Spring 2015 Statistics 153 (Time Series) : Lecture Twenty

Aditya Guntuboyina

09 April 2015

1 Frequency Domain Analysis of Time Series

1.1 The Sinusoid

The sinusoid can be represented in the following three equivalent ways:

1. $R\cos(2\pi ft + \Phi)$. The following terminology is standard. R is called the *amplitude*, f is called the *frequency* and Φ is called the *phase*. The quantity $1/f$ is called the *period* and $2\pi f$ is termed the *angular frequency*.
2. The sinusoid can also be written as $A\cos 2\pi ft + B\sin 2\pi ft$ where $A = R\cos\Phi$ and $B = R\sin\Phi$.
3. Yet another way of representing the sinusoid is to use complex exponentials:
$$\exp(2\pi ift) = \cos(2\pi ft) + i\sin(2\pi ft).$$
Therefore
$$\cos(2\pi ft) = \frac{\exp(2\pi ift) + \exp(-2\pi ift)}{2} \text{ and } \sin(2\pi ft) = \frac{\exp(2\pi ift) - \exp(-2\pi ift)}{2i}.$$
Thus $A\cos 2\pi ft + B\sin 2\pi ft$ can also be written as a linear combination of $\exp(2\pi ift)$ and $\exp(-2\pi ift)$.

1.2 The two most important properties of Sinusoids

Frequency domain analysis of time series is all based on the following two most important properties of sinusoids:

1. Every data set $x_0, \dots, x_{n-1}$ can be represented via sinusoids. We call this data representation.
2. Every stationary process $X_t, t = \dots, -2, -1, 0, 1, 2, \dots$ can be represented via sinusoids. We call this stationary process representation.

1.3 Data Representation

Let us first focus on data representation. We will study stationary process representation in days to come.

Suppose we have a dataset $x_0, \dots, x_{n-1}$ of size n. We can represent this dataset via a vector $(x_0, \dots, x_{n-1})$ of size n. How does one represent this via sinusoids? We will use what are called *Fourier*

Frequencies. A frequency f of the form $f = j/n$ for $0 \leq j \leq (n-1)$ is called a Fourier Frequency. There are n Fourier frequencies (n here is the size of the dataset): $0, 1/n, \dots, (n-1)/n$. Given a Fourier frequency j/n, let us define the vector

$$u^j = (1, \exp(2\pi i j/n), \exp(2\pi i 2j/n), \dots, \exp(2\pi i(n-1)j/n)).$$

This is the sinusoid with frequency j/n evaluated at the time points $t = 0, 1, \dots, (n-1)$.

It turns out that every vector $(x_0, \dots, x_{n-1})$ can be written as a linear combination of $u^0, u^1, \dots, u^n$. This is what we called data representation. The proof of this fact is quite simple and uses basic linear algebra. Indeed one only needs to observe that $u^0, u^1, u^2, \dots, u^{n-1}$ are orthogonal i.e., the dot product between u^k and u^l is zero if $k \neq l$. Recall that the dot product between two complex-valued vectors $(a_1, \dots, a_n)$ and $(b_1, \dots, b_n)$ equals

$$\langle a, b \rangle = \sum_{j=1}^{n} a_j \bar{b_j}.$$

Can you prove that the dot product between u^k and u^l is zero for $k \neq l$?

2 The Discrete Fourier Transform

The fact that every vector $(x_0, \dots, x_{n-1})$ can be written as a linear combination of $u^0, u^1, \dots, u^n$ motivates the definition of the Discrete Fourier Transform (DFT).

The DFT of $x_0, \dots, x_{n-1}$ is given by $b_j, j = 0, 1, \dots, n-1$, where

$$b_j = \sum_{t=0}^{n-1} x_t \exp\left(-\frac{2\pi i j t}{n}\right) \qquad \text{for } j = 0, \dots, n-1.$$

Therefore, $b_0 = \sum x_t$. Also for $1 \leq j \leq n-1$,

$$b_{n-j} = \sum_t x_t \exp\left(-\frac{2\pi i(n-j)t}{n}\right) = \sum_t x_t \exp\left(\frac{2\pi i j t}{n}\right) \exp(-2\pi i t) = \bar{b}_j.$$

Note that this relation only holds if $x_0, \dots, x_{n-1}$ are real numbers.

Thus for $n = 11$, the DFT can be written as:

$$b_0, b_1, b_2, b_3, b_4, b_5, \bar{b}_5, \bar{b}_4, \bar{b}_3, \bar{b}_2, \bar{b}_1.$$

and for $n = 12$, it is

$$b_0, b_1, b_2, b_3, b_4, b_5, b_6 = \bar{b}_6, \bar{b}_5, \bar{b}_4, \bar{b}_3, \bar{b}_2, \bar{b}_1.$$

Note that b_6 is necessarily real because $b_6 = \bar{b}_6$.

The DFT is calculated by the R function `fft()`.

The original data $x_0, \dots, x_{n-1}$ can be recovered from the DFT using:

$$x_t = \frac{1}{n} \sum_{j=0}^{n-1} b_j \exp\left(\frac{2\pi i j t}{n}\right) \qquad \text{for } t = 0, \dots, n-1.$$

Thus for $n = 11$, one can think of the data as the 11 real numbers $x_0, x_1, \dots, x_{10}$ or, equivalently, as one real number b_0 along with 5 complex numbers $b_1, \dots, b_5$.

For $n = 12$, one can think of the data as the 12 real numbers $x_0, \dots, x_{11}$ or, equivalenlty, as two real numbers, b_0 and b_6, along with the 5 complex numbers $b_1, \dots, b_5$.