f k} df = \rho_k$
5. $\int_{-.5}^{.5} S(f)\, df = 1$

- Point 3 shows that if you know the autocorrelations, you can calculate the spectral density
- Point 4 shows that you can calculate the autocorrelations if you know the spectral density

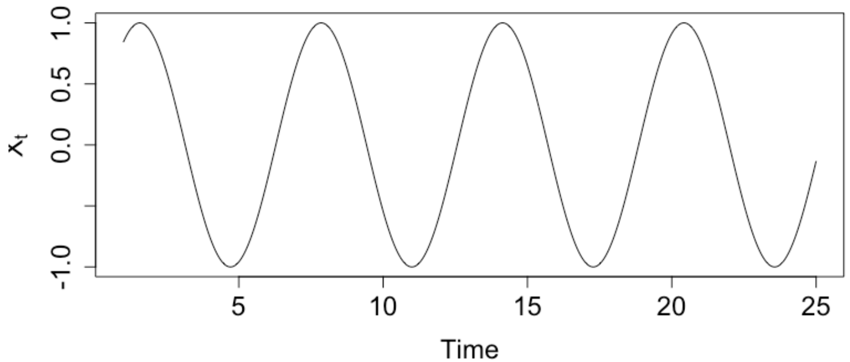
That is: ρ_k and *S(f)* contain equivalent mathematical information (if you know one you can calculate the other.

ρ_k and *S(f)* are called Fourier transform pairs

DataScience@SMU

Examples of the Spectral Density

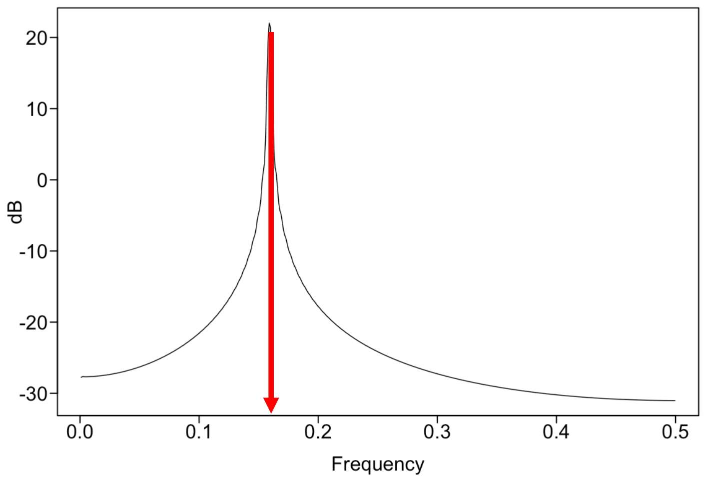
Example: Sin(t)



$$\text{Period} = 2\pi$$

$$\text{Frequency} = \frac{1}{2\pi} = .159$$

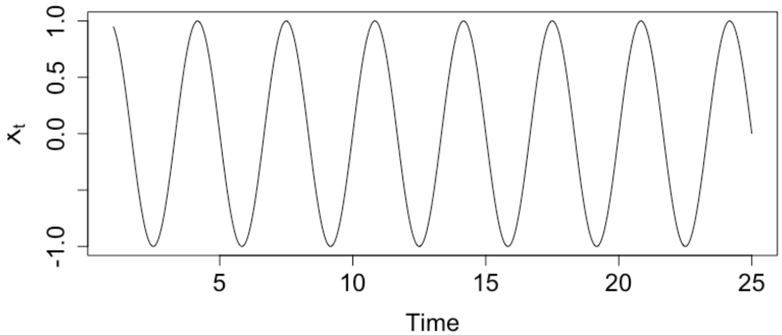
Spectral estimate* for sin(t)



Peak in the spectral density at .159

*More on how the spectrum is estimated later in this unit.

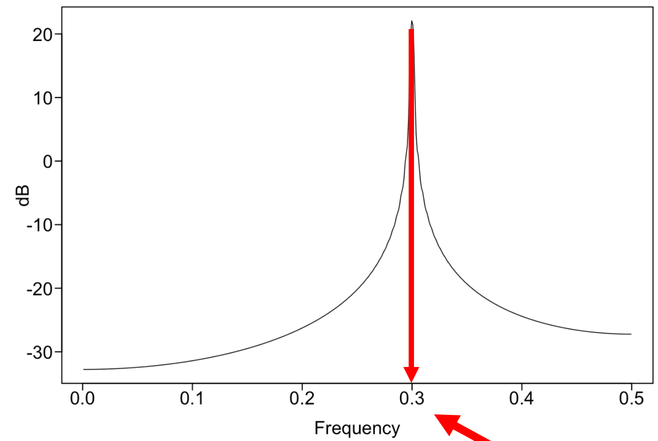
Example: Sin(2π(.3)t)



Period = 3.33

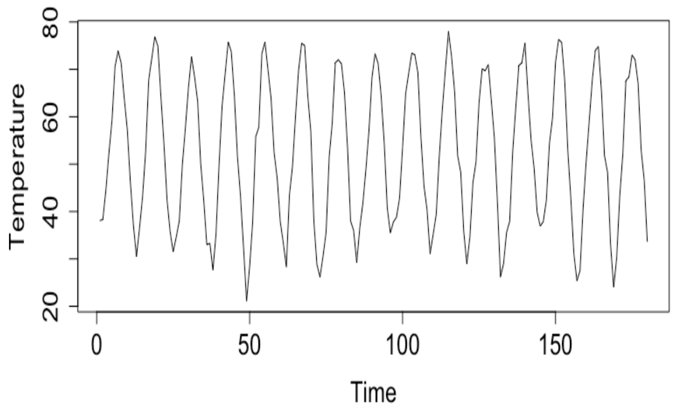
Frequency = .3

Spectral estimate for $\sin(2\pi(.3)t)$



Peak in the spectral density at .3

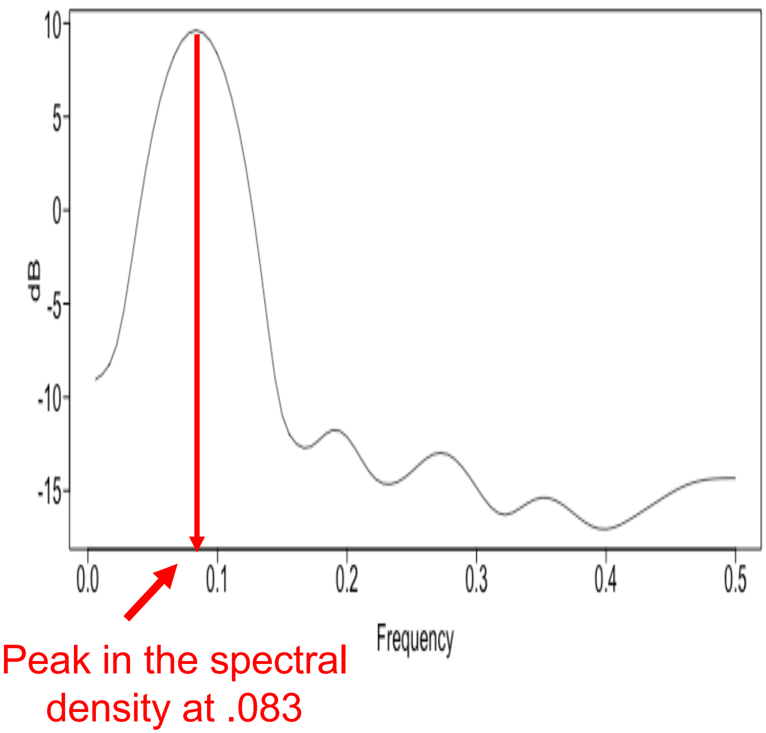
Example: Pennsylvania Temp



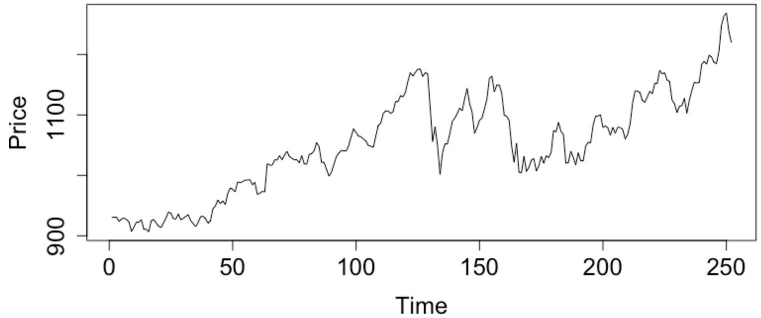
$$\text{Pseudoperiod} = 12 \text{ months}$$

