Analysis of Time Series

Chapter 7: Introduction to time series analysis in the frequency domain

Edward L. Ionides

Contents

1 Frequency components

Frequency components of a time series

1. A time series dataset (like any other sequence of numbers) can be written as a sum of sine and cosine functions with varying frequencies.

2. This is called the **Fourier representation** or **Fourier transform** of the data.

3. The coefficients corresponding to the sine and cosine at each frequency are called **frequency components** of the data.

4. Looking at which frequencies have large and small components can help to identify appropriate models.

5. Looking at the frequency components present in our models can help to assess whether they are doing a good job of describing our data.

1.1 Eigenvalues and eigenvectors of a stationary covariance matrix

What is the spectrum of a time series model?

- We begin by reviewing eigenvectors and eigenvalues of covariance matrices. This eigen decomposition also arises elsewhere in statistics, e.g. principle component analysis.

- A univariate time series model is a vector-valued random variable $Y_{1:N}$ which we suppose has a covariance matrix V which is an $N \times N$ matrix with entries $V_{mn} = \mathrm{Cov}(Y_m, Y_n)$.

- V is a non-negative definite symmetric matrix, and therefore has N non-negative eigenvalues $\lambda_1, \ldots, \lambda_N$ with corresponding eigenvectors $u_1, \ldots, u_N$ such that

$$V u_n = \lambda_n u_n. \tag{1}$$

- A basic property of these eigenvectors is that they are orthogonal, i.e.,

$$u_m^{\mathrm{T}} u_n = 0 \text{ if } m \neq n. \tag{2}$$

- We may work with **normalized** eigenvectors that are scaled such that $u_n^{\mathrm{T}} u_n = 1$.

- We can also check that the components of Y in the directions of different eigenvectors are uncorrelated.

- Since $\mathrm{Cov}(AY, BY) = A\mathrm{Cov}(Y,Y)B^{\mathrm{T}}$, we have

$$\begin{aligned}
\mathrm{Cov}(u_m^{\mathrm{T}} Y, u_n^{\mathrm{T}} Y) &= u_m^{\mathrm{T}} \mathrm{Cov}(Y,Y) u_n \\
&= u_m^{\mathrm{T}} V u_n \\
&= \lambda_n u_m^{\mathrm{T}} u_n \\
&= \begin{cases} \lambda_n & \text{if } m = n \\ 0 & \text{if } m \neq n \end{cases}
\end{aligned}$$

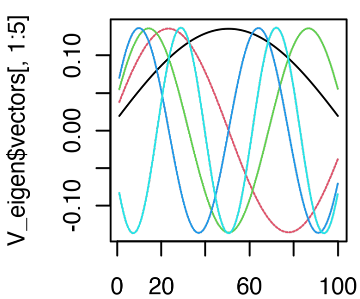
  For the last equality, we have supposed that the eigenvectors are normalized.

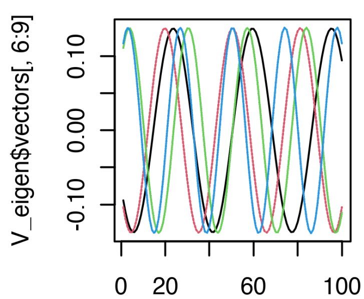
- If we knew V, we could convert the model to a representation where the observable random variables are uncorrelated.

- Transforming the data into its components in the directions of the eigenvectors of the model allows us to use an uncorrelated model. In the Gaussian case, we have independence.

