



Published on STAT 510 (<https://onlinecourses.science.psu.edu/stat510>)

[Home](#) > Lesson 12: Spectral Analysis

Lesson 12: Spectral Analysis

Assignments:

- Read Sections 4.3, 4.5, and 4.6 of your text.
- Read through the Lesson 12 online notes that follow.
- Complete Lesson 12 Assignment.

Overview:

This week we'll cover estimation of the spectral density.

Learning Objectives:

After successfully completing this lesson, you should be able to:

- Estimate the spectral density non-parametrically (Daniell kernel & modified Daniell kernel)
- Identify and interpret bandwidth
- Estimate the spectral density parametrically

12.1: Estimating the Spectral Density

We previously discussed the periodogram, a function/graph that displays information about the periodic components of a time series. Any time series can be expressed as a sum of cosine and sine waves oscillating at the fundamental (harmonic) frequencies $= j/n$, with $j = 1, 2, \dots, n/2$. The periodogram gives information about the relative strengths of the various frequencies for explaining the variation in the time series.

The periodogram is a sample estimate of a population function called the spectral density, which is a frequency domain characterization of a population stationary time series. The spectral density is a frequency domain representation of a time series that is directly related to the autocovariance time domain representation. In essence the spectral density and the autocovariance function contain the same information, but express it in different ways.

[Review Note: The autocovariance is the numerator of the autocorrelation. The autocorrelation is the autocovariance divided by the variance.]

Suppose that $\gamma(h)$ is the autocovariance function of a stationary process and that $f(\omega)$ is the spectral density for the same process. In the notation of the previous sentence, h = time lag and ω = frequency.

The autocovariance and the spectral density have the following relationships:

$$\gamma(h) = \int_{-1/2}^{1/2} e^{2\pi i \omega h} f(\omega) d\omega, \text{ and}$$

$$f(\omega) = \sum_{h=-\infty}^{h=\infty} \gamma(h) e^{-2\pi i \omega h}$$

In the language of advanced calculus, the autocovariance and spectral density are Fourier transform pairs. We won't worry about the calculus of the situation. We'll focus on the estimation of the spectral density – the frequency domain characterization of a series. The Fourier transform equations are only given here to establish that there is a direct link between the time domain representation and the frequency domain representation of a series.

Mathematically, the spectral density is defined for both negative and positive frequencies. However, due to symmetry of the function and its repeating pattern for frequencies outside the range $-1/2$ to $+1/2$, we only need to be concerned with frequencies between 0 and $+1/2$.

The “total” integrated spectral density equals the variance of the series. Thus the spectral density within a particular

interval of frequencies can be viewed as the amount of the variance explained by those frequencies.

Methods for Estimating the Spectral Density

The raw periodogram is a rough sample estimate of the population spectral density. The estimate is “rough”, in part, because we only use the discrete fundamental harmonic frequencies for the periodogram whereas the spectral density is defined over a continuum of frequencies.

One possible improvement to the periodogram estimate of the spectral density is to smooth it using centered moving averages. An additional “smoothing” can be created using tapering methods which weight the ends (in time) of the series less than the center of the data. We’ll not cover tapering in this lesson. Interested parties can see Section 4.5 in the book and various Internet sources.

An alternative approach to smoothing the periodogram is a parametric estimation approach based on the fact that any stationary time series can be approximated by an AR model of some order (although it might be a high order). In this approach a suitable AR model is found, and then the spectral density is estimated as the spectral density for that estimated AR model.

Smoothing Method (Nonparametric Estimation of the Spectral Density)

The usual method for smoothing a periodogram has such a fancy name that it sounds difficult. In fact, it’s merely a centered moving average procedure with a few possible modifications. For a time series, the Daniell kernel with parameter m is a centered moving average which creates a smoothed value at time t by averaging all values between times $t - m$ and $t + m$ (inclusive). For example, the smoothing formula for a Daniell kernel with $m = 2$ is

$$\hat{x}_t = \frac{x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2}}{5}$$

In R, the weighting coefficients for a Daniell kernel with $m = 2$ can be generated with the command `kernel("daniell", 2)`. The result is

```
coef[-2] = 0.2
coef[-1] = 0.2
coef[ 0] = 0.2
coef[ 1] = 0.2
coef[ 2] = 0.2
```

The subscripts for `coef []` refer to the time difference from the center of the average at time t . Thus the smoothing formula in this instance is

$$\hat{x}_t = 0.2x_{t-2} + 0.2x_{t-1} + 0.2x_t + 0.2x_{t+1} + 0.2x_{t+2},$$

which is the same as the formula given above.