$$\text{Frequency} = .0833$$

Spectral estimate for Pennsylvania temp data



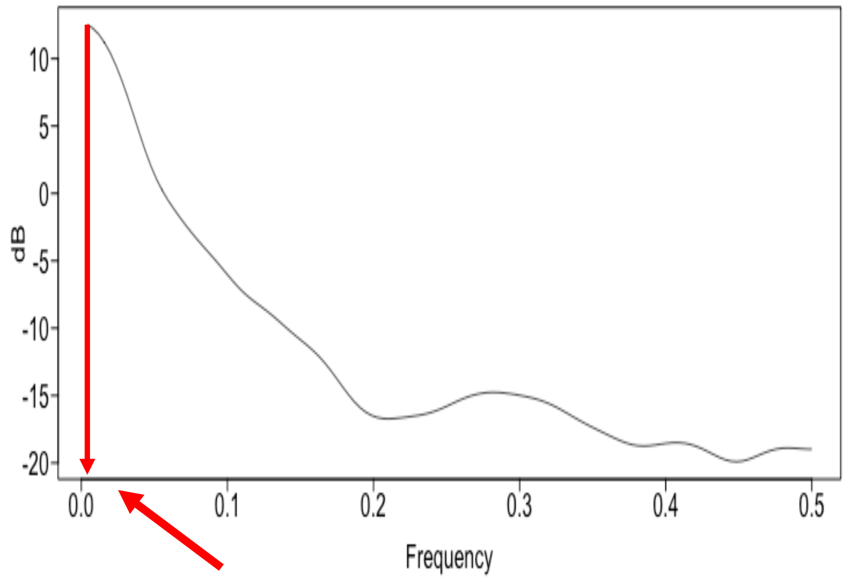
Example: Google Stock Price



$$\text{Period} = \text{Aperiodic}$$

$$\text{Frequency} = NA$$

Spectral estimate for
Google stock price data

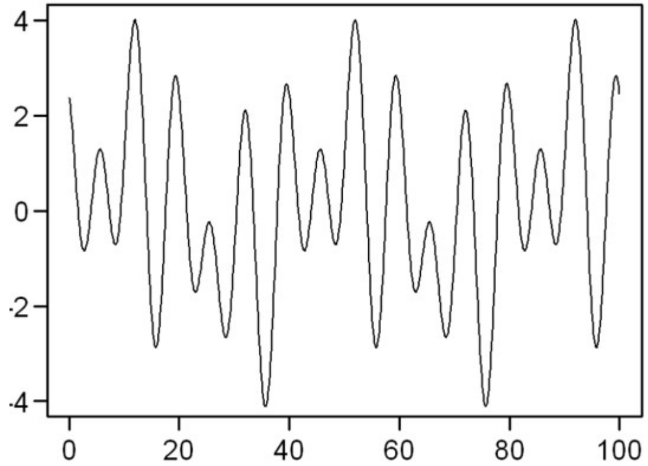


Intuitively, for the length of this data set, the series has not displayed evidence of completing a single period/cycle. Consequently, the shortest the period could be here is 250, making the largest the frequency could be is $\frac{1}{250}$.

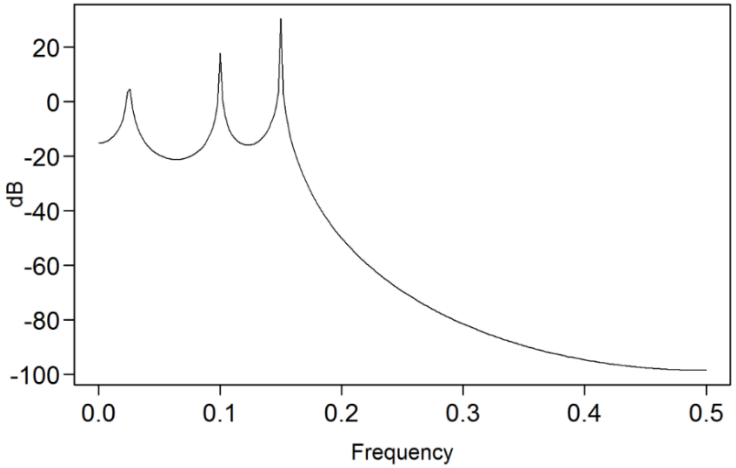
DataScience@SMU

More Examples: Spectral Density

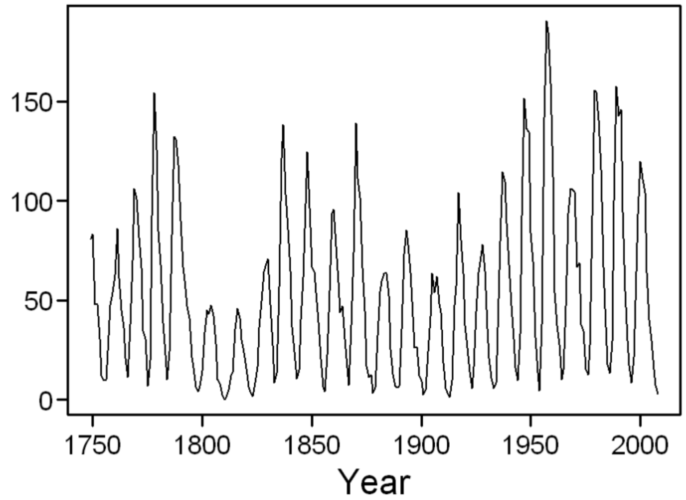
Data set containing sum of 3 sine waves



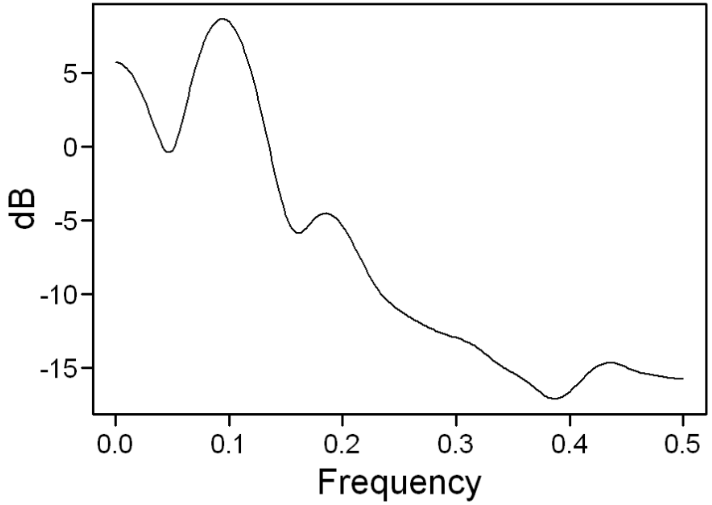
Spectral estimate showing three frequencies



