Spring 2015 Statistics 153 (Time Series) : Lecture Twenty Two

Aditya Guntuboyina

16 April 2015

1 DFT Again

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

Given a data vector $(x_0, \ldots, x_{n-1})$ of length n, its DFT is given by $b_j, j = 0, 1, \ldots, n-1$, where

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

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

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

Two basic facts about the DFT are $b_0 = x_0 + \cdots + x_{n-1}$ and $b_{n-j} = \bar{b}_j$ for all $1 \le j \le n-1$. In R, the DFT is calculated by the function `fft()`.

2 DFT for Sinusoids

To understand the DFT, let us calculate the DFT of the cosine wave $x_t = R\cos(2\pi f_0 t + \phi), t = 0, 1, \ldots, n-1$. In the last class, we saw that we can without loss of generality take $0 \le f_0 \le 1/2$.

2.1 When f_0 is a Fourier frequency

Frequencies of the form k/n are called Fourier frequencies.

Suppose f_0 is of the form k/n for some k where $0 \le k/n \le 1/2$. Then the DFT is given by

$$\begin{aligned} b_j &= \sum_{t=0}^{n-1} R\cos(2\pi(k/n)t + \Phi)\exp\left(-2\pi i(j/n)t\right) \\ &= \frac{Re^{i\Phi}}{2}\sum_{t=0}^{n-1}\exp\left(2\pi it\frac{k-j}{n}\right) + \frac{Re^{-i\Phi}}{2}\sum_{t=0}^{n-1}\exp\left(-2\pi it\frac{j+k}{n}\right). \end{aligned}$$

Note that we do not need to consider the DFT b_j for $j/n > 1/2$. So we assume that $0 \le j/n \le 1/2$. Because the original cosine wave was assumed to have frequency in the range $[0, 1/2]$, we have $0 \le k/n \le 1/2$.

We need to consider the case $j = k$ and $j \neq k$ here. When $j \neq k$, both terms above are zero (because of the geometric sum formula).

When $j = k \neq n/2$, the second term is zero while the first term equals $Re^{i\Phi}n/2$. When $j = k = n/2$ (this can only happen if n is even), we get $b_k = nR(e^{i\Phi} + e^{-i\Phi})/2 = nR\cos\Phi$.

The periodogram $I(j/n)$ therefore equals $nR^2/4$ for $j = k \neq n/2$ and $nR^2\cos^2\Phi$ for $j = k = n/2$. But whenever $j \neq k$ and $j \in \{0, 1, \ldots, n/2\}$, we have $I(j/n) = 0$.

The key is to observe that b_j equals zero when $j \neq k$.

2.2 Multiple Fourier Frequencies

Now consider data that is linear combination of multiple Fourier frequencies:

$$x_t = \sum_{l=1}^{m} R_l \cos(2\pi t(k_l/n) + \Phi_l) \tag{1}$$

where each k_l is an integer satisfying $0 \leq k_l/n \leq 1/2$. Because the definition of the DFT is linear in the data $\{x_t\}$, it follows that the DFT of (1) is given by

$$b_j = \begin{cases} nR_l e^{i\Phi_l}/2 & \text{if } j = k_l \neq n/2 \\ nR_l \cos\Phi_l & \text{if } j = k_l = n/2 \\ 0 & \text{otherwise} \end{cases}$$

for $0 \leq j/n \leq 1/2$. Similarly

$$I(j/n) = \begin{cases} nR_l^2/2 & \text{if } j = k_l \neq n/2 \\ nR_l^2\cos^2\Phi_l & \text{if } j = k_l = n/2 \\ 0 & \text{otherwise} \end{cases}$$

This shows that the DFT picks out the frequencies present in the data. The strength (absolute value) of the DFT at a frequency is proportional to the amplitude (R_l) of the cosine wave at that frequency.

2.3 DFT of a sinusoid at a non-Fourier frequency

The DFT of a sinusoid at a non-Fourier frequency is calculated in the following way: Consider the signal $x_t = e^{2\pi f_0 t}$ where $f_0 \in [0, 1/2]$ is not necessarily of the form k/n for any k. Its DFT is given by

$$b_j := \sum_{t=0}^{n-1} x_t e^{-2\pi i t(j/n)} = \sum_{t=0}^{n-1} e^{2\pi i(f_0 - (j/n))t}.$$

If we denote the function

$$S_n(g) := \sum_{t=0}^{n-1} e^{2\pi i g t} \tag{2}$$

then we can write

$$b_j = S_n(f_0 - (j/n)).$$

The function $S_n(g)$ can clearly be evaluated using the geometric series formula to be

$$S_n(g) = \frac{e^{2\pi i g n} - 1}{e^{2\pi i g} - 1}$$

Because

$$e^{i\theta} - 1 = \cos\theta + i\sin\theta - 1 = 2e^{i\theta/2}\sin\theta/2,$$

we get

$$S_n(g) = \frac{\sin \pi n g}{\sin \pi g} e^{i\pi g(n-1)}$$

Thus the absolute value of the DFT of $y_t = e^{2\pi i f_0 t}$ is given by

$$|b_j| = |S_n(f_0 - (j/n))| = \left|\frac{\sin \pi n(f_0 - (j/n))}{\sin \pi(f_0 - (j/n)))}\right|$$

This expression becomes meaningless when $f_0 = j/n$ but since we are assuming that f_0 is not a Fourier frequency, we do not need to worry about this.

