Tutorial 13: Spectral Analysis (I)

This tutorial deals with Spectral Analysis using R. We need the package 'TSA', which is installed using the command

```
> install.packages("TSA")
```

After the package is installed, click on 'TSA' under 'packages'. Type

```
> ?spectrum
```

to find out how the 'spectrum' function works.

Introductory

The following, by way of worked examples, illustrates the basic ideas of spectral analysis in R.

Firstly, generate a time series from cosine curves with two different frequencies:

```
> t <- 1:100
> cos1 <- cos(2*pi*t*5/100)
> cos2 <- cos(2*pi*t*15/100 + 0.2)
> plot(t,cos1,type="l",lty=1,lwd=2,'ylab'="cosines")
> lines(t,cos2,lty=2,lwd=2)
> y<-2*cos1 + 3*cos2
> plot(t,y,type='o')
```

The periodogram of y is found quite simply:

```
> periodogram(y)
```

The periodogram concentrates on the periods 0.05 and 0.15 as expected. To get a list of the frequencies at which the spectral densities are estimated, together wit the estimates, type

```
> cbind(periodogram(y)$freq,periodogram(y)$spec)
```

Leakage: if the periodicity is not at Fourier frequencies In practice, if the length of the sequence is not an exact multiple of a Fourier frequency, the periodogram may have two large adjacent frequencies, one on either side of the 'true' frequency. This can be helped by 'padding', a technique of adding in 0s. For example, if the observations are $1.5, 2.6, 3.7, ...$ this may be padded by inserting a fixed number of 0s between the observations, to get (for example) $1.5, 0, 0, 2.6, 0, 0, 3.7, 0, 0, \ldots$. This means that the periodogram frequencies are computed on a finer mesh and hence it is more likely that a strong frequency in the true spectrum will be represented by a single strong frequency in the periodogram.

```
> t<- 1:100
> cos1<-cos(2*pi*5.5/100)
> cos1<-cos(2*pi*t*5.5/100)
> cos2<-cos(2*pi*t*15.3/100 + 0.2)
> plot(t,cos1,type='l',lty=1,lwd=2,'ylab'="cosines")
> lines(t,cos2,lty=2)
> y<-2*cos1 + 3*cos2
> plot(t,y,type='o')
> periodogram(y)
> cbind(periodogram(y)$freq, periodogram(y)$spec)
```

Now see what happens when the series is padded with additional zeroes:

```
> spectrum(y,log="no",pad=2)
```

The default method for computing a 'spectrum' is the periodogram. Clearly, there are two well defined peaks.

Spectral Density of the ARMA processes The function `ARMAspec` computes the theoretical densities of the ARMA processes:

```
> ?ARMAspec
```

Now examine the spectral density of a MA process

```
> plot(arima.sim(model=list(ma=0.9),n=100))
> ARMAspec(model=list(ma=0.9))
```

Try with -0.9

```
> plot(arima.sim(model=list(ma=-0.9),n=100))
> ARMAspec(model=list(ma=-0.9))
```

Now try for an AR model:

```
> plot(arima.sim(model=list(ar=0.9),n=100))
> ARMAspec(model=list(ar=0.9))
> plot(arima.sim(model=list(ar=-0.6),n=100))
> ARMAspec(model=list(ar=-0.6))
```

and finally an ARMA(1,1):

```
> plot(arima.sim(model=list(ma = -0.8, ar=0.5),n=100))
> ARMAspec(model=list(ma = -0.8, ar=0.5))
```

Estimating the Spectrum when the data is White Noise When the data is white noise WN$(0, \sigma^2)$, the spectral density is

$$f(\lambda) = \frac{\sigma^2}{2\pi} \qquad \forall \lambda \in (-\pi, \pi].$$

Check that this is, at least approximately, what the estimate looks like.

```
> x <- rnorm(1000)
> sp <- spectrum(x, log="no")
> lines(sp$freq, spectrum(x, log="no",span=21,plot=FALSE)$spec, lwd=3)
```

In R, the default smoothing is the uniform kernel over span $= 2m + 1$ frequencies. It is possible to smooth in other ways, for example triangular smoothing. These involve convolutions.

```
> lines(sp$freq, spectrum(x, log="no",span= c(21,21),plot=FALSE)$spec, lwd=3)
> lines(sp$freq, spectrum(x, log="no",span= c(21,21,21),plot=FALSE)$spec, lwd=3)
```

The periodogram itself is not constant, but the smoothed periodogram looks more like the 'theoretical' spectral density.

More on the AR

```
> x <- arima.sim(model=list(ar=-0.5),n=200)
> sp<-spectrum(x,log="no")
> lines(sp$freq,ARMAspec(model=list(ar=-0.5),freq=sp$freq,plot=F)$spec,lwd=3)
```

Now try a smoothed kernel estimate:

```
> x <- arima.sim(model=list(ar=-0.5),n=1000)
> sp<-spectrum(x,log="no", span = 60)
> lines(sp$freq,ARMAspec(model=list(ar=-0.5),freq=sp$freq,plot=F)$spec,lwd=3)
```

Exercises

1. **Air Passenger Data** Make a spectral analysis of the Air Passenger data, using the following commands and interpret the output.

    ```
    > data(AirPassengers)
    > AP <- AirPassengers
    > plot(AP)
    > boxplot(AP~cycle(AP))
    > x <- diff(log(AP))
    > plot(x)
    > spectrum(as.matrix(x),log="no")
    > axis(1,at=c(1*12/length(x)))
    ```

2. The data for this exercise is contained in the file `SUNSPOTS.TSM` on the course page.

   First, find the average $\overline{x}$ and create the mean corrected series $Y_t = X_t - \overline{x}$. Fit an AR(2) model:

$$Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = Z_t \qquad \{Z_t\} \sim WN(0, \sigma^2).$$

   Plot the spectral density of the fitted model and find the frequency at which it achieves its maximum value. What is the corresponding period?

   Now consider the series $Z_t = (1 - B)^n X_t$. Plot these for $n = 1, 2, 3, 4$. Do the results correspond to the results of Exercise 3 Page 214?

3. The data for this exercise is found in `LAKE.TSM`. It consists of the lake levels in feet (reduced by 570) of Lake Huron for July of each year from 1875 to 1972 inclusive. In the class of ARIMA models, choose the model which fits the data best. Try

```
> install.packages("forecast")
```

   click on 'forecast' under 'packages' to activate it and investigate `auto.arima`.

   (a) Give approximate confidence bounds for the components of ϕ and θ and

   (b) Examine the residuals. Do they correspond to $\mathrm{WN}(0, \sigma^2)$?

   Now examine the spectral density for the Lake Huron data. Do this in two ways: firstly, from the data, build a periodogram, using techniques of padding or smoothing if they seem appropriate. Secondly, construct the spectral density for the fitted ARIMA model. Do the answers from the two approaches correspond?