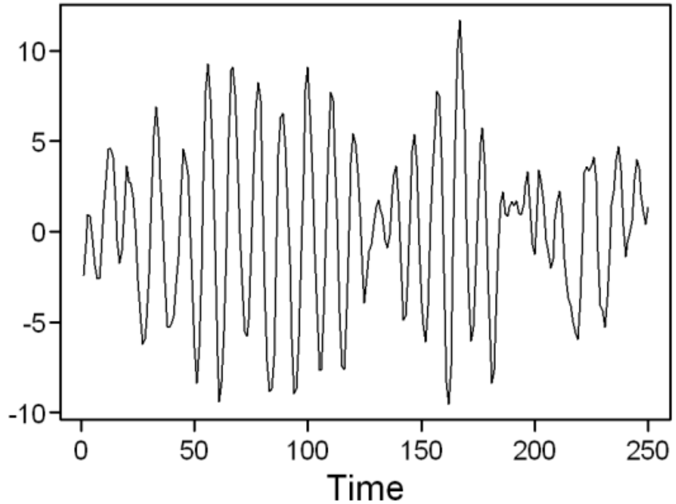
Sunspot data



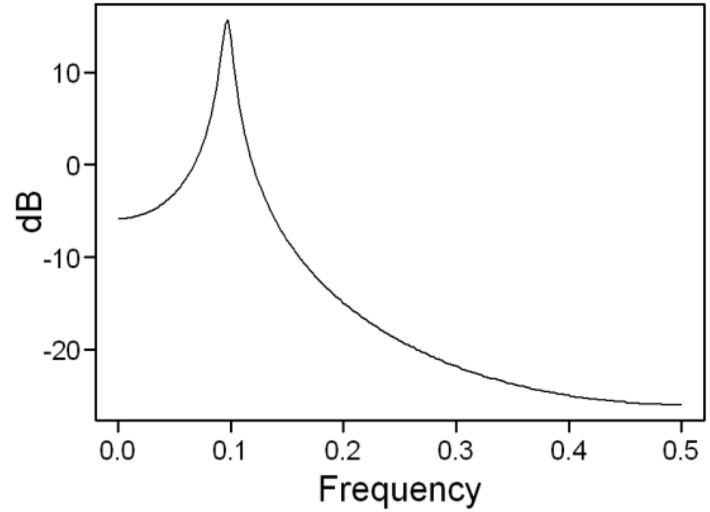
Spectral estimate



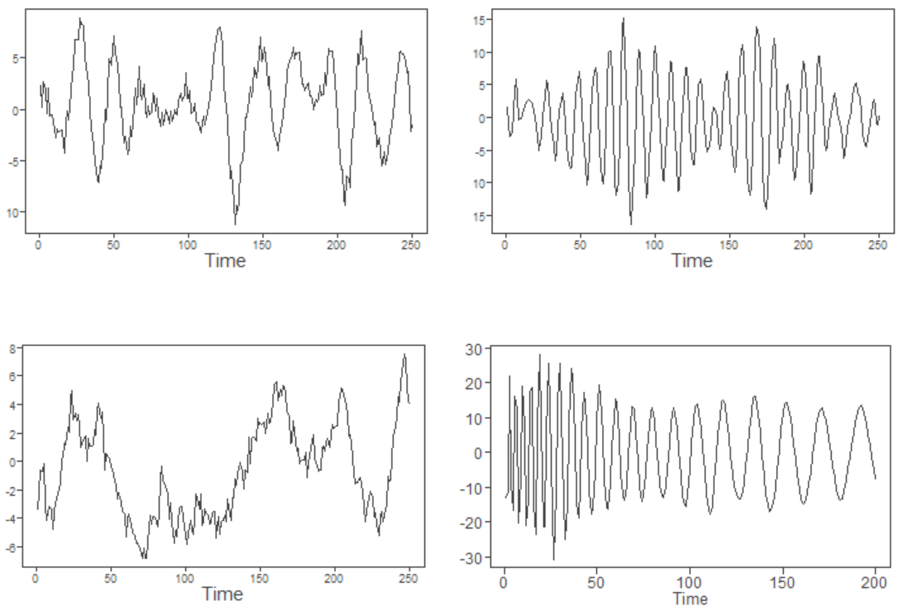
Simulated data

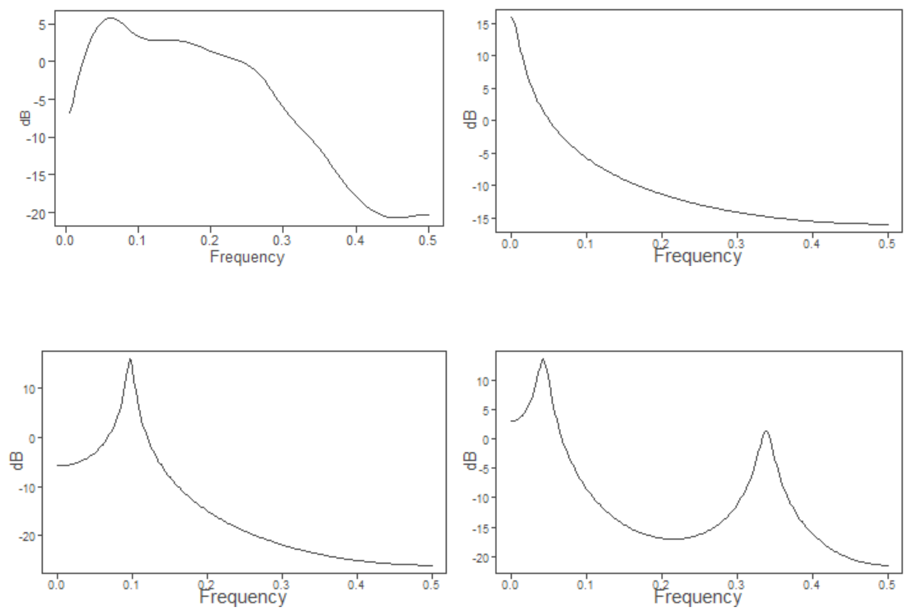


Spectral estimate

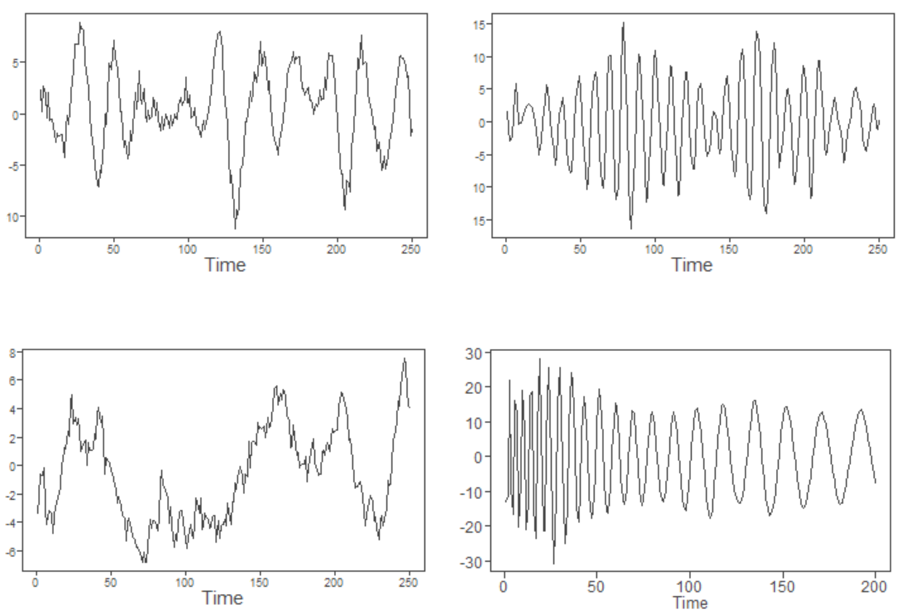


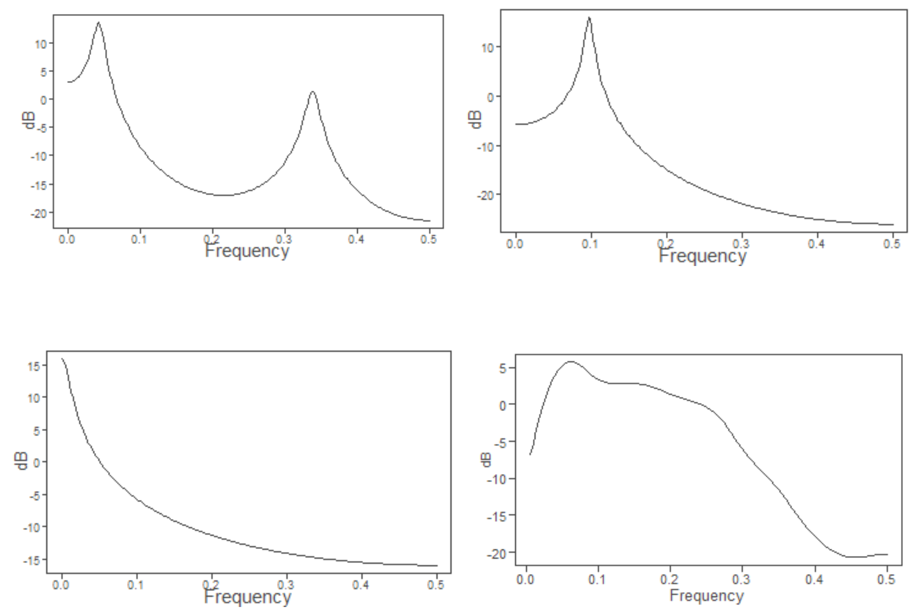
Instructor Concept Check:
4 Time Series Realizations





Instructor Concept Check:
4 Time Series Realizations





DataScience@SMU

The Nyquist Frequency

Nyquist Frequency