Eigenvectors for the covariance matrix of an AR(1) model with $N = 100$ and $\phi = 0.8$

```
N <- 100;  phi <- 0.8;  sigma <- 1
V <- matrix(NA,N,N)
for(m in 1:N) for(n in 1:N) V[m,n]<-sigma^2*phi^abs(m-n)/(1-phi^2)
V_eigen <- eigen(V,symmetric=TRUE)
matplot(V_eigen$vectors[,1:5],type="l")
matplot(V_eigen$vectors[,6:9],type="l")
```





Eigenvalues for the covariance matrix of an AR(1) model with $N = 100$ and $\phi = 0.8$

- We see that the eigenvectors, plotted as functions of time, look like sine wave oscillations.
- The eigenvalues are

```
round(V_eigen$values[1:9],2)

[1] 24.59 23.44 21.73 19.70 17.57 15.51 13.61 11.91 10.42
```

- We see that the eigenvalues are decreasing. For this model, the components of $Y_{1:N}$ with highest variance correspond to long-period oscillations.
- Are the sinusoidal eigenvectors a special feature of this particular time series model, or something more general?

The eigenvectors for a long stationary time series model

- Suppose $\{Y_n, -\infty < n < \infty\}$ has a stationary autocovariance function γ_h.
- We write Γ for the infinite matrix with entries

$$\Gamma_{m,n} = \gamma_{m-n} \quad \text{for all integers } m \text{ and } n. \tag{3}$$

- An infinite eigenvector is a sequence $u = \{u_n, -\infty < n < \infty\}$ with corresponding eigenvalue λ such that

$$\Gamma u = \lambda u, \tag{4}$$

or, writing out the matrix multiplication explicitly,

$$\sum_{n=-\infty}^{\infty} \Gamma_{m,n} u_n = \lambda u_m \quad \text{for all } m. \tag{5}$$

- We look for a sinusoidal solution, $u_n = e^{2\pi i\omega n}$, where ω is cycles per unit time.

$$\begin{aligned}
\textstyle\sum_{n=-\infty}^{\infty} \Gamma_{m,n} u_n &= \textstyle\sum_{n=-\infty}^{\infty} \gamma_{m-n} u_n \\
&= \textstyle\sum_{h=-\infty}^{\infty} \gamma_h u_{m-h} \quad \text{setting } h = m-n \\
&= \textstyle\sum_{h=-\infty}^{\infty} \gamma_h e^{2\pi i\omega(m-h)} \\
&= e^{2\pi i\omega m} \textstyle\sum_{h=-\infty}^{\infty} \gamma_h e^{-2\pi i\omega h} \\
&= u_m \lambda(\omega) \quad \text{for } \lambda(\omega) = \textstyle\sum_{h=-\infty}^{\infty} \gamma_h e^{-2\pi i\omega h}
\end{aligned}$$

Question 7.1. Why does this calculation show that $u_n(\omega) = e^{2\pi i\omega n}$ is an eigenvector for Γ for any choice of ω.

- The eigenvalue at frequency ω is

$$\lambda(\omega) = \sum_{h=-\infty}^{\infty} \gamma_h \, e^{-2\pi i \omega h} \tag{6}$$

- Viewed as a function of ω, this is called the **spectral density function**.
- $\lambda(\omega)$ is the **Fourier transform** of γ_h.
- An integral version of (6) is used in applied math and engineering:

$$\lambda(\omega) = \int_{-\infty}^{\infty} \gamma(x) \, e^{-2\pi i \omega x} \, dx. \tag{7}$$

- We obtain (6) from (7) when $\gamma(h)$ has a point mass γ_h when h is an integer, and $\gamma(x) = 0$ for non-integer x.
- It was convenient to do this calculation with complex exponentials. However, writing

$$e^{2\pi i \omega n} = \cos(2\pi\omega n) + i \sin(2\pi\omega n), \tag{8}$$

and noting that γ_h is real, we see that the real and imaginary parts of $\lambda(\omega) = \sum_{h=-\infty}^{\infty} \gamma_h e^{-2\pi i \omega h}$ give us two real eigenvectors, $\cos(2\pi\omega n)$ and $\sin(2\pi\omega n)$.

Question 7.2. Review: how would you demonstrate the correctness of the identity $e^{2\pi i \omega} = \cos(2\pi\omega) + i\,\sin(2\pi\omega)$.