The modified Daniell kernel is such that the two endpoints in the averaging receive half the weight that the interior points do. For a modified Daniell kernel with $m = 2$, the smoothing is

$$\hat{x}_t = \frac{x_{t-2} + 2x_{t-1} + 2x_t + 2x_{t+1} + x_{t+2}}{8} = 0.125x_{t-2} + 0.25x_{t-1} + 0.25x_t + 0.25x_{t+1} + 0.125x_{t+2}$$

In R, the command `kernel("modified.daniell", 2)` will list the weighting coefficients just used.

Either the Daniell kernel or the modified Daniell kernel can be convoluted (repeated) so that the smoothing is applied again to the smoothed values. This produces a more extensive smoothing by averaging over a wider time interval. For instance, to repeat a Daniell kernel with $m = 2$ on the smoothed values that resulted from a Daniell kernel with $m = 2$, the formula would be

$$\hat{\hat{x}}_t = \frac{\hat{x}_{t-2} + \hat{x}_{t-1} + \hat{x}_t + \hat{x}_{t+1} + \hat{x}_{t+2}}{5}$$

This is the average of the smoothed values within two time periods of time t , in either direction.

In R, the command `kernel("daniell", c(2, 2))` will supply the coefficients that would be applied to as weights in averaging the original data values for a convoluted Daniell kernel with $m = 2$ in both smoothings. The result is

```
> kernel("daniell",c(2,2))
coef[-4] = 0.04
coef[-3] = 0.08
coef[-2] = 0.12
coef[-1] = 0.16
coef[ 0] = 0.20
coef[ 1] = 0.16
coef[ 2] = 0.12
coef[ 3] = 0.08
coef[ 4] = 0.04
```

This generates the smoothing formula

$$\hat{x}_t = 0.04x_{t-4} + 0.08x_{t-3} + 0.12x_{t-2} + 0.16x_{t-1} + 0.20x_t + 0.16x_{t+1} + 0.12x_{t+2} + 0.08x_{t+3} + 0.04x_{t+4}.$$

A convolution of the modified method in which the end points have less weight is also possible. The command `kernel("modified.daniell",c(2,2))` gives these coefficients:

```
coef[-4] = 0.01563
coef[-3] = 0.06250
coef[-2] = 0.12500
coef[-1] = 0.18750
coef[ 0] = 0.21875
coef[ 1] = 0.18750
coef[ 2] = 0.12500
coef[ 3] = 0.06250
coef[ 4] = 0.01563
```

Thus the center values are weighted slightly more heavily than in the unmodified Daniell kernel.

When we smooth a periodogram, we are smoothing across a frequency interval rather than a time interval. Remember that the periodogram is determined at the fundamental frequencies $\omega_j = j/n$ for $j = 1, 2, \dots, n/2$. Let $I(\omega_j)$ denote the periodogram value at frequency $\omega_j = j/n$. When we use a Daniell kernel with parameter m to smooth a periodogram, the smoothed value $\hat{I}(\omega_j)$ is a weighted average of periodogram values for frequencies in the range $(j-m)/n$ to $(j+m)/n$.

Bandwidth

There are $L = 2m+1$ fundamental frequency values in the range $(j-m)/n$ to $(j+m)/n$, the range of values used for smoothing. The bandwidth for the smoothed periodogram is defined as

$$B_\omega = \frac{L}{n}$$

The bandwidth is a measure of the width of the frequency interval(s) used for smoothing the periodogram.

When unequal weights are used in the smoothing, the bandwidth definition is modified. Denote the smoothed periodogram value at $\omega_j = j/n$ as

$$\hat{I}(\omega_j) = \sum_{k=-m}^{+m} h_k I\left(\omega_j + \frac{k}{n}\right).$$

The h_k are the possibly unequal weights used in the smoothing. The bandwidth formula is then modified to

$$B_\omega = \frac{L_h}{n} = \frac{1/\sum h_k^2}{n}$$

Actually, this formula works for equal weights too.

