Time series modelling

B.Hadj-Amar, N. Cunningham, S.Ip

March 8, 2016

Abstract

1 Spectral Analysis and Harmonic Regression

1.1 Methodology

Let us consider the periodic model

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

where Z_t is white noise, A is the amplitude, ϕ is a phase shift that determine the initial point of the cosine function and ω is a fixed frequency. Using trigonometric identities we rewrite X_t:

$$X_t = \beta_1\cos(2\pi\omega t) + \beta_2\sin(2\pi\omega t) + Z_t, \tag{1}$$

where $\beta_1 = A\cos(\phi)$ and $\beta_2 = -A\sin(\phi)$.

We use the periodogram as a tool to pick the frequencies that drive the data; then we estimate for each frequency, its corresponding regressions coefficients by using least squares estimates; finally, we can easily obtain the estimate of the amplitude A and the phase shift ϕ:

$$A = \sqrt{\hat{\beta}_1^2 + \hat{\beta}_2^2}, \qquad \phi = \tan^{-1}\Big(\frac{\hat{\beta}_2}{\hat{\beta}_1}\Big) \tag{2}$$

Measure the squared correlation of the data with sines and cosine oscillating at frequency $\omega_j = j/T$

As it discussed in [1], the *Spectral Representation Theorem* states that any weakly stationary time series can be approximated as the sum of sinusoidal waveforms oscillating at different frequencies:

$$X_t = \mu + \sum_{k=1}^{K} \big\{\beta_{k1}\cos(2\pi\omega_k t) + \beta_{k2}\sin(2\pi\omega_k t)\big\}, \tag{3}$$

where $\{\omega_k\}_{k=1}^{K}$ is a collection of distinct fixed frequencies.

In this scenario, least squares estimates from this regression can be effectively obtained by associating the data to the frequency domain via the Discrete Fourier Transform (DFT), and then obtaining a statistic called the *periodogram*, which is an unbiased estimator of the *spectral density function.*

1.2 Application

We now apply the techniques summarised above on a dataset consisting of two different measurements of 4 healthy patients (whose ID are 2, 8, 24, 26). These observations are temperature and rest activity.

Temperature

We initially focus on the temperature of patient 8, whose time series is given in Figure 1.

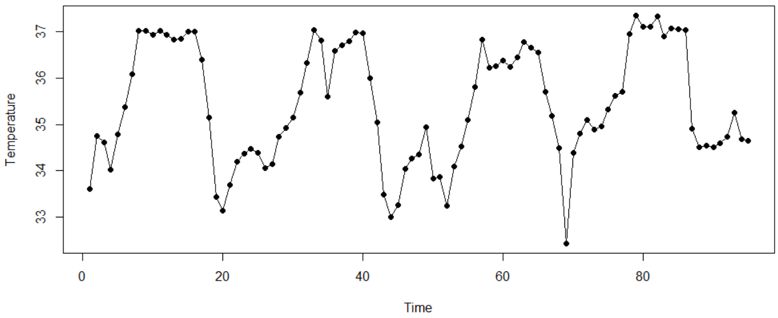


Figure 1: Time series of the temperature of patient 8.

The periodicity of this time series suggests applying an harmonic regression to fit the data. The first step for working on spectral analysis is considering a zero mean process. Therefore, we performed our analysis on the residuals, by subtracting the trend.

In order to find the frequencies that drive the data, we explored the periodogram; notice that it is recommended to investigate behaviours of the smoothed periodogram, which is a *consistent* estimator of the spectral density function.

In Figure 2, we show the periodogram for this time series, and two different smoothing version of it, using uniform weights and *Daniell* weights. We can notice how the main frequency which drives the data is around 0.04, i.e 1/24 hours.

