STAT 153 & 248 - Time Series
Lecture Eight

Spring 2025, UC Berkeley

Aditya Guntuboyina

February 13, 2025

1 DFT

Given a time series dataset $y_0, y_1, \dots, y_{n-1}$ of length n, its Discrete Fourier Transform (DFT) is given by $b_0, b_1, \dots, b_{n-1}$ where

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

An alternative expression for b_j is:

$$b_j = \left[\sum_{t=0}^{n-1} y_t \cos\left(\frac{2\pi jt}{n}\right)\right] - i\left[\sum_{t=0}^{n-1} y_t \sin\left(\frac{2\pi jt}{n}\right)\right]. \tag{2}$$

So b_j is a complex number with real part $\sum_t y_t \cos(2\pi jt/n)$ and imaginary part $-\sum_t y_t \sin(2\pi jt/n)$. The following are useful facts to know about the DFT.

1. b_0 is always equal to $y_0 + \dots + y_{n-1}$. To see this, just plug in $j = 0$ in (1) or (2).

2. For each $j = 1, \dots, n-1$, the DFT term b_{n-j} equals the complex conjugate of b_j:

$$b_{n-j} = \bar{b}_j. \tag{3}$$

   The reason for the above is

$$b_{n-j} = \sum_t y_t \exp\left(-\frac{2\pi i(n-j)t}{n}\right) = \sum_t y_t \exp\left(\frac{2\pi ijt}{n}\right)\exp(-2\pi it) = \bar{b}_j,$$

   where, in the above, we used that $\exp(-2\pi it) = 1$ (because t is an integer) and that $\exp\left(\frac{2\pi ijt}{n}\right)$ is the complex conjugate of $\exp\left(-\frac{2\pi ijt}{n}\right)$. Note that, for the above argument, it is crucial that $y_0, \dots, y_{n-1}$ are real. If some of $y_0, \dots, y_{n-1}$ are complex, the relation (3) is no longer true.

Because of (3), the DFT terms for later indices j are determined as the complex conjugates for the DFT indices for earlier indices. For example, when $n = 11$, 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 when $n = 12$, the term b_6 is necessarily real because $b_6 = \bar{b}_6$. The data and their DFT provide equivalent information. When $n = 11$, the data consists of 11 real numbers while the DFT consists of one real number (b_0) and 5 complex numbers. On the other hand, when $n = 12$, the data consists of 12 real numbers while the DFT consists of two real numbers (b_0 and b_6) and 5 complex numbers.

The formula (1) tells how to compute the DFT $\{b_j\}$ from the data $\{y_t\}$. It is also possible to write a formula for recovering the data $\{y_t\}$ from its DFT $\{b_j\}$. This formula is given by:

$$y_t = \frac{1}{n} \sum_{j=0}^{n-1} b_j \exp\left(\frac{2\pi i j t}{n}\right). \tag{4}$$

Note that the two formulae (1) and (4) are quite similar; the differences being in the sign of the exponent in the complex exponential and the presence of the factor $1/n$ in (4). (1) is known as the DFT formula while (4) is known as the Inverse DFT formula.

Formula (4) becomes

$$y = \frac{1}{n} \sum_{j=0}^{n-1} b_j u^j \tag{5}$$

in vector form (note $y = (y_0, y_1, \ldots, y_{n-1})^T$) and this encapsulates the idea that y can be written as a linear combination of the basis sinusoids $u^0, u^1, \ldots, u^{n-1}$.

The most important fact about the DFT is that it can be computed efficiently (in time $O(n \log n)$) using an algorithm called FFT (Fast Fourier Transform). Note that a naive computation of $b_0, \ldots, b_{n-1}$ directly using the formula (1) will take $O(n^2)$ time. The FFT is a divide-and-conquer algorithm that efficiently computes the DFT by using clever recursions (see e.g., `https://en.wikipedia.org/wiki/Fast_Fourier_transform` for details).

2 The Periodogram

The Periodogram is a way of visualizing the DFT. The DFT consists of complex numbers so it is difficult to visualize it directly. The common visualization consists of looking at the squared absolute values of the DFT. More precisely, the periodogram is defined by

$$I\left(\frac{j}{n}\right) := \frac{|b_j|^2}{n} \qquad \text{for } 0 < \frac{j}{n} \leq \frac{1}{2}.$$

One visualizes the size of the DFT terms by plotting the periodogram. Note that $j = 0$ is not plotted as b_0 is simply the sum of the data values and does not provide any information on the sinusoidal components present in the data.

Because

$$b_j = \sum_{t=0}^{n-1} y_t \exp\left(-\frac{2\pi i j t}{n}\right) = \sum_{t=0}^{n-1} y_t \cos\frac{2\pi j t}{n} - i \sum_{t=0}^{n-1} y_t \sin\frac{2\pi j t}{n},$$

we can write the periodogram as:

$$I\left(\frac{j}{n}\right) = \frac{1}{n}\left[\left(\sum_{t=0}^{n-1} y_t \cos\frac{2\pi j t}{n}\right)^2 + \left(\sum_{t=0}^{n-1} y_t \sin\frac{2\pi j t}{n}\right)^2\right] \qquad \text{for } 0 < \frac{j}{n} \leq \frac{1}{2}. \tag{6}$$

3 Utility of the Periodogram

The periodogram is a very commonly used tool for time series data analysis. It has the following two main uses:

1. If we want to fit the single sinusoidal model
$$y_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t) + \epsilon_t \tag{7}$$
to the data, then the periodogram allows efficient computation of $RSS(f)$ for Fourier frequencies $f \in (0, 0.5)$ via the following formula:
$$RSS(f) = \sum_{t=0}^{n-1} (y_t - \bar{y})^2 - 2I(f).$$
For data sets with large n, directly computing $RSS(f)$ over a fine grid might be computationally infeasible. In such cases, one can restrict to Fourier frequencies and compute $RSS(f)$ using the above formula (note that, due to the FFT algorithm, $I(f)$ for Fourier frequencies f can be computed very efficiently).