4. $$\int_{-.5}^{.5} S(f)e^{2\pi i f k} df = \rho_k$$

5. $$\int_{-.5}^{.5} S(f)\, df = 1$$

Nyquist Frequency

You may be asking yourself: What's the deal with the $-.5$ to $.5$ limits in the integrals in points 4 and 5?

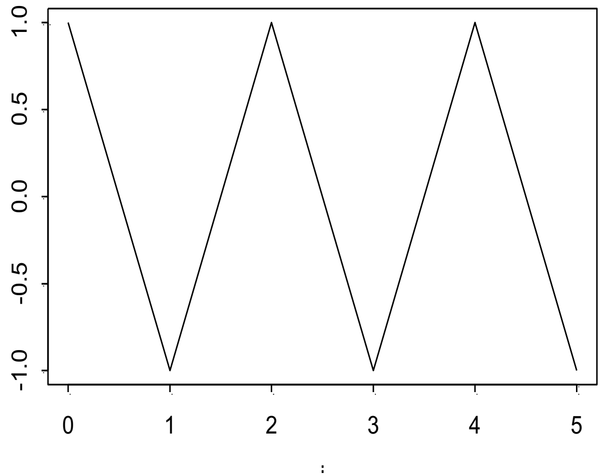
Consider a discrete signal X_k, $k = 0, 1, 2, \ldots$

If X_k is periodic with period j, then j is the smallest value such that $X_k = X_{k+j}$ for all k.

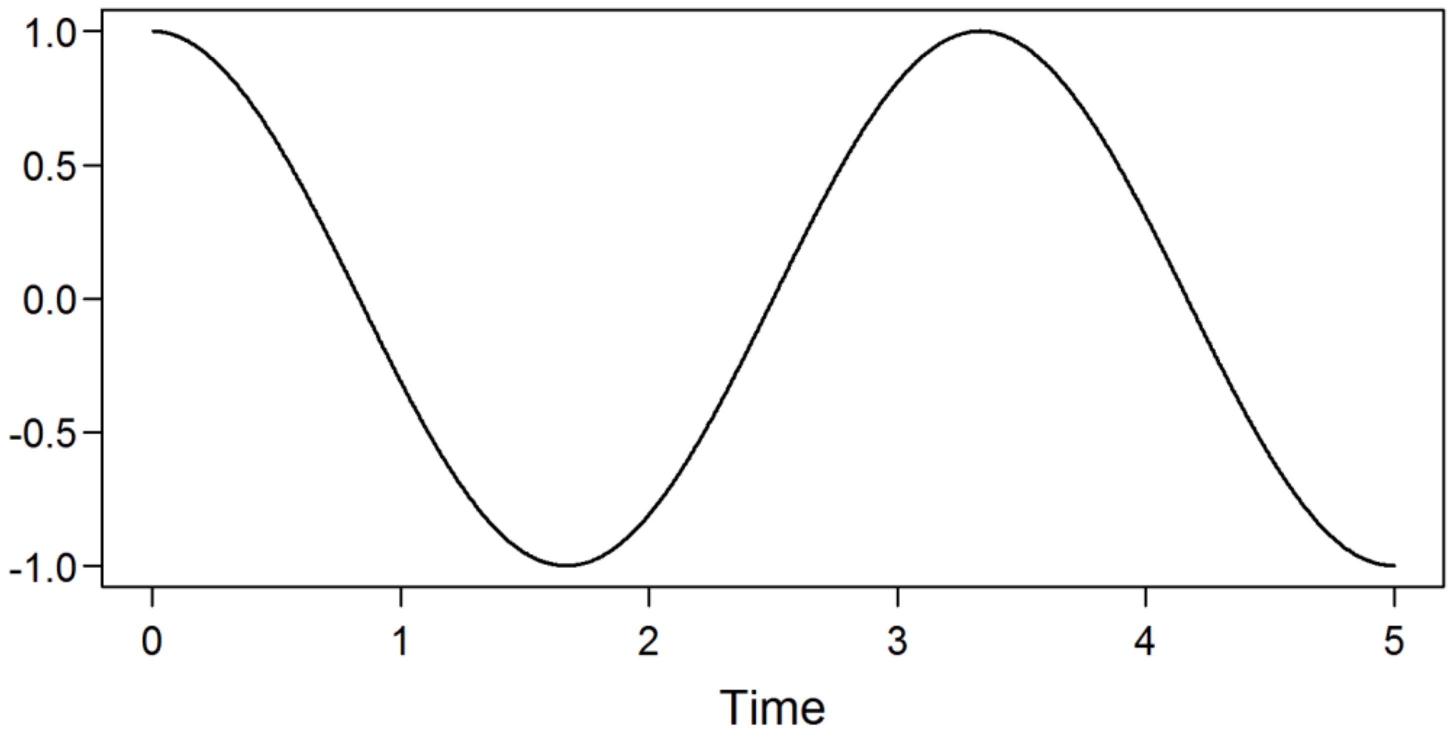
Question: What is the smallest possible period when data are observed at the integers?

