Spring 2013 Statistics 153 (Time Series) : Lecture Nineteen

Aditya Guntuboyina

4 April 2013

1 DFT Recap

Given data $x_0, \dots, x_{n-1}$, their DFT is given by $b_0, b_1, \dots, b_{n-1}$ where

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

Two key things to remember are:

1. $b_0 = x_0 + \dots + x_{n-1}$
2. $b_{n-j} = \bar{b}_j$ for $1 \le j \le n-1$

For odd values of n, say $n = 11$, the DFT is comprised of the real number b_0 and the $(n-1)/2$ complex numbers $b_1, \dots, b_{(n-1)/2}$.

For even values of n, say $n = 12$, the DFT consists of two real numbers b_0 and $b_{n/2}$ and the $(n-2)/2$ complex numbers $b_1, \dots, b_{(n-2)/2}$.

The original data $x_0, x_1, \dots, x_{n-1}$ can be recovered from the DFT by the formula:

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

This formula can be written succintly as:

$$x = \frac{1}{n} \sum_{j=0}^{n-1} b_j u^j$$

where $x = (x_0, \dots, x_{n-1})$ denotes the data vector and u^j denotes the vector obtained by evaluating the sinusoid $\exp(2\pi ijt/n)$ at times $t = 0, 1, \dots, n-1$. We have seen in the last class that $u^0, u^1, \dots, u^{n-1}$ are orthogonal with $u^j \cdot u^j = n$ for each j.

As a result, we have the sum of squares identity:

$$n \sum_t x_t^2 = \sum_{j=0}^{n-1} |b_j|^2 .$$

The absolute values of b_j and b_{n-j} are equal because $b_{n-j} = \bar{b}_j$ and hence we can write the above sum of squares identity in the following way. For $n = 11$,

$$n \sum_t x_t^2 = b_0^2 + 2|b_1|^2 + 2|b_2|^2 + 2|b_3|^2 + 2|b_4|^2 + 2|b_5|^2$$

and for $n = 12$,

$$n \sum_t x_t^2 = b_0^2 + 2|b_1|^2 + 2|b_2|^2 + 2|b_3|^2 + 2|b_4|^2 + 2|b_5|^2 + b_6^2.$$

Note that there is no need to put an absolute value on b_6 because it is real.

Because $b_0 = \sum_t x_t = n\bar{x}$, we have

$$n \sum_t x_t^2 - b_0^2 = n \sum_t x_t^2 - n^2 \bar{x}^2 = n \sum_t (x_t - \bar{x})^2.$$

Thus the sum of squares identity can be written for n odd, say $n = 11$, as

$$\sum_t (x_t - \bar{x})^2 = \frac{2}{n}|b_1|^2 + \frac{2}{n}|b_2|^2 + \frac{2}{n}|b_3|^2 + \frac{2}{n}|b_4|^2 + \frac{2}{n}|b_5|^2$$

and, for n even, say $n = 12$, as

$$\sum_t (x_t - \bar{x})^2 = \frac{2}{n}|b_1|^2 + \frac{2}{n}|b_2|^2 + \frac{2}{n}|b_3|^2 + \frac{2}{n}|b_4|^2 + \frac{2}{n}|b_5|^2 + \frac{1}{n} b_6^2.$$

2 DFT of the Cosine Wave

Let $x_t = R\cos(2\pi f_0 t + \phi)$ for $t = 0, \ldots, n-1$. We have seen in R that when f_0 is a Fourier frequency (i.e., of the form k/n for some k), the DFT has exactly one spike but when f_0 is not a Fourier frequency, there is leakage. We prove this here.

We can, without loss of generality, assume that $0 \le f_0 \le 1/2$ because:

1. If $f_0 < 0$, then we can write $\cos(2\pi f_0 t + \phi) = \cos(2\pi(-f_0)t - \phi)$. Clearly, $-f_0 \ge 0$.
2. If $f_0 \ge 1$, then we write

   $$\cos(2\pi f_0 t + \phi) = \cos(2\pi[f_0]t + 2\pi(f - [f_0])t + \phi) = \cos(2\pi(f - [f_0])t + \phi),$$

   because $\cos(\cdot)$ is periodic with period 2π. Clearly $0 \le f - [f_0] < 1$.
3. If $f_0 \in [1/2, 1)$, then

   $$\cos(2\pi f_0 t + \phi) = \cos(2\pi t - 2\pi(1 - f_0)t + \phi) = \cos(2\pi(1 - f_0)t - \phi)$$

   because $\cos(2\pi t - x) = \cos x$ for all integers t. Clearly $0 < 1 - f_0 \le 1/2$.