The bandwidth should be sufficient to smooth our estimate, but if we use a bandwidth that is too great, we'll smooth out the periodogram too much and miss seeing important peaks. In practice, it usually takes some experimentation to find the bandwidth that gives a suitable smoothing.

The bandwidth is predominately controlled by the number of values that are averaged in the smoothing. In other words,

the m parameter for the Daniell kernel and whether the kernel is convoluted (repeated) affect the bandwidth.

Note: The bandwidths R reports with its plots don't match the values that would be calculated using the formulas above. Please see the footnote on p. 197 of your text for an explanation.

R Code

Averaging/smoothing the periodogram with a Daniell kernel can be accomplished in R using a sequence of two commands. The first defines a Daniell kernel and the second creates the smoothed periodogram.

As an example, suppose that the observed series is named x and we wish to smooth the periodogram using a Daniell kernel with $m = 4$. The commands are

```
k = kernel("daniell", 4)
spec.pgram(x, k, taper=0, log = "no")
```

The first command creates the weighting coefficients needed for the smoothing and stores them in a vector named k . (It's arbitrary to call it k . It could be called anything.) The second command asks for a spectral density estimate based on the periodogram for the series x , using the weighting coefficients stored in k , with no taper, and the plot will be on an ordinary scale, not a log scale.

If a convolution is desired, the kernel command could be modified to something like `k = kernel("daniell", c(4,4))`.

There are two possible ways to achieve a modified Daniell kernel. You can either change the kernel command to refer to the "modified.daniell" rather than "daniell" or you can skip using the kernel command and use a spans parameter in the `spec.pgram` command.

The spans parameter gives the length ($=2m+1$) of the desired modified Daniell kernel. For instance, a modified Daniell kernel with $m = 4$ has length $L = 2m+1 = 9$ so the we could use the command

```
spec.pgram(x, spans=9, taper = 0, log="no")
```

Two passes of a modified Daniell kernel with $m = 4$ on each pass can be done using

```
spec.pgram(x, spans=c(9,9), taper = 0, log="no")
```

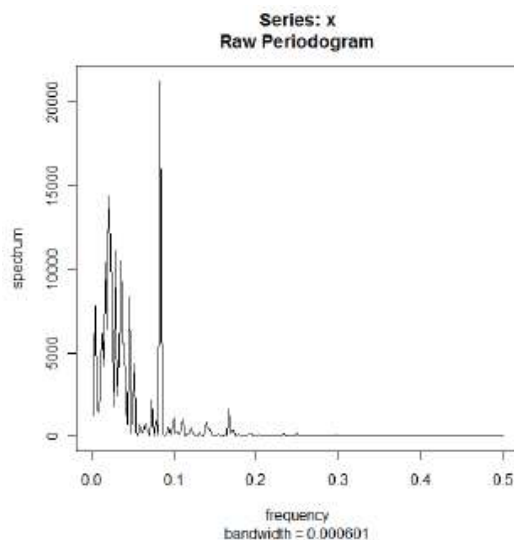
Example: This example will use the fish recruitment series that's used in several places in the text, including several places in chapter 4. The series consists of $n = 453$ monthly values of a measure of a fish population in a southern hemisphere location. The data are in the file `recruit.dat`.

The raw periodogram can be created using the command (or it could be created using the method given in Lesson 6).

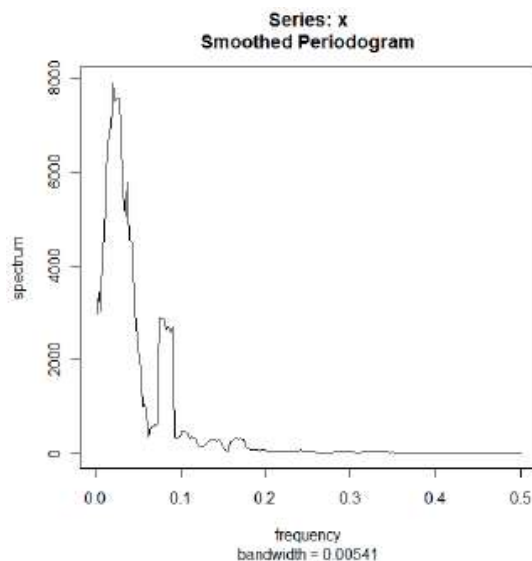
```
spec.pgram(x, taper=0, log="no")
```

Note that in the command just given we have omitted the parameter that gives weights for smoothing.