For Discrete Data

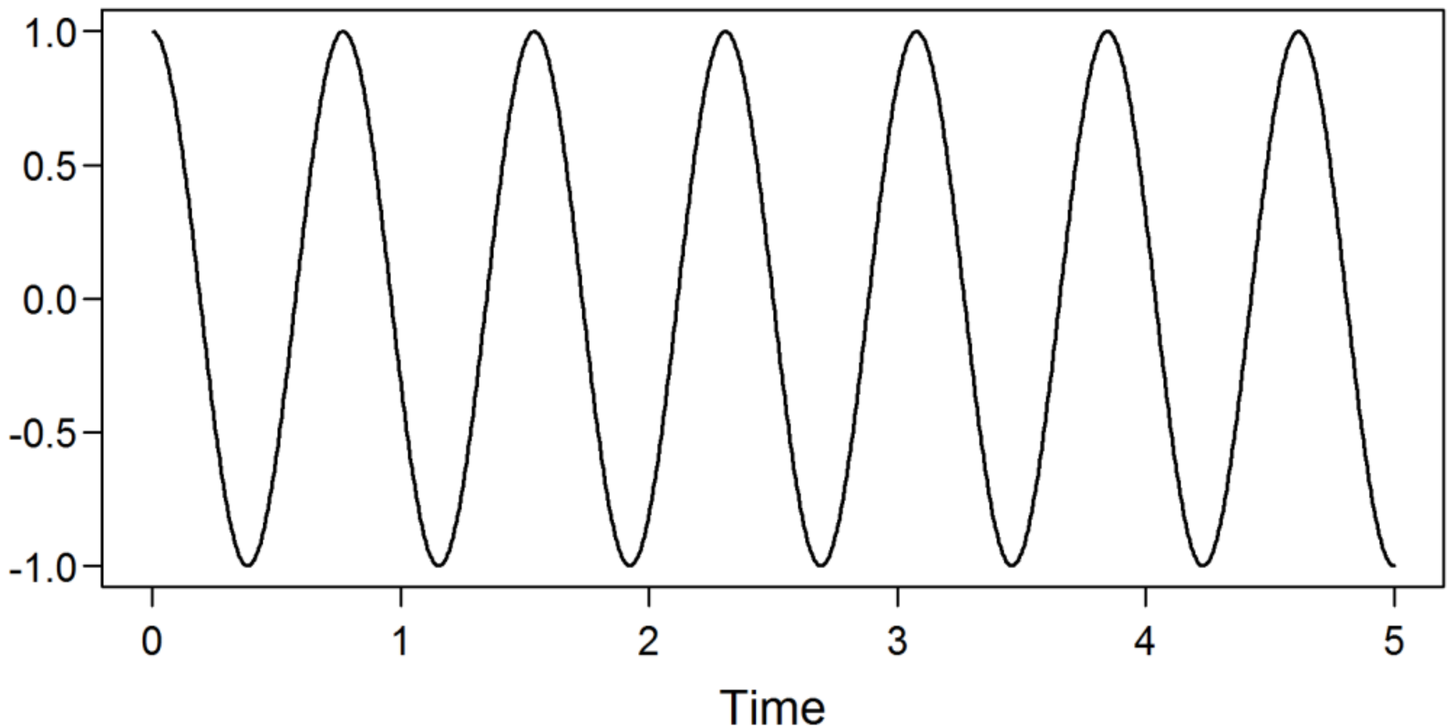
- Shortest observable period $= 2$ units