- Assuming that this computation for an infinite sum represents a limit of increasing dimension for finite matrices, we have found that the eigenvectors for any long, stationary time series model are approximately sinusoidal.
- For the finite time series situation, we only expect N eigenvectors for a time series of length N. We have one eigenvector for $\omega = 0$, two eigenvectors corresponding to sine and cosine functions with frequency

$$\omega_n = n/N, \text{ for } 0 < n < N/2, \tag{9}$$

and, if N is even, a final eigenvector with frequency

$$\omega_{(N/2)} = 1/2. \tag{10}$$

- These sine and cosine vectors are the **Fourier basis**.
- The time series $y^*_{1:N}$ is the **time domain** representation of the data. Transforming to the Fourier basis gives the **frequency domain** representation.

2 The Fourier transform

Frequency components and the Fourier transform

- The **frequency components** of $Y_{1:N}$ are the components in the directions of these eigenvectors, given by

$$
\begin{aligned}
C_n &= \frac{1}{\sqrt{N}} \sum_{k=1}^{N} Y_k \cos(2\pi\omega_n k) \text{ for } 0 \le n \le N/2, \\
S_n &= \frac{1}{\sqrt{N}} \sum_{k=1}^{N} Y_k \sin(2\pi\omega_n k) \text{ for } 1 \le n \le N/2.
\end{aligned}
$$

- Similarly, the **frequency components** of data $y^*_{1:N}$ are

$$
\begin{aligned}
c_n &= \frac{1}{\sqrt{N}} \sum_{k=1}^{N} y^*_k \cos(2\pi\omega_n k) \text{ for } 0 \le n \le N/2, \\
s_n &= \frac{1}{\sqrt{N}} \sum_{k=1}^{N} y^*_k \sin(2\pi\omega_n k) \text{ for } 1 \le n \le N/2.
\end{aligned}
$$

- The frequency components of the data can be written as real and imaginary parts of the **discrete Fourier transform**,

$$
\begin{aligned}
d_n &= \frac{1}{\sqrt{N}} \sum_{k=1}^{N} y^*_k e^{-2\pi i n/N} \\
&= c_n - i s_n
\end{aligned}
$$

- The normalizing constant of $1/\sqrt{N}$ is convenient for a central limit theorem.
- Various choices about signs and factors of 2π, $\sqrt{2\pi}$ and $\sqrt{N}$ can be made in the definition of the Fourier transform. For example, the `fft` command in R does not include this constant.
- `fft` is an implementation of the fast Fourier transform algorithm, which enables computation of all the frequency components with order $N\log(N)$ computation. Computing the frequency components may appear to require a matrix multiplication involving order N^3 additions and multiplications. When $N = 10^5$ or $N = 10^6$ this difference becomes important!
- The first frequency component, C_0, is a special case, since it has mean $\mu = \mathbb{E}[Y_n]$ whereas the other components have mean zero.
- In practice, we subtract a mean before computing the periodogram, which is equivalent to removing the frequency component for frequency zero.
- The frequency components $(C_{0:N/2}, S_{1:N/2})$ are asymptotically uncorrelated. They are constructed as a sum of a large number of terms, with the usual $1/\sqrt{N}$ scaling for a central limit theorem. So, it may not be surprising that a central limit theorem applies, giving asymptotic justification for the following normal approximation.
- Moving to the frequency domain (i.e., transforming the data to its frequency components) has **decorrelated** the data. Statistical techniques based on assumptions of independence are appropriate when applied to frequency components.

2.1 A normal approximation

Normal approximation for the frequency components

- $(C_{1:N/2}, S_{1:N/2})$ are approximately independent, mean zero, Normal random variables with

$$\mathrm{Var}(C_n) = \mathrm{Var}(S_n) \approx 1/2\lambda(\omega_n). \tag{11}$$

- $C_0/\sqrt{N}$ is approximately Normal, mean μ, independent of $(C_{1:N/2}, S_{1:N/2})$, with

$$\mathrm{Var}(C_0/\sqrt{N}) \approx \lambda(0)/N. \tag{12}$$

- It follows from the normal approximation that, for $1 \le n \le N/2$,

$$C_n^2 + S_n^2 \approx \lambda(\omega_n)\frac{\chi_2^2}{2}, \tag{13}$$

  where χ_2^2 is a chi-squared random variable on two degrees of freedom.

