7. Introduction to time series analysis in the frequency domain

Edward Ionides

2/7/2018

Contents

Objectives

- This course emphasizes time domain analysis of time series, but we also want to be able to present and interpret the frequency domain properties of our time series models and data.

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

7.1 What is the spectrum of a time series model?

- We're going to start by reviewing eigenvectors and eigenvalues of covariance matrices. This eigen decomposition also arises elsewhere in Statistics, such as the principle component analysis technique in multivariate 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

$$V u_n = \lambda_n u_n.$$

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

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

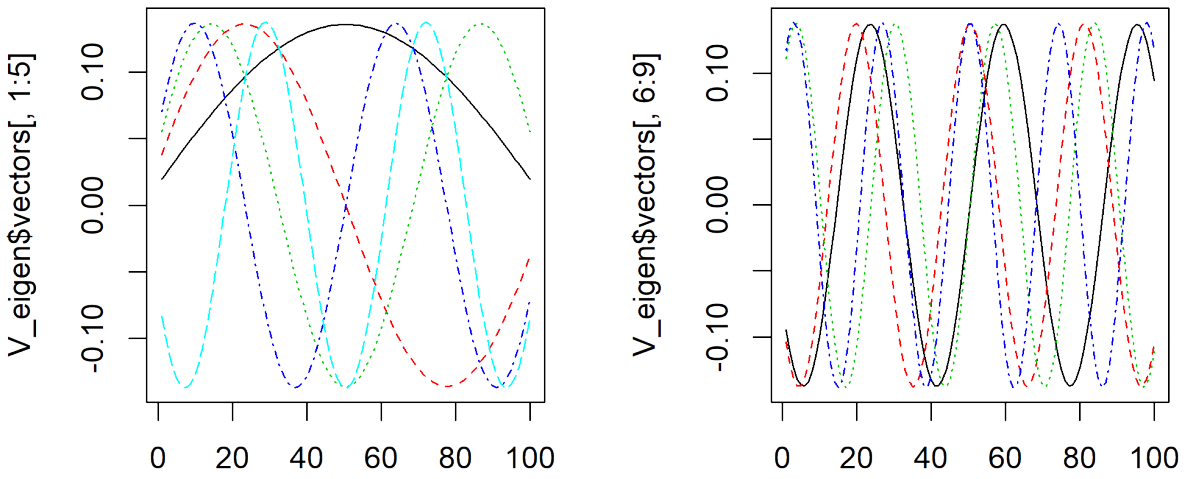
- We may work with **normalized** eigenvectors that are scaled such that $u_n^T u_n = 1$.
- We can also check that the components of Y in the directions of different eigenvectors are uncorrelated. Since $\mathrm{Cov}(AY, BY) = A\mathrm{Cov}(Y, Y)B^T$, we have

$$\begin{aligned} \mathrm{Cov}(u_m^T Y, u_n^T Y) &= u_m^T \mathrm{Cov}(Y,Y) u_n && (1)\\ &= u_m^T V u_n && (2)\\ &= \lambda_n u_m^T u_n \qquad = \begin{cases} \lambda_n & \text{if } m = n \\ 0 & \text{if } m \neq n \end{cases} && (3) \end{aligned}$$

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

- Thus, if we knew V, we could convert the model to a representation where the observable random variables are uncorrelated.
- Specifically, we could transform the data into its components in the directions of the eigenvectors of the model. An uncorrelated (or, in the Gaussian case, independent) model would then become appropriate for this transformation of the data.
- Let's see how to do that for a stationary time series model, say 100 observations from an AR(1) model with autoregressive coefficient 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)
oldpars <- par(mfrow=c(1,2))
matplot(V_eigen$vectors[,1:5],type="l")
matplot(V_eigen$vectors[,6:9],type="l")
```