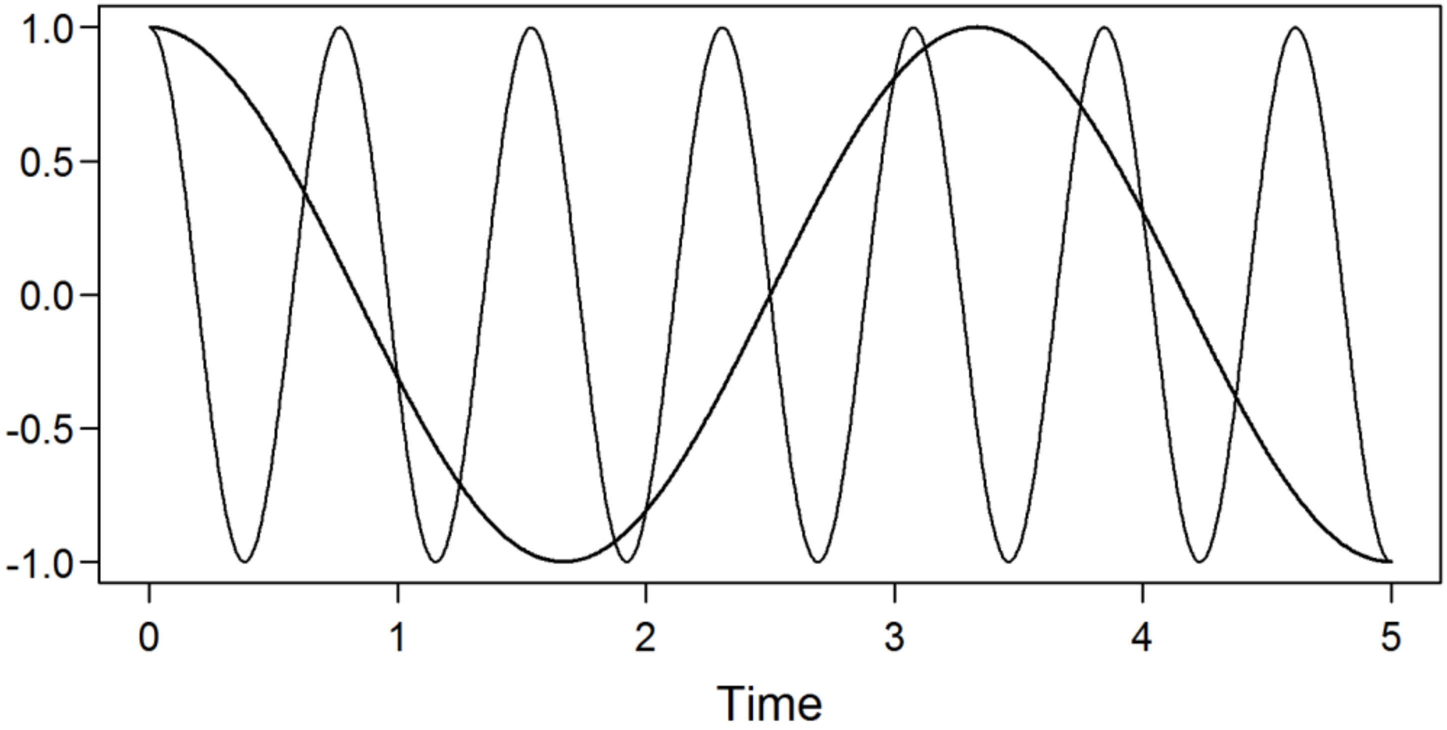
- Highest observable frequency $= .5$

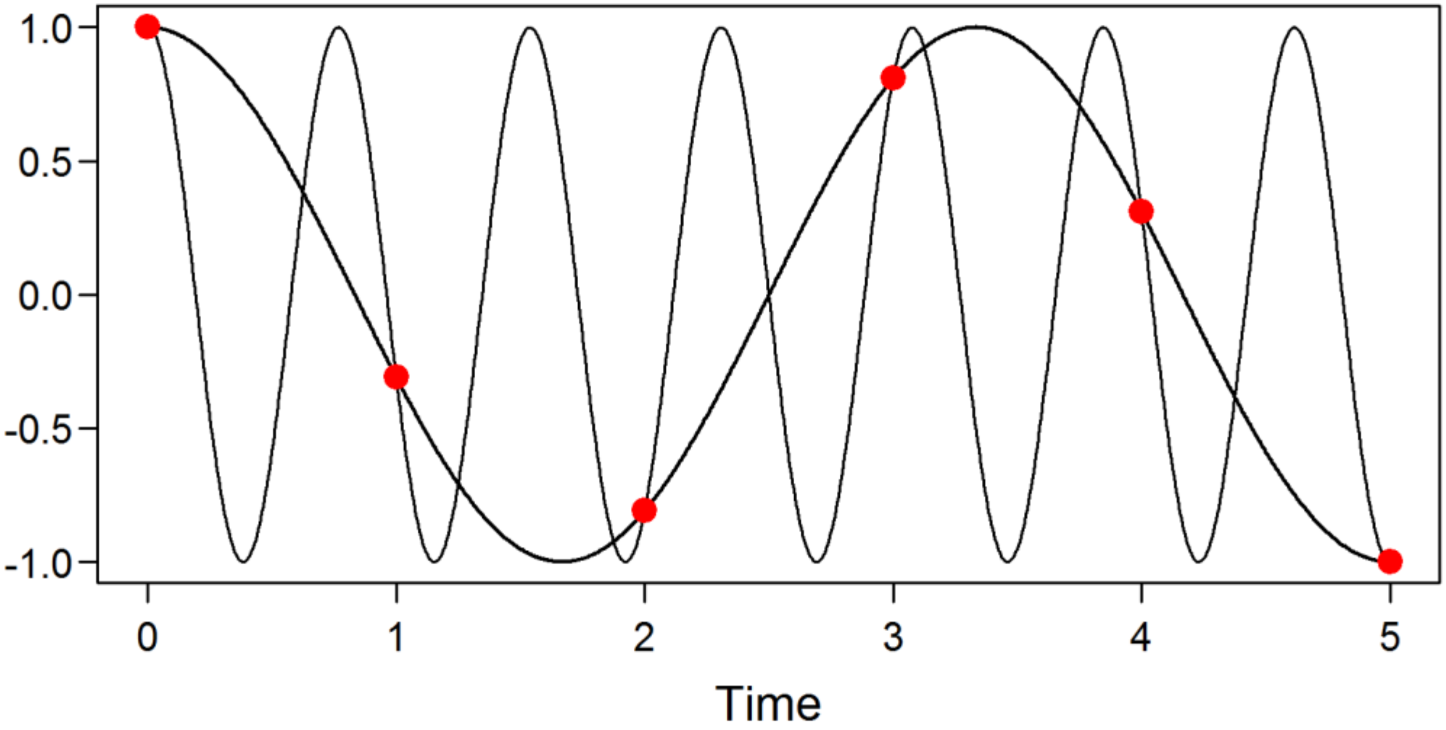


$$\cos(2\pi(.3)t)$$

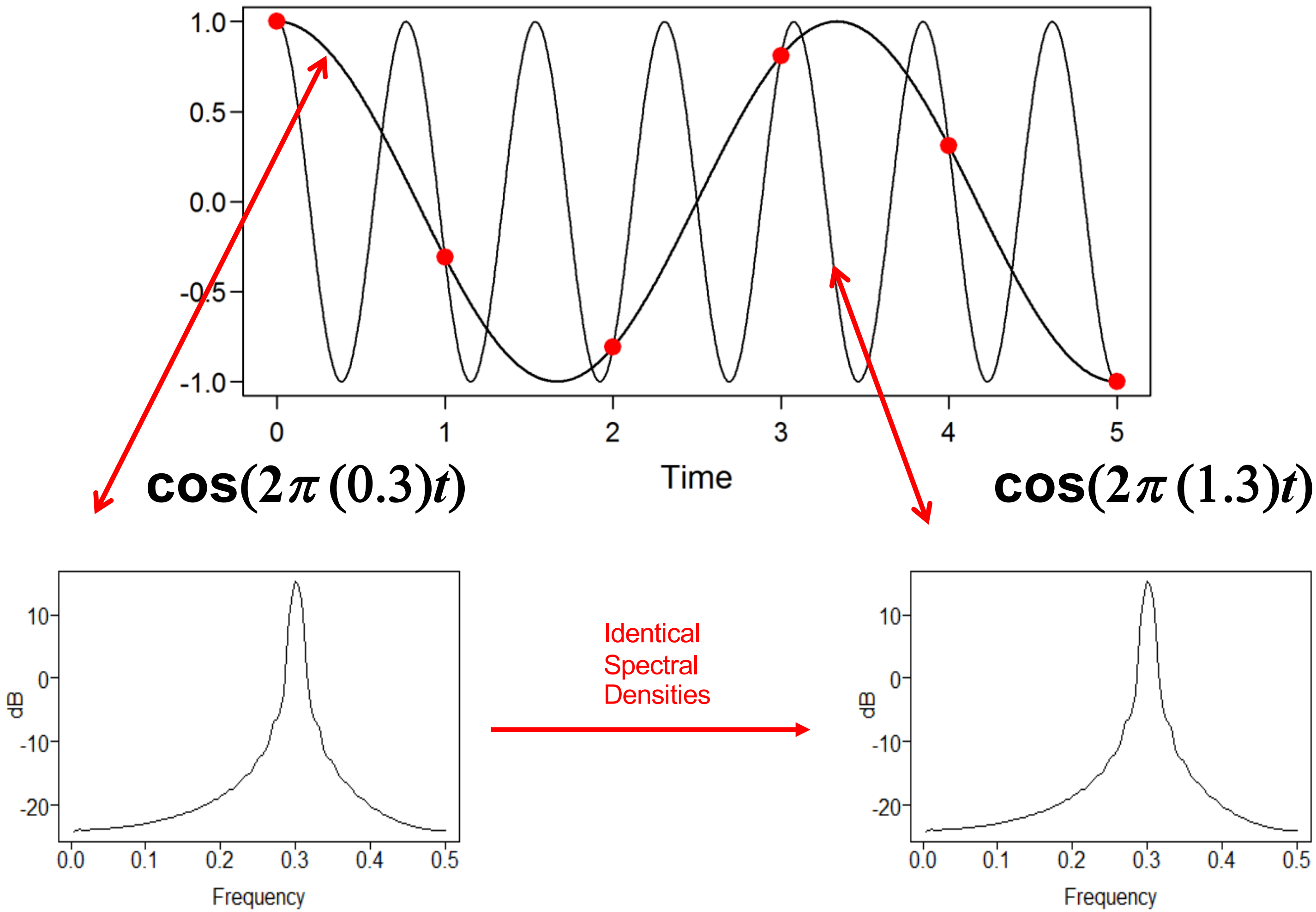


$\cos(2\pi(1.3)t)$


1.0
0.5
0.0
-0.5
-1.0
0
1
2
3
4
5
Time



Note: At $t = 0, \pm 1, \pm 2, \ldots$ these two curves are indistinguishable —*aliasing.*


1.0
0.5
0.0
-0.5
-1.0
0
1
2
3
4
5
Time
$\cos(2\pi(0.3)t)$
$\cos(2\pi(1.3)t)$
10
0
-10
-20
dB
0.0
0.1
0.2
0.3
0.4
0.5
Frequency
Identical
Spectral
Densities

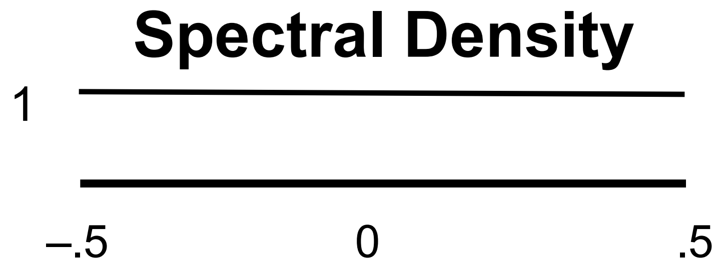
Spectral Density: White Noise

Example: Let a_t be discrete white noise with zero mean and variance σ_a^2.