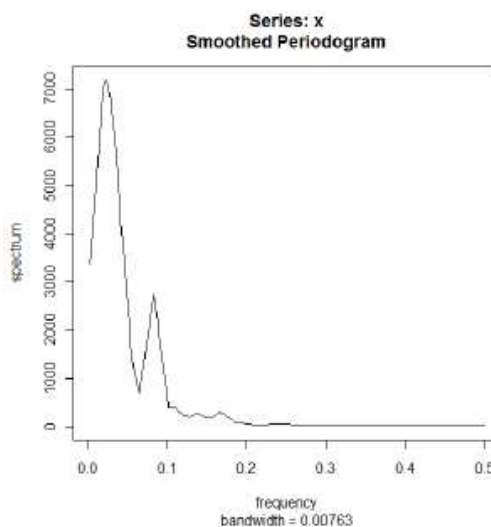
The raw periodogram follows:



The next plot is a smoothed periodogram using a Daniell kernel with $m = 4$. Note that one effect of the smoothing is that the dominant peak in the unsmoothed version is now the second tallest peak. This happened because the peak is so sharply defined in the unsmoothed version that when we average it with a few surrounding values the height is reduced.



The next plot is a smoothed periodogram using two passes of a Daniell kernel with $m = 4$ on each pass. Note how it is even more smoothed than previously.



To learn where the two dominant peaks are located, assign a name to the `spec.pgram` output and then you can list it. For instance,

```
specvalues = spec.pgram(x, k, taper=0, log="no")
specvalues
```

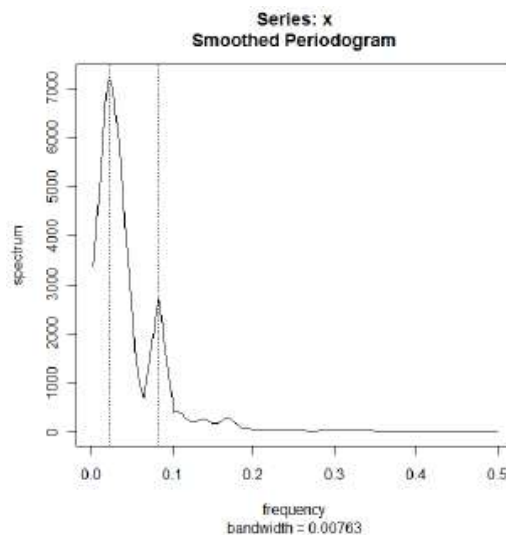
You can sift through the output to find the frequencies at which the peaks occur. The frequencies and spectral density estimates are listed separately, but in the same order. Identify the maximum spectral densities and then find the corresponding frequencies.

Here, the first peak is at a frequency $\approx .0229$. The period (number of months) associated with this cycle = $1/.0229 = 43.7$ months, or about 44 months. The second peak occurs at a frequency ≈ 0.083333 . The associated period = $1/.08333 = 12$ months. The first peak is associated with an El Nino weather effect. The second is the usual 12 month seasonal effect.

These two commands will put vertical dotted lines onto the (estimated) spectral density plot at the approximate locations of the peak densities.

```
abline(v=1/44, lty="dotted")
abline(v=1/12, lty = "dotted")
```

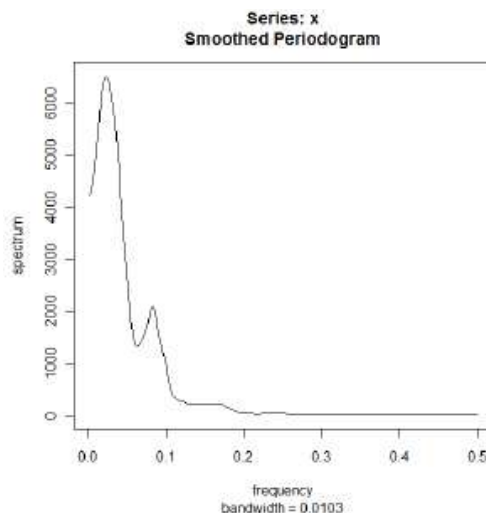
Here's the resulting plot:



We've smoothed enough, but for demonstration purposes, the next plot is the result of

```
spec.pgram(x, spans=c(13,13), taper=0, log="no")
```