- Taking logs, we have

$$\log\left(C_n^2 + S_n^2\right) \approx \log\lambda(\omega_n) + \log(\chi_2^2/2). \tag{14}$$

3 The periodogram to estimate the spectral density

- These results motivate consideration of the **periodogram**,

$$I_n = {c_n}^2 + {s_n}^2 = \left|d_n\right|^2 \tag{15}$$

  as an estimator of the spectral density.

- $\log I_n$ can be modeled as an estimator of the log spectral density with independent, identically distributed errors.

- We see from the normal approximation that a signal-plus-white-noise model is appropriate for estimating the log spectral density using the log periodogram.

Interpreting the spectral density as a power spectrum

- The power of a wave is proportional to the square of its amplitude.
- The spectral density gives the mean square amplitude of the components at each frequency, and therefore gives the expected power.
- The spectral density function can therefore be called the **power spectrum**.

Question 7.3. Consider the AR(1) model, $\phi(B)Y_n = \epsilon_n$ with $\phi(B) = 1 - \phi_1 B$ and $\epsilon_n \sim \mathrm{WN}(\sigma^2)$, i.e., white noise with variance σ^2. Show that the spectrum of Y is

$$\lambda(\omega) = \frac{\sigma^2}{\left|\phi\left(e^{2\pi i\omega}\right)\right|^2} = \frac{\sigma^2}{1 + \phi^2 - 2\phi\cos(2\pi\omega)}. \tag{16}$$

ARMA models have a rational spectrum

- The calculation for the AR(1) model generalizes. We give the result without proof.
- Let Y_n be an ARMA(p,q) model, $\phi(B)Y_n = \psi(B)\epsilon_n$ with $\epsilon_n \sim \mathrm{WN}(\sigma^2)$. The spectrum of Y is

$$\lambda(\omega) = \sigma^2 \left| \frac{\psi\big(e^{2\pi i\omega}\big)}{\phi\big(e^{2\pi i\omega}\big)} \right|^2 . \tag{17}$$

- The so-called **rational spectrum** of ARMA models is computationally convenient.
- A stationary, causal ARMA model cannot have roots on the unit circle. If a root approaches the unit circle, the denominator in (17) becomes close to zero.

4 Frequency domain data analysis

Michigan winters revisited: Frequency domain methods

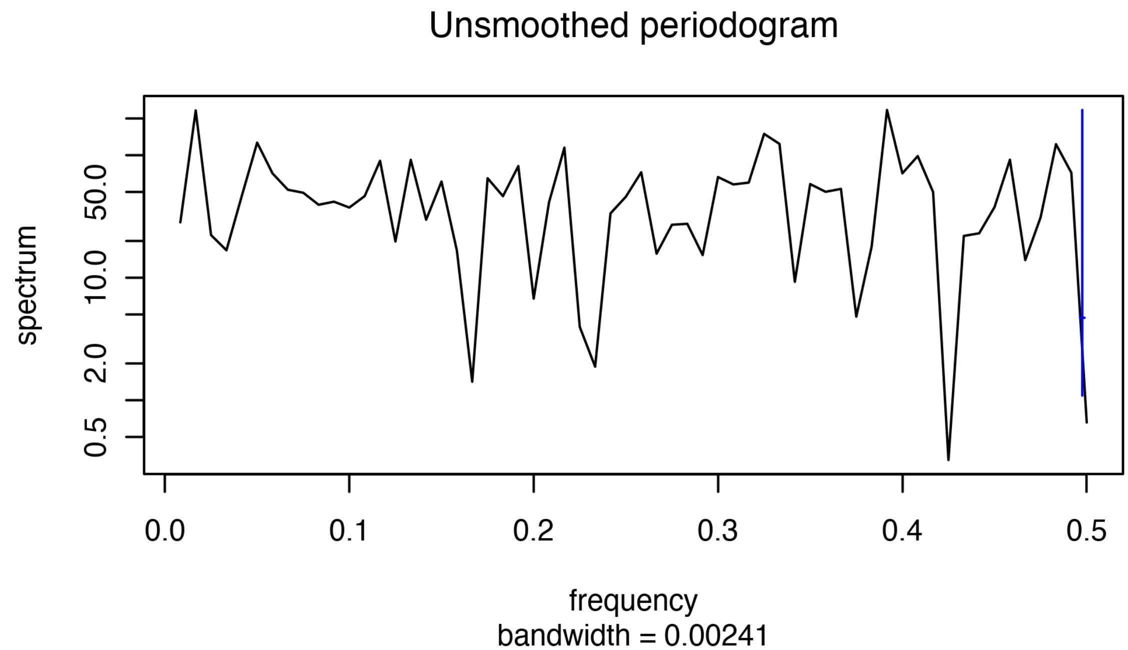
```
y <- read.table(file="ann_arbor_weather.csv",header=TRUE)
head(y[,1:9],3)

 Year Low High Hi_min Lo_max Avg_min Avg_max Mean Precip
 1900  -7   50     36     12      18    34.7 26.3   1.06
 1901  -7   48     37     20      17    31.8 24.4   1.45
 1902  -4   41     27     11      15    30.4 22.7   0.60
```