In this case, $\rho_k = 0 \; for \; all \; k$.

Recall: $$S(f) = 1 + 2\sum_{k=1}^{\infty} \rho_k cos2\pi fk$$

$$Therefore\text{: } S(f) = 1$$



That is, all frequencies are “equally present” in white noise.

DataScience@SMU

Smoothing and Windows

Spectral Estimation in Practice

Note that the quality of $\hat{\ }_k$ decreases as k increases

For example: Given a realization of length n

$$\hat{\rho}_1 = \frac{\frac{1}{n}\sum_{t=1}^{n-1}(X_t - \bar{X})(X_{t+1} - \bar{X})}{\hat{\sigma}_X^2}$$

because there are $n \quad 1$ pairs of points separated by 1 time point e.g. x_1 and x_2, x_2 and x_3, and $x_{n \quad 1}$ and x_n

However,

$$\hat{\rho}_{n-1} = \frac{\frac{1}{n}(X_1 - \bar{X})(X_n - \bar{X})}{\hat{\sigma}_X^2}$$

since x_1 and x_n are the only pair of points separated by $n \quad 1$ time points

Smoothing Methods

-- used to “minimize impact” of $\hat{}_k$ as k increases

$$\hat{S}(f) = \lambda_0 1 + 2\sum_{k=1}^{M} \lambda_k \hat{\rho}_k cos2\pi f k \quad |f| \leq .5$$

Implementation involves choice of M and a window function

Note: A common choice for M is $2\sqrt{n}$

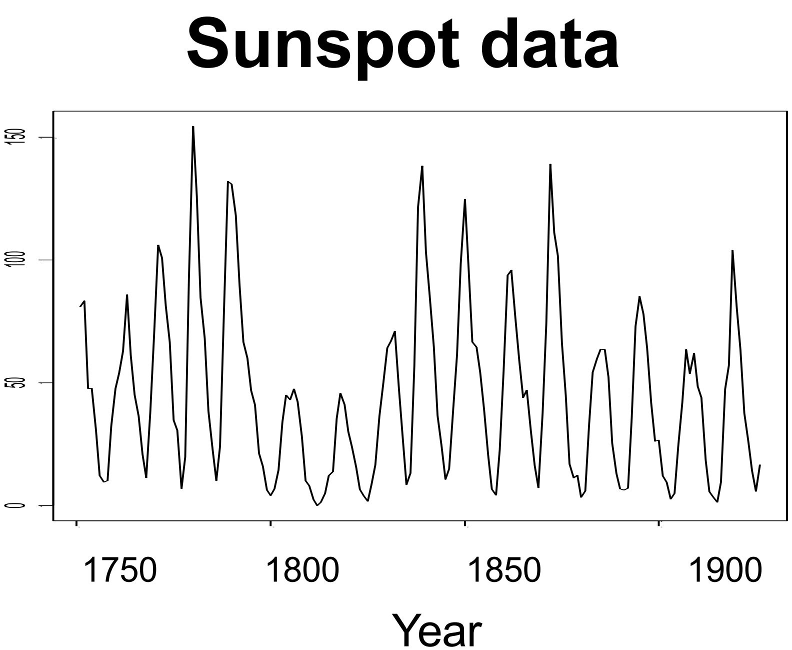
Windows: $\lambda_k' s$

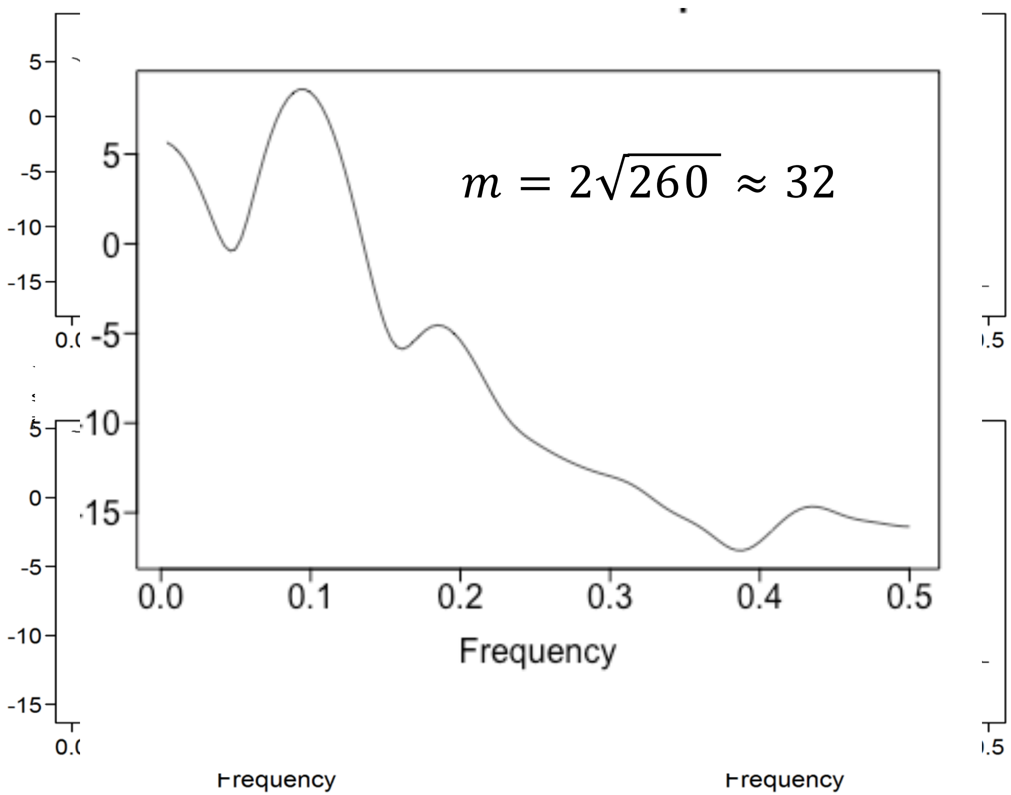
Designed to approach zero as k approaches $n-1$

There are many windows:

`tswge` uses the “Parzen” window

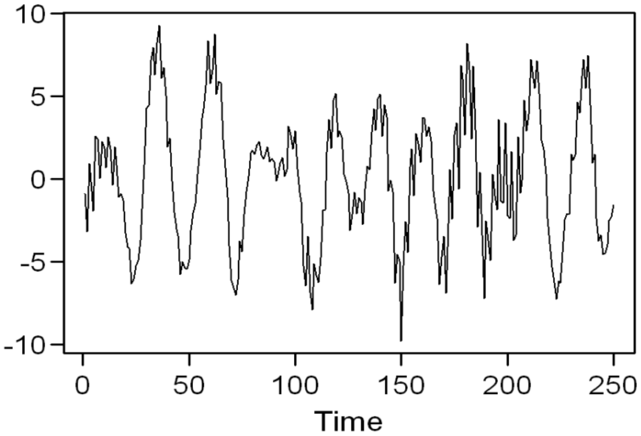
Smoothed Spectral Estimates for Sunspot: Data with Different Truncation Points