```
par(oldpars)
```

- 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?

7.1.1 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

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

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

$$\Gamma u = \lambda u,$$

or, writing out the matrix multiplication explicitly,

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

- Now, let's look for a sinusoidal solution, $u_n = e^{i\omega n}$. Then,

$$\begin{aligned}
\sum_{n=-\infty}^{\infty} \Gamma_{m,n} u_n &= \sum_{n=-\infty}^{\infty} \gamma_{m-n} u_n && (4)\\
&= \sum_{h=-\infty}^{\infty} \gamma_h u_{m-h} \quad \text{setting } h = m-n && (5)\\
&= \sum_{h=-\infty}^{\infty} \gamma_h e^{i\omega(m-h)} && (6)\\
&= e^{i\omega m} \sum_{h=-\infty}^{\infty} \gamma_h e^{-i\omega h} && (7)\\
&= u_m \lambda \text{ for } \lambda = \sum_{h=-\infty}^{\infty} \gamma_h e^{-i\omega h} && (8)
\end{aligned}$$

- This calculation shows that

$$u_n(\omega) = e^{i\omega n}$$

is an eigenvector for Γ for any choice of ω. The corresponding eigenvalue function,

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

is called the **spectral density function**. It is calculated as the **Fourier transform** of γ_h at *frequency* ω.

- It was convenient to do this calculation with complex exponentials. However, writing

$$e^{i\omega n} = \cos(\omega n) + i\sin(\omega n)$$

we see that the real and imaginary parts of this calculation in fact give us two real eigenvectors, $\cos(\omega n)$ and $\sin(\omega n)$.

- Assuming that this computation for an infinite sum represents a limit of increasing dimension for finite matrices, we have found that the eigenfunctions 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

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

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

$$\omega_{(N/2)} = \pi.$$

- These sine and cosine vectors are called the **Fourier basis**.

7.2 Frequency components of the data and their representation via the Fourier transform

- The **frequency components** of $Y_{1:N}$ are the components in the directions of these eigenvectors. Equivalently, we could say they are the representation of $Y_{1:N}$ in the Fourier basis. Specifically, we write

$$C_n = \frac{1}{\sqrt{N}}\sum_{k=1}^{N} Y_k \cos(\omega_n k) \text{ for } 0 \le n \le N/2, \tag{9}$$

$$S_n = \frac{1}{\sqrt{N}}\sum_{k=1}^{N} Y_k \sin(\omega_n k) \text{ for } 1 \le n \le N/2. \tag{10}$$

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

$$c^*_n = \frac{1}{\sqrt{N}}\sum_{k=1}^{N} y^*_k \cos(\omega_n k) \text{ for } 0 \le n \le N/2, \tag{11}$$

$$s^*_n = \frac{1}{\sqrt{N}}\sum_{k=1}^{N} y^*_k \sin(\omega_n k) \text{ for } 1 \le n \le N/2. \tag{12}$$

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

$$d^*_n = \frac{1}{\sqrt{N}}\sum_{k=1}^{N} y^*_k e^{2\pi i n/N} \tag{13}$$

$$= c^*_n + i s^*_n \tag{14}$$

- Here, we have made a decision to introduce a normalizing constant of $1/\sqrt{N}$. There are various choices about signs and factors of 2π, $\sqrt{2\pi}$ and $\sqrt{N}$ that can—and are—made in the definition of the Fourier transform in various situations. One should try to be consistent, and also be careful: the `fft` command in R, for example, doesn't 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. At first consideration, computing the frequency components appears to require a matrix multiplication involving order N^2 additions and multiplications. When $N = 10^5$ or $N = 10^6$ this difference becomes important!
- The first frequency component, C_0, is something of 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.

7.2.1 Normal approximation for the frequency components

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

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

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

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

- 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 become applicable.

  Converting data to its frequency components and looking at the periodogram (frequency domain analysis) decorrelates the data

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

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

  where χ_2^2 is a chi-squared random variable on two degrees of freedom.
- Taking logs, we have

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

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

$$I_n = (c_n^*)^2 + (s_n^*)^2 = \left|d_n^*\right|^2$$

  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.

7.2.2 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**.

7.2.3 Question: compute the spectrum of an AR(1) model.

From the book:

Property 4.2 The Spectral Density

If the autocovariance function, $\gamma(h)$, of a stationary process satisfies

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty, \tag{4.15}$$

then it has the representation

$$\gamma(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i\omega h} f(\omega)\, d\omega \quad h = 0, \pm 1, \pm 2, \ldots \tag{4.16}$$

as the inverse transform of the spectral density,

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i\omega h} \quad -1/2 \le \omega \le 1/2. \tag{4.17}$$

ANS.

$$\begin{aligned}
\lambda(\omega) &= \sum_{h=-\infty}^{\infty} \gamma_h e^{-i\omega h} \\
&= \sum_{h=-\infty}^{\infty} \frac{\sigma^2 \rho^{|h|}}{1-\rho^2} e^{-i\omega h} \\
&= \frac{\sigma^2}{1-\rho^2} + \underbrace{\sum_{h=1}^{\infty} \frac{\sigma^2}{1-\rho^2} \rho^h \left[e^{-i\omega h} + e^{i\omega h}\right]}_{\text{use geometric sum formula}} \\
&= \frac{\sigma^2}{1-\rho^2} \left[1 + \sum_{h=1}^{\infty} \left(\rho e^{-i\omega}\right)^h + \sum_{h=1}^{\infty} \left(\rho e^{i\omega}\right)^h\right]
\end{aligned}$$

7.2.4 Question: compute the spectrum of the MA(q) moving mean,

$$Y_n = \frac{1}{q+1} \sum_{k=0}^{q} \epsilon_{n-k}.$$

7.2.5 Review question: how would you demonstrate the correctness of the identity,

$$e^{i\omega} = \cos(\omega) + i \sin(\omega).$$

Use Taylor series expansion

7.3 Some data analysis using the frequency domain

7.3.1 Michigan winters revisited

- Recall the Ann Arbor January weather data,