Thus given a cosine wave $R\cos(2\pi f t + \phi)$, one can always write it as $R\cos(2\pi f_0 t + \phi')$ with $0 \le f_0 \le 1/2$ and a phase ϕ' that is possibly different from ϕ. This frequency f_0 is said to be an **alias** of f. From now on, whenever we speak of the cosine wave $R\cos(2\pi f_0 t + \phi)$, we assume that $0 \le f_0 \le 1/2$.

If $\phi = 0$, then we have $x_t = R\cos(2\pi f_0 t)$. When $f_0 = 0$, then $x_t = R$ and so there is no oscillation in the data at all. When $f_0 = 1/2$, then $x_t = R\cos(\pi t) = R(-1)^t$ and so $f_0 = 1/2$ corresponds to the maximum possible oscillation.

What is the DFT of $x_t = R\cos(2\pi f_0 t + \phi)$ for $0 \le f_0 \le 1/2$? The formula is

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

Suppose $f = j/n$ and we shall calculate

$$b(f) = \sum_{t=0}^{n-1} x_t \exp\left(-2\pi i f t\right).$$

The easiest way to calculate this DFT is to write the cosine wave in terms of complex exponentials:

$$x_t = \frac{R}{2}\left(e^{2\pi i f_0 t} e^{i\phi} + e^{-2\pi i f_0 t} e^{-i\phi}\right).$$

It is therefore convenient to first calculate the DFT of the complex exponential $e^{2\pi i f_0 t}$.

2.1 DFT of $y_t = e^{2\pi i f_0 t}$

The DFT of $y_t = e^{2\pi i f_0 t}$ is given by

$$\sum_{t=0}^{n-1} y_t e^{-2\pi i f t} = \sum_{t=0}^{n-1} e^{2\pi i (f_0 - f) t}$$

where $f = j/n$. Let us denote this by $S_n(f_0 - f)$ i.e.,

$$S_n(f_0 - f) = \sum_{t=0}^{n-1} e^{2\pi i (f_0 - f) t}. \tag{1}$$

This can clearly be evaluated using the geometric series formula to be

$$S_n(f_0 - f) = \frac{e^{2\pi i (f_0 - f) n} - 1}{e^{2\pi i (f_0 - f)} - 1}$$

It is easy to check that

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

As a result

$$S_n(f_0 - f) = \frac{\sin \pi n (f_0 - f)}{\sin \pi (f_0 - f)} e^{i\pi (f_0 - f)(n-1)}$$

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

$$|S_n(f - f_0)| = \left| \frac{\sin \pi n (f_0 - f)}{\sin \pi (f_0 - f))} \right| \qquad \text{where } f = j/n$$

This expression becomes meaningless when $f_0 = f$. But when $f_0 = f$, the value of $S_n(f_0 - f)$ can be directly be calculated from (1) to be equal to n.

The behavior of $|S_n(f - f_0)|$ can be best understood by plotting the function $g \mapsto (\sin \pi n g)/(\sin \pi g)$. This explains leakage.

The behavior of the DFT of the cosine wave can be studied by writing it in terms of the DFT of the complex exponential.

If f_0 is not of the form k/n for any j, then the term $S_n(f - f_0)$ is non-zero for all f of the form j/n. This situation where one observes a non-zero DFT term b_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 because of the form of the function $S_n(f - f_0)$.

There are two problems with Leakage:

1. Fourier analysis is typically used to separate out the effects due to different frequencies; so leakage is an undesirable phenomenon.

2. Leakage at j/n due to a sinusoid at frequency f_0 can mask the presence of a true sinusoid at frequency j/n.

How to get rid of leakage? The easy answer is to choose n appropriately (ideally, n should be a multiple of the periods of all oscillations). For example, if it is monthly data, then it is better to have whole year's worth of data. But this is not always possible. We will study a leakage-reducing technique later.

