

4.1: Introduction to Spectral Analysis

Many of the time series discussed in the previous chapters displayed strong periodic components: The sunspot numbers of Example 1.1.1, the number of trapped lynx of Example 1.1.2 and the Australian wine sales data of Example 1.4.1. Often, there is an obvious choice for the period d of this cyclical part such as an annual pattern in the wine sales. Given d , one could then proceed by removing the seasonal effects as in Section 1.4. In the first two examples it is, however, somewhat harder to determine the precise value of d . In this chapter, a general method is therefore discussed to deal with the periodic components of a time series. To complicate matters, it is usually the case that several cyclical patterns are simultaneously present in a time series. As an example recall the southern oscillation index (SOI) data which exhibits both an annual pattern and a so-called El Niño pattern.

The sine and cosine functions are the prototypes of periodic functions. They are going to be utilized here to describe cyclical behavior in time series. Before doing so, a *cycle* is defined to be one complete period of a sine or cosine function over a time interval of length 2π . Define also the *frequency*

$$\omega = \frac{1}{d}$$

as the number of cycles per observation, where d denotes the period of a time series (that is, the number of observations in a cycle). For monthly observations with an annual period, $d = 12$ and hence $\omega = 1/12 = 0.083$ cycles per observation. Now reconsider the process

$$X_t = R \sin(2\pi\omega t + \varphi)$$

as introduced in Example 1.2.2, using the convention $\lambda = 2\pi\omega$. To include randomness in this process, choose the amplitude R and the phase φ to be random variables. An equivalent representation of this process is given by

$$X_t = A \cos(2\pi\omega t) + B \sin(2\pi\omega t),$$

with $A = R \sin(\varphi)$ and $B = R \cos(\varphi)$ usually being independent standard normal variates. Then, $R^2 = A^2 + B^2$ is a χ -squared random variable with 2 degrees of freedom and $\varphi = \tan^{-1}(B/A)$ is uniformly distributed on $(-\pi, \pi]$. Moreover, R and φ are independent. Choosing now the value of ω one particular periodicity can be described. To accommodate more than one, it seems natural to consider mixtures of these periodic series with multiple frequencies and amplitudes:

$$X_t = \sum_{j=1}^m [A_j \cos(2\pi\omega_j t) + B_j \sin(2\pi\omega_j t)], \quad t \in \mathbb{Z},$$

where A_1, \dots, A_m and B_1, \dots, B_m are independent random variables with zero mean and variances $\sigma_1^2, \dots, \sigma_m^2$, and $\omega_1, \dots, \omega_m$ are *distinct* frequencies. It can be shown that $(X_t : t \in \mathbb{Z})$ is a weakly stationary process with lag- h ACVF

$$\gamma(h) = \sum_{j=1}^m \sigma_j^2 \cos(2\pi\omega_j h), \quad h \in \mathbb{Z}.$$

The latter result yields in particular that $\gamma(0) = \sigma_1^2 + \dots + \sigma_m^2$. The variance of X_t is consequently the sum of the component variances.

Example 4.1.1. Let $m = 2$ and choose $A_1 = B_1 = 1$, $A_2 = B_2 = 4$ to be constant as well as $\omega_1 = 1/12$ and $\omega_2 = 1/6$. This means that

$$X_t = X_t^{(1)} + X_t^{(2)} = [\cos(2\pi t/12) + \sin(2\pi t/12)] + [4 \cos(2\pi t/6) + 4 \sin(2\pi t/6)]$$

is the sum of two periodic components of which one exhibits an annual cycle and the other a cycle of six months. For all processes involved, realizations of $n = 48$ observations (4 years of data) are displayed in Figure 4.1. Also shown is a fourth time series plot which contains the X_t distorted by standard normal independent noise, \tilde{X}_t . The corresponding R code is:

```
>t=1:48
>x1=cos(2*pi*t/12)+sin(2*pi*t/12)
>x2=4*cos(2*pi*t/6)+4*sin(2*pi*t/6)
```

```
>x=x1+x2
>tildex=x+rnorm(48)
```

Note that the squared amplitude of $X_t^{(1)}$ is $1^2 + 1^2 = 2$. The maximum and minimum values of $X_t^{(1)}$ are therefore $\pm\sqrt{2}$. Similarly, we obtain $\pm\sqrt{32}$ for the second component.

For a statistician it is now important to develop tools to recover the periodicities from the data. The branch of statistics concerned with this problem is called spectral analysis. The standard method in this area is based on the *periodogram* which is introduced now. Suppose for the moment that the frequency parameter $\omega_1 = 1/12$ in Example 4.1.1 is known. To obtain estimates of A_1 and B_1 , one could try to run a regression using the explanatory variables $Y_{t,1} = \cos(2\pi t/12)$ or $Y_{t,2} = \sin(2\pi t/12)$ to compute the least squares estimators

$$\hat{A}_1 = \frac{\sum_{t=1}^n X_t Y_{t,1}}{\sum_{t=1}^n Y_{t,1}^2} = \frac{2}{n} \sum_{t=1}^n X_t \cos(2\pi t/12),$$

$$\hat{B}_1 = \frac{\sum_{t=1}^n X_t Y_{t,2}}{\sum_{t=1}^n Y_{t,2}^2} = \frac{2}{n} \sum_{t=1}^n X_t \sin(2\pi t/12).$$