```
system("head ann_arbor_weather.csv",intern=TRUE)
```

```
##  [1] "## https://weather-warehouse.com/WeatherHistory/PastWeatherData_AnnArborUnivOfMi_AnnArbor_MI_Ja
##  [2] "## accessed by ELI on Nov 2, 2015"
##  [3] "##"
##  [4] "## Full column descriptions:"
##  [5] "## Lowest Temperature (F)"
##  [6] "## Highest Temperature (F)"
##  [7] "## Warmest Minimum Temperature (F)"
##  [8] "## Coldest Maximum Temperature (F)"
##  [9] "## Average Minimum Temperature (F)"
## [10] "## Average Maximum Temperature (F)"
```

```
y <- read.table(file="ann_arbor_weather.csv",header=TRUE)
head(y)
```

```
##   Year Low High Hi_min Lo_max Avg_min Avg_max Mean Precip Snow Hi_Pricip
## 1 1900  -7   50     36     12    18.0    34.7 26.3   1.06  4.0       0.28
## 2 1901  -7   48     37     20    17.0    31.8 24.4   1.45 10.1       0.40
## 3 1902  -4   41     27     11    15.0    30.4 22.7   0.60  6.0       0.25
## 4 1903  -7   50     36     12    15.1    29.6 22.4   1.27  7.3       0.40
## 5 1904 -11   38     31      6     8.2    22.9 15.3   2.51 11.0       0.67
## 6 1905  -3   47     32     14    10.9    25.9 18.4   1.64  7.9       0.84
##   Hi_Snow
## 1     1.1
## 2     3.2
## 3     2.5
## 4     3.2
## 5     2.1
## 6     2.5
```