   Remember also that $RSS(f)$ is key to doing inference for f in the model (7). The MLE of f is obtained by minimizing $RSS(f)$. The Bayesian posterior is given by:
$$\propto \left(\frac{1}{RSS(f)}\right)^{(n-3)/2} |X_f^T X_f|^{-1/2} I\{0 < f < 1/2\}.$$
When f is a Fourier frequency lying in $(0, 1/2)$, we saw in Lecture 6 that
$$X_f^T X_f = \begin{pmatrix} n & 0 & 0 \\ 0 & n/2 & 0 \\ 0 & 0 & n/2 \end{pmatrix} \quad \text{and} \quad (X_f^T X_f)^{-1} = \begin{pmatrix} 1/n & 0 & 0 \\ 0 & 2/n & 0 \\ 0 & 0 & 2/n \end{pmatrix}$$
so that $|X_f^T X_f| = n^3/8$. Importantly, this term does not depend on f. Thus if we restrict to Fourier frequencies, then the Bayesian posterior simplifies to
$$\propto \left(\frac{1}{RSS(f)}\right)^{(n-3)/2} I\{0 < f < 1/2\}.$$

2. The periodogram can suggest alternative models for the data. For example, if the periodogram has two prominent peaks, then this suggests the model:
$$y_t = \beta_0 + \beta_1 \cos(2\pi f_1 t) + \beta_2 \sin(2\pi f_1 t) + \beta_3 \cos(2\pi f_2 t) + \beta_4 \sin(2\pi f_2 t) + \epsilon_t. \tag{8}$$
Formal inference for this model proceeds very similarly to (7). The main difference is that the definition of RSS should now be changed to:
$$RSS(f_1, f_2)$$
$$= \min_{\beta_j, 0 \le j \le 4} \sum_{t=1}^{n} \left(y_t - \beta_0 - \beta_1 \cos(2\pi f_1 t) - \beta_2 \sin(2\pi f_1 t) - \beta_3 \cos(2\pi f_2 t) - \beta_4 \sin(2\pi f_2 t)\right)^2$$
The analysis now proceeds as before with this modified definition of RSS. The MLE of f_1 and f_2 is obtained by minimizing $RSS(f_1, f_2)$ over f_1, f_2, and the Bayesian posterior is given by
$$\propto \left(\frac{1}{RSS(f_1, f_2)}\right)^{(n-5)/2} |X_f^T X_f|^{-1/2}$$

where X_f is now given by

$$X_{f_1,f_2} = \begin{pmatrix} 1 & \cos(2\pi f_1(1)) & \sin(2\pi f_1(1)) & \cos(2\pi f_2(1)) & \sin(2\pi f_2(1)) \\ 1 & \cos(2\pi f_1(2)) & \sin(2\pi f_1(2)) & \cos(2\pi f_2(2)) & \sin(2\pi f_2(2)) \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & \cos(2\pi f_1(n)) & \sin(2\pi f_1(n)) & \cos(2\pi f_2(n)) & \sin(2\pi f_2(n)) \end{pmatrix}$$

In practice, evaluation and minimization of $RSS(f_1, f_2)$ can be done either on a joint grid for f_1 and f_2, or some sequential algorithm. It may be helpful to note here that if f_1 and f_2 are both Fourier frequencies and both lie strictly between 0 and 0.5, then

$$RSS(f_1, f_2) = \sum_t (y_t - \bar{y})^2 - 2I(f_1) - 2I(f_2).$$

Thus if one restricts to Fourier frequencies, then inference under the model (8) can be easily via the periodogram. If we work with arbitrary frequencies, grid-based minimization of RSS and evaluation of the Bayesian posterior would be computationally expensive.

4 Other Nonlinear Regression Models

While our focus so far has been on sinusoidal models, the methodology can be applied in the same way to some other nonlinear regression models. Here are some examples:

1. Consider the model:

$$y_t = \beta_0 + \beta_1 t + \beta_2 \cos(2\pi f t) + \beta_3 \sin(2\pi f t) + \epsilon_t \tag{9}$$

The difference between (7) and (9) is the presence of $\beta_1 t$. The RSS for this model is:

$$RSS(f) = \min_{\beta_0,\beta_1,\beta_2,\beta_3} \sum_{t=1}^{n} \left(y_t - \beta_0 - \beta_1 t - \beta_2 \cos(2\pi f t) - \beta_3 \sin(2\pi f t)\right)^2.$$

2. Consider the model:

$$y_t = \beta_0 + \beta_1 t + \beta_2 (t - s)_+ + \epsilon_t \tag{10}$$

This is sometimes called the broken-stick regression model because the function $t \mapsto \beta_0 + \beta_1 t + \beta_2 (t - s)_+$ resembles a broken stick. RSS for this model is:

$$RSS(s) = \min_{\beta_0,\beta_1,\beta_2} \sum_{t=1}^{n} \left(y_t - \beta_0 - \beta_1 t - \beta_2 (t - s)_+\right)^2.$$

Homework Two will contain some other examples of these models.