The behavior of this DFT can be best understood by plotting the function $g \mapsto (\sin \pi n g)/(\sin \pi g)$.

When f_0 is not of the form k/n for any k, the term $S_n(f_0 - j/n)$ is non-zero for all j. This situation when one observes a non-zero DFT term $_j$ because of the presence of a sinusoid at a frequency f_0 different from j/n is referred to as **Leakage**. Leakage due to a sinusoid with frequency f_0 not of the form k/n is present in all DFT terms b_j but the magnitude of the presence decays as j/n gets far from f_0 (this is due to the form of the function $S_n(f_0 - j/n)$).

The behavior of the DFT of the cosine wave $x_t = R\cos(2\pi f_0 t + \Phi)$ can be understood from the DFT of $x_t = e^{2\pi f_0 t}$.

3 Interpreting the Periodogram

The DFT writes the given data in terms of sinusoids with frequencies of the form k/n. Frequencies of the form k/n are called Fourier frequencies.

Suppose that we are given a dataset $x_0, \dots, x_{n-1}$. We have calculated its DFT: $b_0, b_1, \dots, b_{n-1}$ and we have plotted $|b_j|^2/n$ for $j = 1, \dots, (n-1)/2$ for odd n and for $j = 1, \dots, n/2$ for even n.

If we see a single spike in this plot, say at b_k, we are sure that the data is a sinusoid with frequency k/n.

If we get two spikes, say at b_{k_1} and b_{k_2}, then the data is slightly more complicated: it is a linear combination of two sinusoids at frequencies k_1/n and k_2/n with the strengths of these sinusoids depending on the size of the spikes.

Multiple spikes indicate that the data is made up of many sinusoids at Fourier frequencies and, in general, this means that the data is more complicated.

However, sometimes one can see multiple spikes in the periodogram even when the structure of the data is not very complicated. A typical example is leakage due to the presence of a sinusoid at a non-Fourier frequency.

4 Periodogram and the Sample Autocovariance Function

The periodogram and the sample autocovariance function are connected by the following formula:

$$I(j/n) = \frac{|b_j|^2}{n} = \sum_{|h|<n} \hat{\gamma}(h) \exp\left(-\frac{2\pi i j h}{n}\right) \qquad \text{for } j = 1, \dots, n-1 \tag{3}$$

where $\hat{\gamma}(h)$ is the sample autocovariance function. We shall prove this in the next class. This is a rather important formula. It states that the periodogram can be computed directly from the sample autocovariance function. What happens if we replace $\hat{\gamma}(h)$ by the true autocovariance function $\gamma(h)$ (assuming that the data come from a stationary process with autocovariance $\gamma(h)$)? We shall come back to this question later.

5 Process Representation

Next, we shall study process representation which says that every stationary process can be written in terms of sines and cosines.

The simplest stationary model using sines and cosines is

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

where $0 \le f \le 1/2$ is a fixed constant and A and B are uncorrelated random variables with mean 0 and variance σ^2. It is not hard to check that $\{X_t\}$ is stationary with mean 0 and variance σ^2.

More complicated stationary models with sines and cosines can be constructed by taking linear combinations of the form

$$X_t = \sum_{j=1}^{m} \left(A_j \cos(2\pi f_j t) + B_j \sin(2\pi f_j t)\right) \tag{4}$$

where $0 \le f_j \le 1/2$ are fixed constants and $A_1, B_1, A_2, B_2, \dots, A_m, B_m$ are all uncorrelated random variables with mean zero and

$$\text{var}(A_j) = \text{var}(B_j) = \sigma_j^2.$$

Let $\sum_{j=1}^{m} \sigma_j^2 = \sigma^2$ so that the variance of the process $\{X_t\}$ equals σ^2.

It turns out that the model (4) can approximate any stationary model provided m is large enough and $\lambda_1, \dots, \lambda_m$ and $\sigma_1^2, \dots, \sigma_m^2$ are chosen appropriately. For example, the choices

$$\lambda_j = \frac{j}{2m} \text{ and } \sigma_j^2 = \frac{\sigma^2}{m} \qquad \text{for } j = 1, \dots, m$$

for m large lead to a very good approximation of the white noise model.