- We have to deal with the NA measurement for 1955. A simple approach is to replace the NA by the mean.
- What other approaches can you think of for dealing with this missing observation?
- What are the strengths and weaknesses of these approaches?

```
low <- y$Low
low[is.na(low)] <- mean(low, na.rm=TRUE)
```

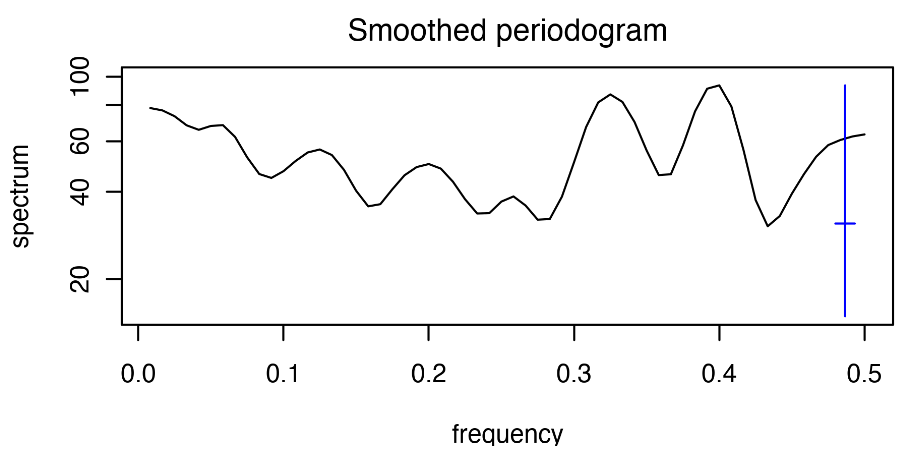
4.1 Smoothing the periodogram

```
spectrum(low, main="Unsmoothed periodogram")
```



- To smooth, we use the default periodogram smoother in R

```
spectrum(low, spans=c(3,5,3), main="Smoothed periodogram",
  ylim=c(15,100))
```



Question 7.4. What is the default periodogram smoother in R?

Question 7.5. How should we use it?

Question 7.6. Why is that default chosen?