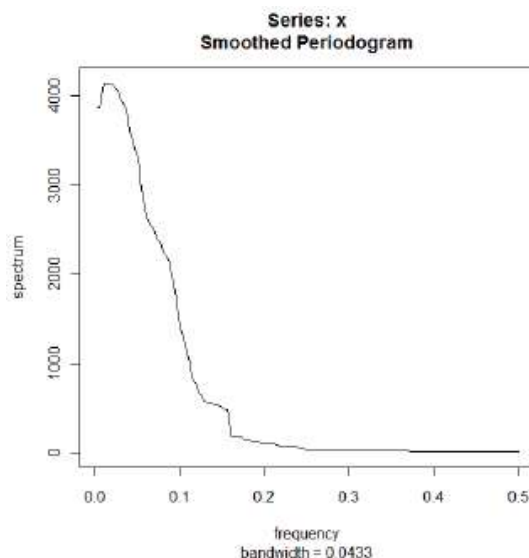
This uses two passes of a modified Daniell kernel with length $L = 13$ (so $m = 6$) each time. The plot is a bit smoother, but not by much. The peaks, by the way, are in exactly the same places as in the plot immediately above.



It's definitely possible to smooth too much. Suppose that we were to use a modified Daniell kernel of total length = 73 ($m = 36$). The command is

```
spec.pgram(x, spans=73, taper=0, log="no")
```

The result follows. The peaks are gone!

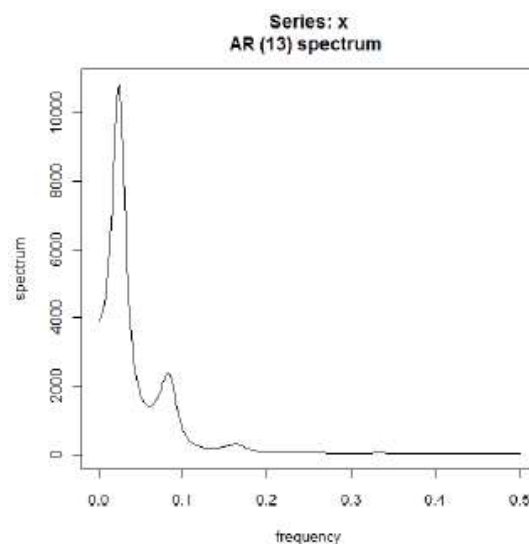


Parametric Estimation of the Spectral Density

The smoothing method of spectral density estimation is called a nonparametric method because it doesn't use any parametric model for the underlying time series process. An alternative method is a parametric method which entails finding the best fitting AR model for the series and then plotting the spectral density of that model. This method is supported by a theorem which says that the spectral density of any time series process can be approximated by the spectral density of an AR model (of some order, possibly a high one).

In R, parametric estimation of the spectral density is easily done with the command/function `spec.ar`. A command like `spec.ar(x, log="no")` will cause R to do all of the work. Again, to identify peaks we can assign a name to the `spec.ar` results by doing something like `specvalues=spec.ar(x, log="no")`.

For the fish recruitment example, the following plot is the result. Note that the density plotted is that of an AR(13) model. We can certainly find more parsimonious ARIMA models for these data. We're just using the spectral density of that model to approximate the spectral density of the observed series.



The appearance of the estimated spectral density is about the same as before. The estimated El Niño peak is located at a slightly different place – the frequency is about 0.024 for a cycle of about $1/0.024 = \text{about } 42 \text{ months}$.

De-trending

A series should be de-trended prior to a spectral analysis. A trend will cause such a dominant spectral density at a low frequency that other peaks won't be seen. By default, the R command `spec.pgram` performs a de-trending using a linear trend model. That is, the spectral density is estimated using the residuals from a regression done where the y-variable = observed data and the x-variable = t . If a different type of trend is present, a quadratic for instance, then a polynomial regression could be used to de-trend the data before the estimated spectral density is explored. Note, however, that the

R command `spec.ar`, however does not perform a de-trending by default.

Application of Smoothers to Raw Data

Note that the smoothers described here could also be applied to raw data. The Daniell kernel and its modifications are simply moving average (or weighted moving average) smoothers.

Source URL: <https://onlinecourses.science.psu.edu/stat510/node/57>