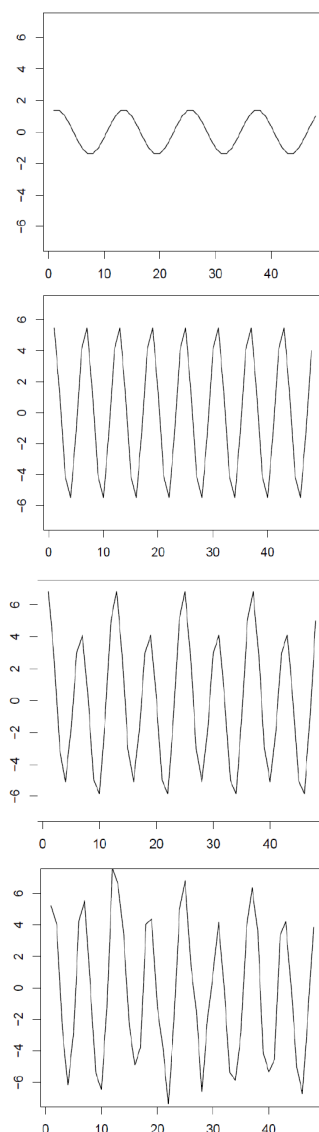


Figure 4.1: Time series plots of $(X_t^{(1)})$, $(X_t^{(2)})$, (X_t) and \tilde{X}_t

Since, in general, the frequencies involved will not be known to the statistician prior to the data analysis, the foregoing suggests to pick a number of potential ω 's, say j/n for $j = 1, \dots, n/2$ and to run a long regression of the form

$$X_t = \sum_{j=0}^{n/2} [A_j \cos(2\pi jt/n) + B_j \sin(2\pi jt/n)]. \quad (4.1.1)$$

This leads to least squares estimates \hat{A}_j and \hat{B}_j of which the "significant" ones should be selected. Note that the regression in 4.1.1 is a perfect one because there are as many unknowns as variables! Note also that

$$P(j/n) = \hat{A}_j^2 + \hat{B}_j^2$$

is essentially (up to a normalization) an estimator for the correlation between the time series X_t and the corresponding sum of the periodic cosine and sine functions at frequency j/n . The collection of all $P(j/n)$, $j = 1, \dots, n/2$, is called the *scaled periodogram*. It can be computed quickly via an algorithm known as the fast Fourier transform (FFT) which in turn is based on the discrete Fourier transform (DFT)

$$d(j/n) = \frac{1}{\sqrt{n}} \sum_{t=1}^n X_t \exp(-2\pi i jt/n).$$

The frequencies j/n are called the Fourier or fundamental frequencies. Since $\exp(-ix) = \cos(x) - i \sin(x)$ and $|z|^2 = z\bar{z} = (a + ib)(a - ib) = a^2 + b^2$ for any complex number $z = a + ib$, it follows that

$$I(j/n) = |d(j/n)|^2 = \frac{1}{n} \left(\sum_{t=1}^n X_t \cos(2\pi jt/n) \right)^2 + \frac{1}{n} \left(\sum_{t=1}^n X_t \sin(2\pi jt/n) \right)^2.$$

The quantity $I(j/n)$ is referred to as the *periodogram*. It follows immediately that the periodogram and the scaled periodogram are related via the identity $4I(j/n) = nP(j/n)$.

Example 4.1.2. Using the expressions and notations of Example 4.1.1, the periodogram and the scaled periodogram are computed in R as follows:

```
>t=1:48
>l=abs(fft(x)/sqrt(48))^2
>P=4*I/48
>f=0:24/48
>plot(f, P[1:25], type="l")
>abline(v=1/12)
>abline(v=1/6)
```

The corresponding (scaled) periodogram for (\tilde{X}_t) can be obtained in a similar fashion. The scaled periodograms are shown in the left and middle panel of Figure 4.2. The right panel displays the scaled periodogram of another version of (\tilde{X}_t) in which the standard normal noise has been replaced with normal noise with variance 9. From these plots it can be seen that the six months periodicity is clearly visible in the graphs (see the dashed vertical lines at $x=1/6$. The less pronounced annual cycle (vertical line at $x=1/12$ is still visible in the first two scaled periodograms but is lost if the noise variance is increased as in the right plot. Note, however, that the y-scale is different for all three plots.

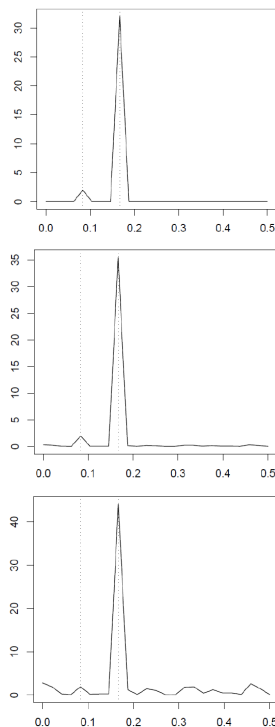


Figure 4.2: The scaled periodograms of (X_t) , $(\tilde{X}_t^{(1)})$, $(\tilde{X}_t^{(2)})$

In the ideal situation that we observe the periodic component without additional contamination by noise, we can furthermore see why the periodogram may be useful in uncovering the variance decomposition from above. We have shown in the lines preceding Example 4.1.1 that the squared amplitudes of $X_t^{(1)}$ and $X_t^{(2)}$ are 2 and 32, respectively. These values are readily read from the scaled periodogram in the left panel of Figure 4.2. The contamination with noise alters these values.

In the next section, it is established that the time domain approach (based on properties of the ACVF, that is, regression on past values of the time series) and the frequency domain approach (using a periodic function approach via fundamental frequencies, that is, regression on sine and cosine functions) are equivalent. Some details are given on the spectral density (the population counterpart of the periodogram) and on properties of the periodogram itself.

Contributors and Attributions

- Alexander Aue (Department of Statistics, University of California, Davis)
- Integrated by Brett Nakano (statistics, UC Davis)

This page titled [4.1: Introduction to Spectral Analysis](#) is shared under a [not declared](#) license and was authored, remixed, and/or curated by [Alexander Aue](#).