4.2 Tapering before calculating the periodogram

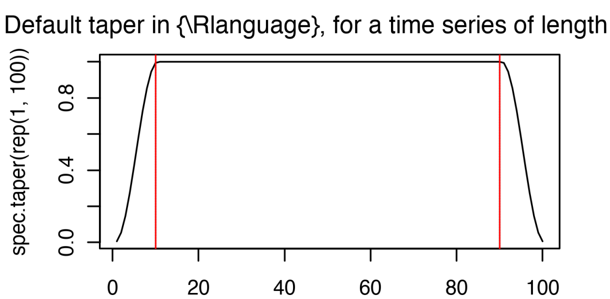
More details on computing and smoothing the periodogram

- To see what R actually does to compute and smooth the periodogram, type `?spectrum`.
- This will lead you to type `?spec.pgram`.
- You will see that, by default, R removes a linear trend, fitted by least squares. This may often be a sensible thing to do. Why?
- You will see that R then multiplies the data by a quantity called a **taper**, computed by `spec.taper`.
- The taper smooths the ends of the time series and removes high-frequency artifacts arising from an abrupt start and end to the time series.
- Formally, from the perspective of the Fourier transform, the time series takes the value zero outside the observed time points $1:N$. The sudden jump to and from zero at the start and end produces unwanted effects in the frequency domain.

The default taper in R smooths the first and last $p = 0.1$ fraction of the time points, by modifying the detrended data $y^*_{1:N}$ to tapered version $z_{1:N}$ defined by

$$z_n = \begin{cases} y^*_n\big(1-\cos(\pi n/Np)\big)/2 & \text{if } 1 \leq n < Np \\ y^*_n & \text{if } Np \leq n \leq N(1-p) \\ y^*_n\big(1-\cos(\pi[N+1-n]/Np)\big)/2 & \text{if } N(1-p) < n \leq N \end{cases}$$

```
plot(spec.taper(rep(1,100)),type="l",
  main="Default taper in {\\Rlanguage}, for a time series of length 100")
abline(v=c(10,90),lty="dotted",col="red")
```

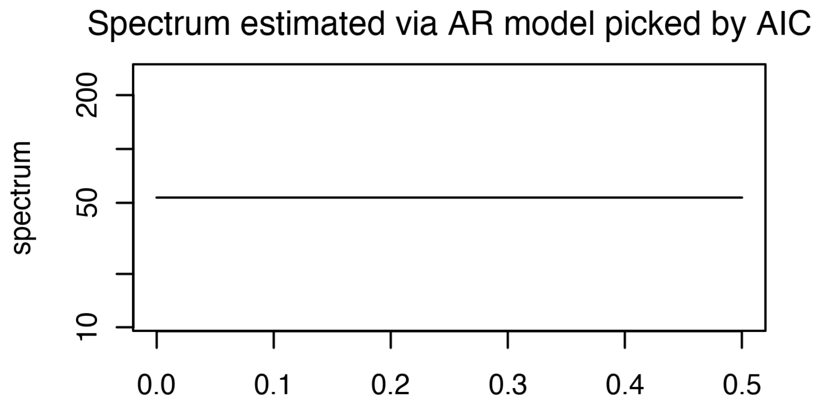


4.3 Fitting an AR model to estimate the spectrum

Spectral density estimation by fitting a model

Another standard way to estimate the spectrum is to fit an AR(p) model with p selected by AIC.

```
spectrum(low,method="ar",
  main="Spectrum estimated via AR model picked by AIC")
```



Units of frequency and period

- When we call ω the frequency in cycles per unit time, we really mean **cycles per unit observation**.
- Suppose the time series consists of equally spaced observations, with $t_n - t_{n-1} = \Delta$ years. Then, the frequency is ω/Δ **cycles per year**.
- The **period** of an oscillation is the time for one cycle,

$$\text{period} = \frac{1}{\text{frequency}}. \tag{18}$$

- When the observation intervals have a time unit (years, seconds, etc) we usually use that unit for the period, and its inverse for the frequency.

Further reading

- Sections 4.1 to 4.3 of Shumway and Stoffer (2017) cover similar topics to this chapter.

License, acknowledgments, and links

- The materials builds on previous courses.
- Compiled on February 13, 2021 using R version 4.0.3.

Back to course homepage

References

Shumway RH, Stoffer DS (2017). *Time Series Analysis and its Applications: With R Examples*. Springer. URL http://www.stat.pitt.edu/stoffer/tsa4/tsa4.pdf. 27