```
low <- y$Low
```

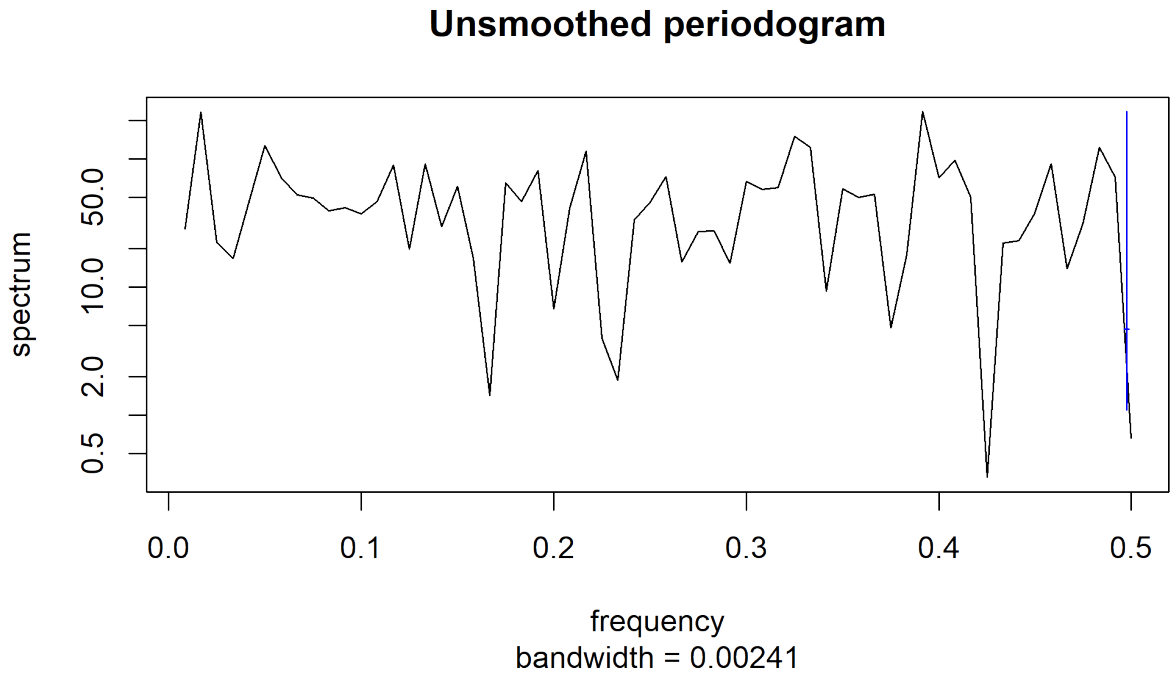
- We have to deal with the NA measurement for 1955. A simple approach is to replace the NA by the mean.
  - What other approaches can you think of for dealing with this missing observation?

    Can fit it with a suitable mean, can remove it.

  - What are the strengths and weaknesses of these approaches?

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

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



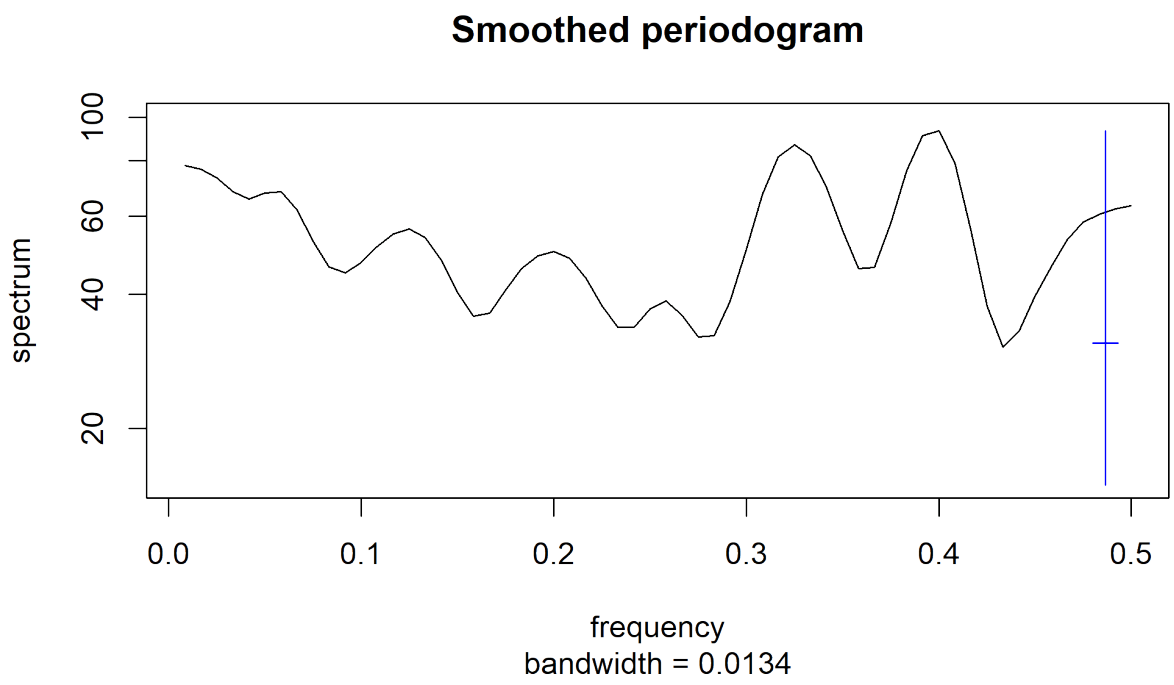
Here, "frequency" is $\omega_n = 2\pi n/N$, for $0 < n < N/2$. Frequency in "cycles per data point"

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