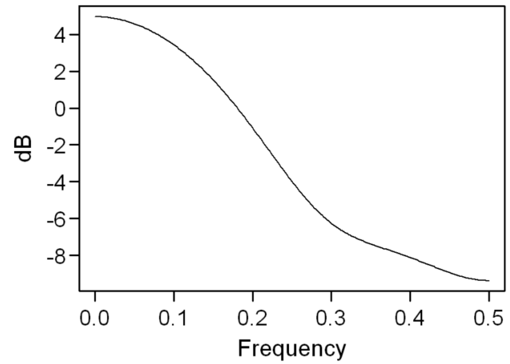


Smoothed Spectral Estimates for Data in Fig. 1.21 with Different Truncation Points

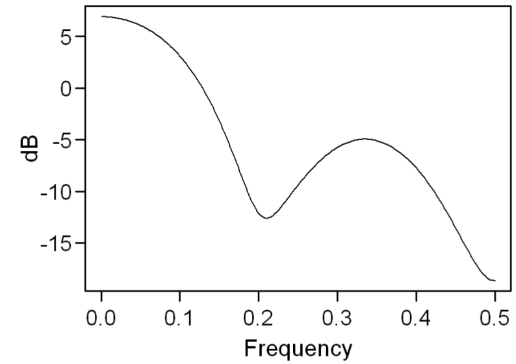
Fig. 1.21



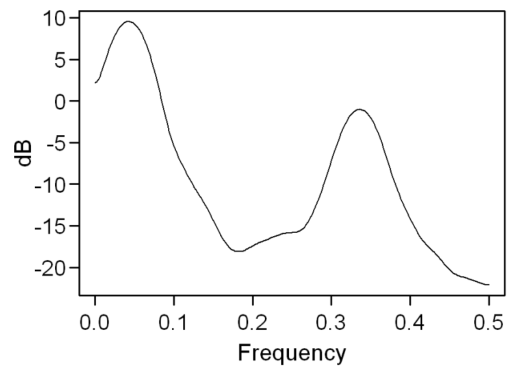
(a) Realization



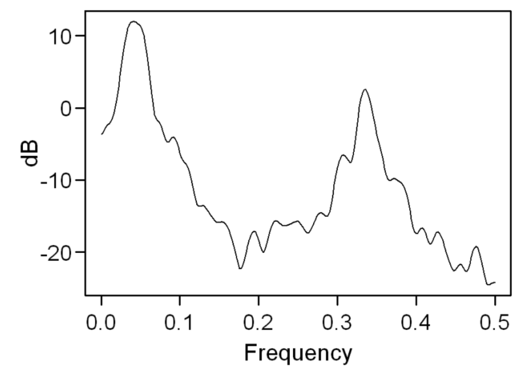
(a) Truncation Point M = 5



(b) Truncation Point M = 10



(c) Truncation Point M = 31



(d) Truncation Point M = 100

$$2\sqrt{250} = 31$$

Other Windows

1. Bartlett (triangular) window

$$\lambda_k = 1 - \left(\frac{k}{M}\right), \quad k = 0, 1, \ldots, M$$
$$= 0, \quad k > M$$

2. Tukey window

$$\lambda_k = \frac{1}{2}\left[1 + \cos\left(\frac{k\pi}{M}\right)\right], \quad k = 0, 1, \ldots, M$$
$$= 0, \quad k > M.$$

3. Parzen window

$$\lambda_k = 1 - 6\left(\frac{k}{M}\right)^2 + 6\left(\frac{|k|}{M}\right)^3, \quad 0 < k \le \frac{M}{2}$$
$$= 2\left[1 - \left(\frac{|k|}{M}\right)\right]^3, \quad M/2 < k \le M$$
$$= 0, \quad k > M.$$

The Parzen window is what we will use in this class.

Fun Fact: Professor Parzen was a personal friend of Dr. Woodward!

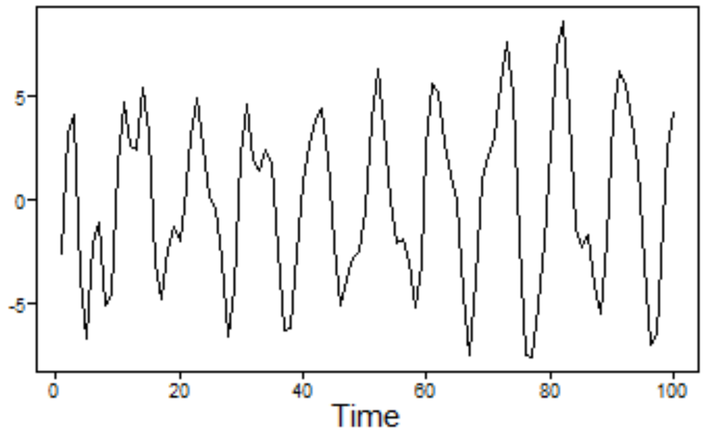
DataScience@SMU

Log of the Spectral Density

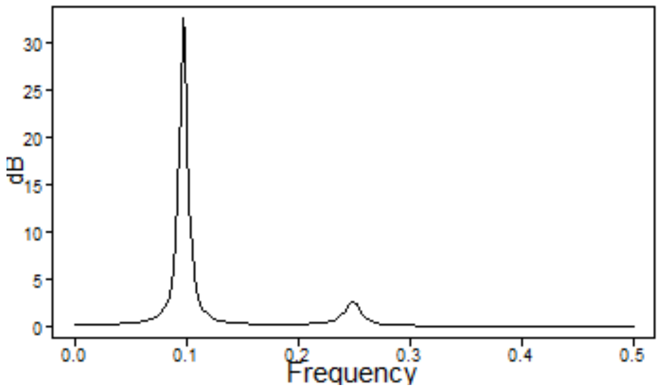
Log of the Spectral Density

In practice, we often plot the log of the spectral density in order to accentuate the secondary peaks.

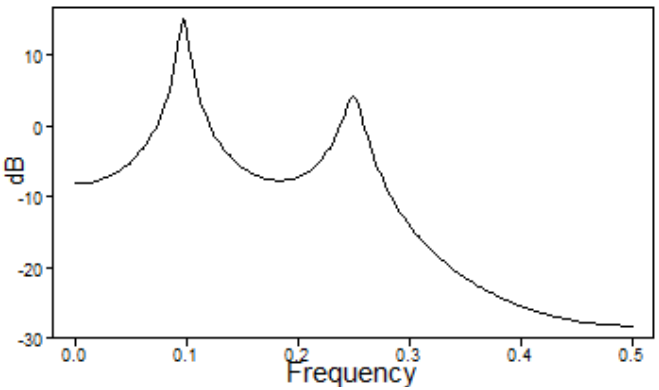
Data with dominant frequency at $f = .1$ and secondary frequency at $f = .25$



Spectral density



$10log_{10}$ spectral density



DataScience@SMU

tswge and the Spectral Density

Slide Deck Note

```
data(patemp)
parzen.wge(patemp)
data(airlog)
parzen.wge(airlog)
parzen.wge(airlog,trunc = 70)
data(bat)
parzen.wge(bat)
data(sunspot.classic)
parzen.wge(sunspot.classic)
plotts.sample.wge(patemp)
plotts.sample.wge(airlog)
plotts.sample.wge(bat)
plotts.sample.wge(sunspot.classic)
#ICC: Estimate the frequency in the Canadian lynx data
data(llynx)
parzen.wge(llynx)
plotts.sample.wge(llynx)
```

DataScience@SMU