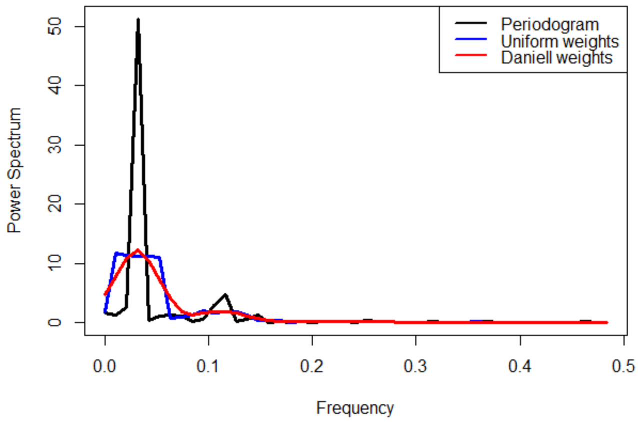
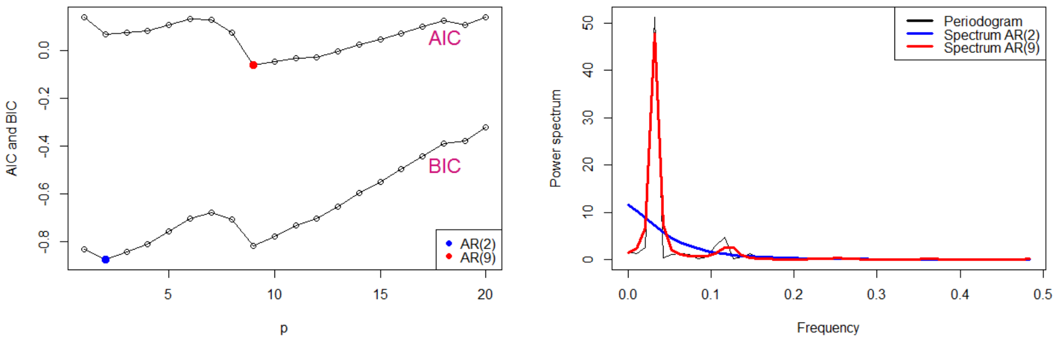


Figure 2: Periodogram and smoothed periodograms (uniform and Daniell weights) for temperature of patient 8. This figure is best viewed in colours.

Furthermore, it can be shown that the spectrum of any (weakly) stationary time series can be approximated through the spectrum of an AR(p) model, whose explicit form is known. In order to find the *best* lag parameter p we fit several AR(p) to the data, for increasing values of p. We then calculated two measures of quality of the model, BIC and AIC; the smallest value for BIC or AIC is a good suggestion for choosing the value of p. We give in Figure 3 an idea of this procedure to find the value of p; hence, we show the spectrum of an AR(2) and AR(9).



(a) AIC and BIC (b) Spectrum AR(2), AR(9)

Figure 3: AR(p) approach to obtain the correct frequency representation. Main frequencies that drive the data are around 1/24, and 1/8

Once the correct frequencies are selected, we can apply the harmonics regression explained in Section 1.1. The resulting model is given in Figure 4.

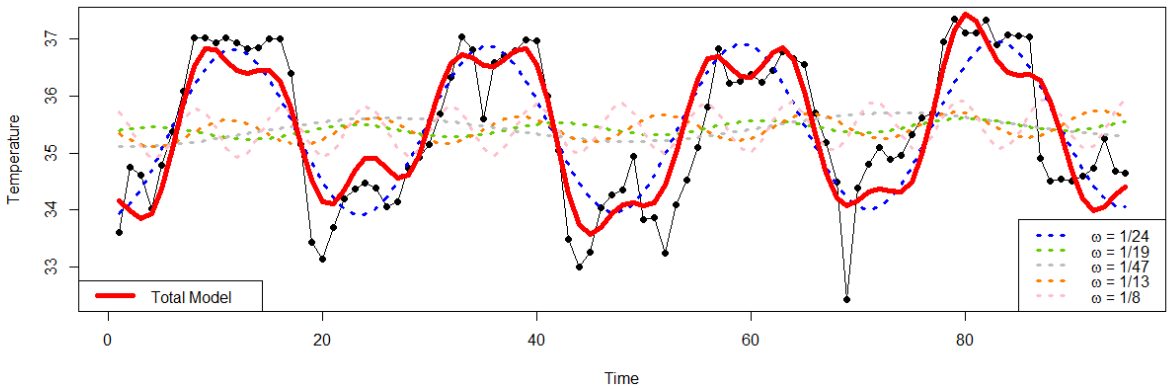


Figure 4: Harmonic regression for temperature of patient 8, using 5 different harmonics. Dotted lines represent the fitting for single harmonics; thick red line is the superposition of the dotted harmonics, which is the final fit.