7.3.2 Question: how does R smooth?

- What is the default periodogram smoother in R?
- How should we use it?
- Why is that default chosen?

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

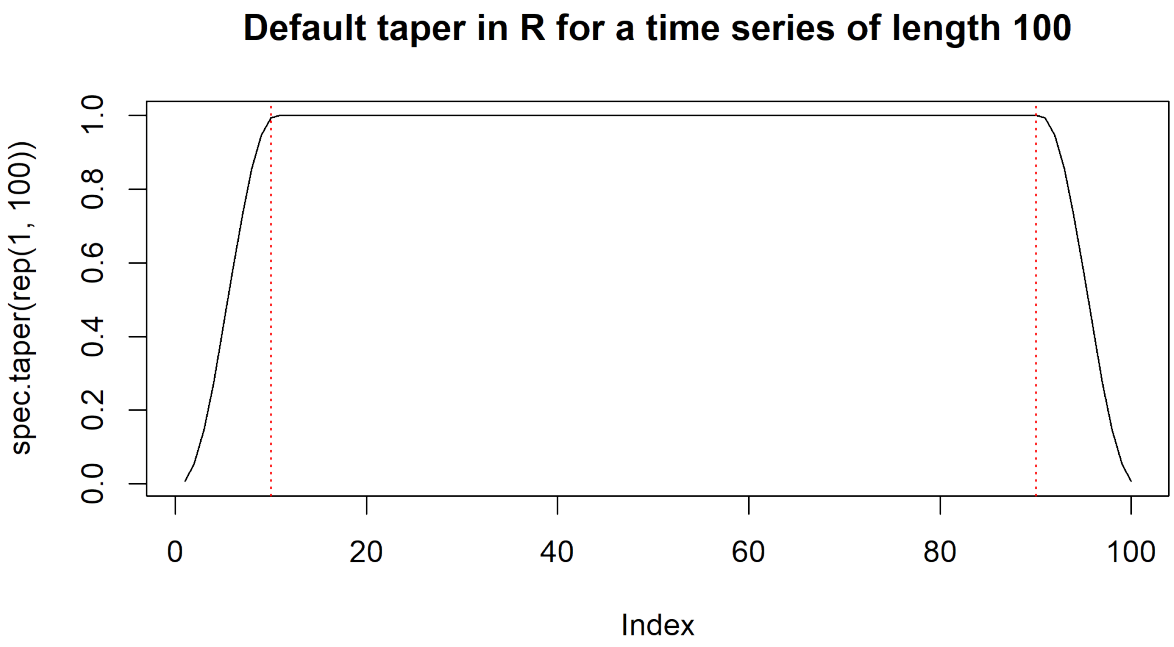


7.3.3 More details on computing and smoothing the periodogram

- 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

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

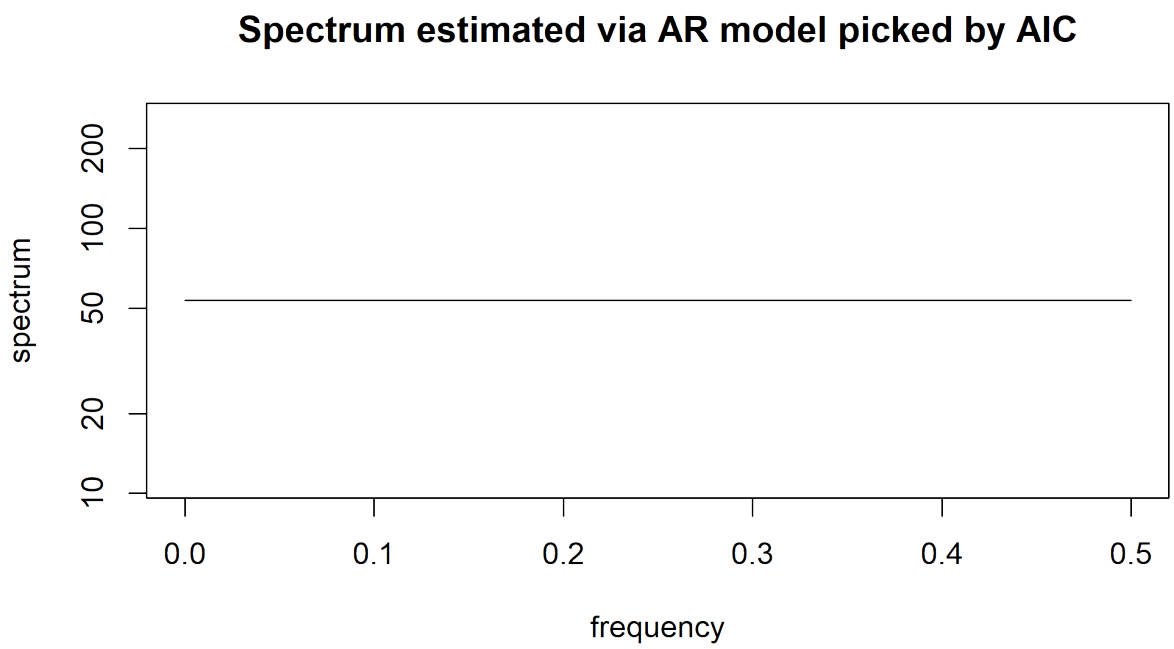
```
plot(spec.taper(rep(1,100)),type="l",
  main="Default taper in R for a time series of length 100")
abline(v=c(10,90),lty="dotted",col="red")
```



7.3.4 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")
```