3 DFT of a Periodic Series

Suppose that the data $x_0, x_1, \dots, x_{n-1}$ is periodic with period h i.e., $x_{t+hu} = x_t$ for all integers t and u. Let n be an integer multiple of h i.e., $n = kh$. For example, suppose we have monthly data collected over 10 years in which case: $h = 12$, $k = 10$ and $n = 120$.

Suppose that DFT of the data $x_0, \dots, x_{n-1}$ is $b_0, b_1, \dots, b_{n-1}$. Suppose also that the DFT of the data in the first cycle: $x_0, x_1, \dots, x_{h-1}$ is $\beta_0, \beta_1, \dots, \beta_{h-1}$.

We shall express b_j in terms of $\beta_0, \dots, \beta_{h-1}$. Let $f = j/n$ for simplicity.

By definition

$$b_j = \sum_{t=0}^{n-1} x_t \exp\left(-2\pi i t f\right).$$

Break up the sum into

$$\sum_{t=0}^{h-1} + \sum_{t=h}^{2h-1} + \dots + \sum_{t=(k-1)h}^{kh-1}$$

The lth term above can be evaluated as:

$$\begin{aligned}\sum_{t=(l-1)h}^{lh-1} x_t \exp\left(-2\pi i f t\right) &= \sum_{s=0}^{h-1} x_s \exp\left(-2\pi i f (s + (l-1)h)\right) \\ &= \exp\left(-2\pi i f (l-1)h\right) \sum_{s=0}^{h-1} x_s \exp(-2\pi i f s).\end{aligned}$$

Therefore

$$
\begin{aligned}
b_j &= \sum_{l=1}^{k} \exp\left(-2\pi i f(l-1)h\right) \sum_{s=0}^{h-1} x_s \exp(-2\pi i f s) \\
&= \sum_{s=0}^{h-1} x_s \exp(-2\pi i f s) \sum_{l=1}^{k} \exp\left(-2\pi i f(l-1)h\right) \\
&= S_k(fh) \sum_{s=0}^{h-1} x_s \exp(-2\pi i f s) \\
&= S_k(jh/n) \sum_{s=0}^{h-1} x_s \exp(-2\pi i j s/n) \\
&= S_k(j/k) \sum_{s=0}^{h-1} x_s \exp(-2\pi i (j/k) s/h)
\end{aligned}
$$

Thus $b_j = 0$ if j is not a multiple of k and when j is a multiple of k, then $|b_j| = k|\beta_{j/k}|$.

Thus the original DFT terms $\beta_0, \beta_1, \ldots, \beta_{h-1}$ now appear as $b_0 = k\beta_0, b_k = k\beta_1, b_{2k} = k\beta_2$ etc. until $b_{(h-1)k} = k\beta_{h-1}$. All other b_js are zero.

4 DFT and Sample Autocovariance Function

We show below that

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

where $\hat{\gamma}(h)$ is the sample autocovariance function. This gives an important connection between the dft and the sample autocovariance function.

To see this, observe first, by the formula for the sum of a geometric series, that

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

In other words, if the data is constant i.e., $x_0 = \cdots = x_{n-1}$, then b_0 equals nx_0 and b_j equals 0 for all other j. Because of this, we can write:

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

Therefore, for $j = 1, \dots, n-1$, we write

$$
\begin{aligned}
|b_j|^2 = b_j \bar{b}_j &= \sum_{t=0}^{n-1} \sum_{s=0}^{n-1} (x_t - \bar{x})(x_s - \bar{x}) \exp\left(-\frac{2\pi i j t}{n}\right) \exp\left(\frac{2\pi i j s}{n}\right) \\
&= \sum_{t=0}^{n-1} \sum_{s=0}^{n-1} (x_t - \bar{x})(x_s - \bar{x}) \exp\left(-\frac{2\pi i j (t-s)}{n}\right) \\
&= \sum_{h=-(n-1)}^{n-1} \sum_{t,s: t-s=h} (x_t - \bar{x})(x_{t-h} - \bar{x}) \exp\left(-\frac{2\pi i j h}{n}\right) \\
&= n \sum_{|h|<n} \hat{\gamma}(h) \exp\left(-\frac{2\pi i j h}{n}\right).
\end{aligned}
$$