7.3.5 Fitting a signal plus white noise model

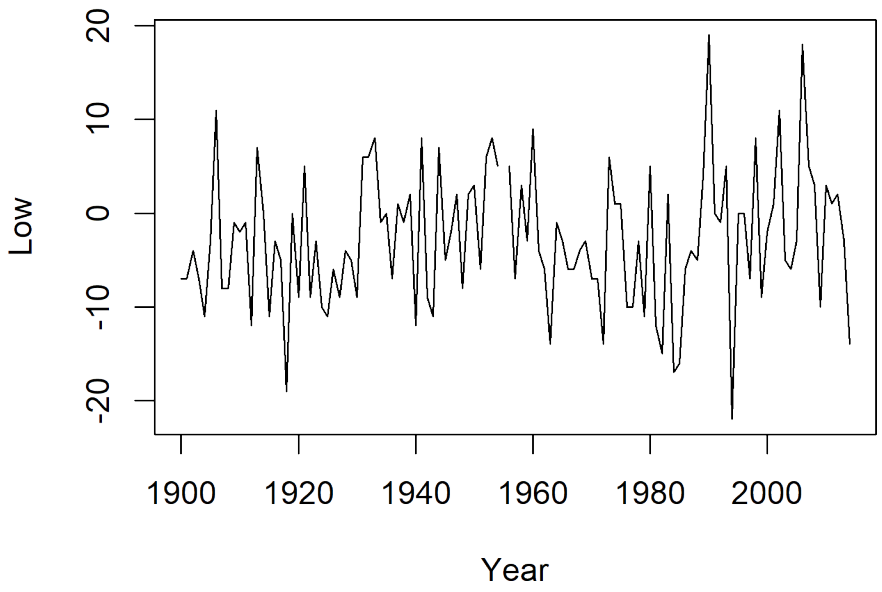
- Since this time series is well modeled by white noise, we could fit a signal plus white noise model. This might be a more sensitive way to look for a trend.
- Let's try some low-order polynomial trend specifications.

```
lm0 <- lm(Low~1,data=y)
lm1 <- lm(Low~Year,data=y)
lm2 <- lm(Low~Year+I(Year^2),data=y)
lm3 <- lm(Low~Year+I(Year^2)+I(Year^3),data=y)
poly_aic <- matrix( c(AIC(lm0),AIC(lm1),AIC(lm2),AIC(lm3)), nrow=1,
   dimnames=list("<b>AIC</b>", paste("order",0:3)))
require(knitr)
kable(poly_aic,digits=1)
```

	order 0	order 1	order 2	order 3
AIC	780.9	781.6	783.6	783.8

- Still no evidence suggesting anything other than a white noise model.
- Now, let's remind ourselves what the data look like.

```
plot(Low~Year,data=y,type="l")
```



- Our eye may see a trend, and recall that it looks a bit like the global temperature trend.
- Let's try fitting global temperature as an explanatory variable.

```
Z <- read.table("Global_Temperature.txt",header=TRUE)
global_temp <- Z$Annual[Z$Year %in% y$Year]
lm_global <- lm(Low~global_temp,data=y)
AIC(lm_global)
```

```
## [1] 779.8877
```

- Got it! We have an explantion of the trend that makes scientific sense. However, the model is prefered by AIC but is not quite statistically significant viewed as a 2-sided test against a null of white noise via a t-statistic for the estimated coefficient:

```
summary(lm_global)
```

```
##
## Call:
## lm(formula = Low ~ global_temp, data = y)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.1554  -4.5079  -0.1595   4.1068  20.3959
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -3.0413     0.6916  -4.397 2.51e-05 ***
## global_temp   3.7396     2.1530   1.737   0.0851 .
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.273 on 112 degrees of freedom
##   (1 observation deleted due to missingness)
## Multiple R-squared:  0.02623,    Adjusted R-squared:  0.01754
## F-statistic: 3.017 on 1 and 112 DF,  p-value: 0.08515
```

1-sided p-value is $0.043 < 0.05$, 2–sided p-value is 0.085

7.3.6 Question: is a 2-sided test or a 1-sided test more reasonable here?

- What is the p-value for the 1-sided test?

- Perhaps we now have a satisfactory understanding of Ann Arbor January low temperatures: random, white noise, variation around the global mean temperature trend.
- With noisy data, uncovering what is going on can take careful data analysis together with scientific reasoning.
- What could we do to improve the signal to noise ratio in this analysis?

7.3.7 Question: Why might you expect the estimated coefficient to be statistically insignificant in this analysis, even if the model with global mean temperature plus trend is correct?

Hint: notice the standard error on the coefficient, together with consideration of possible values of the coefficient in a scientifically plausible model.

7.4 Units of frequency and period

- It is always good practice to be explicit about the units of quantities. For frequency domain analysis we can consider various options for units of frequency.
- For a frequency component corresponding to $\sin(2\pi\omega n)$ or $\cos(2\pi\omega n)$, we say that the frequency is ω **cycles per unit time**.
- Suppose the time series consists of equally spaced observations, with $t_n - t_{n-1} = \Delta$ years. Then we say that the frequency is ω/Δ **cycles per year**.
- For a frequency component corresponding to $\sin(\nu t)$ or $\cos(\nu t)$, we say that the frequency is ν **radians per unit time**.

- The **period** of an oscillation is the time for one cycle. So, when frequency is measured in cycles per time, we have

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

  Thus, for a frequency component corresponding to $\sin(2\pi\omega n)$ or $\cos(2\pi\omega n)$, the period is $1/\omega$ observation intervals.

- When the observation intervals have constant time length (years, seconds, etc) we should use those units for the period.