STAT 153 & 248 - Time Series
Lecture Nine

Fall 2025, UC Berkeley

Aditya Guntuboyina

September 26, 2025

1 Recap: previous two lectures

We observe time series data $y_0, \dots, y_{n-1}$ (note we are now using Python indexing starting from 0). We want to fit the sinusoidal model:

$$y_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t) + \epsilon_t \qquad \text{with } \epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2). \tag{1}$$

The main parameter is the frequency f (which we restrict to the interval $[0, 0.5]$). The other parameters $(\beta_0, \beta_1, \beta_2, \sigma)$ are nuisance parameters.

The key role in the inference of f is played by the RSS:

$$RSS(f) := \min_{\beta_0, \beta_1, \beta_2} \sum_{t=1}^{n} (y_t - \beta_0 - \beta_1 \cos(2\pi f t) - \beta_2 \sin(2\pi f t))^2 = \|y - X_f \hat{\beta}_f\|^2$$

where

$$y = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \quad \text{and} \quad X_f = \begin{pmatrix} 1 & \cos(2\pi f(1)) & \sin(2\pi f(1)) \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & \cos(2\pi f(n)) & \sin(2\pi f(n)) \end{pmatrix} \quad \text{and} \quad \hat{\beta}_f = (X_f^T X_f)^{-1} X_f^T y.$$

$RSS(f)$ is simply the residual sum of squares in the linear regression model obtained by fixing the frequency parameter f. It describes how well the sinusoid with frequency f fits the observed data $\{y_t\}$.

To estimate f, we minimize $RSS(f)$. For uncertainty quantification for f, we can use the posterior formula:

$$\text{posterior}(f) \propto \left(\frac{1}{RSS(f)} \right)^{(n-3)/2} |X_f^T X_f|^{-1/2} I\{0 < f < 1/2\}.$$

For both these tasks (minimization of $RSS(f)$ and calculation of the posterior of f), we need to discretize and restrict f to a finite set of values in the range $(0, 1/2)$. Here there are two options:

1. Take a dense grid of values in $(0, 1/2)$

2. Work with the Fourier frequencies j/n which are in the range $(0, 1/2)$.

The advantage of taking the second option (Fourier Frequencies) is that it admits very fast computation of $RSS(f)$ via the following:

1. Calculate the DFT of the data:
$$b_j = \sum_{t=0}^{n-1} y_t \exp\left(-\frac{2\pi i j t}{n}\right) \qquad \text{for } j = 0, 1, \dots, n-1.$$
This can be done via a very fast algorithm called the FFT (in python, use `np.fft.fft()`).

2. Calculate the periodogram:
$$I(j/n) := \frac{|b_j|^2}{n} \qquad \text{for } 0 < \frac{j}{n} < 0.5$$

3. Use the formula:
$$RSS(j/n) = \sum_{t=0}^{n-1} (y_t - \bar{y})^2 - 2I(j/n).$$
This formula was proved in Lecture 7. The key observation is that when f is a Fourier frequency lying in $(0, 0.5)$, we have
$$X_f^T X_f = \begin{pmatrix} n & 0 & 0 \\ 0 & n/2 & 0 \\ 0 & 0 & n/2 \end{pmatrix} \tag{2}$$

4. Minimize $RSS(j/n)$ over j to obtain $\hat{f}$.

5. Calculate the posterior of f via:
$$\text{posterior}(j/n) \propto \left(\frac{1}{RSS(j/n)}\right)^{(n-3)/2} I\{0 < j/n < 0.5\}.$$

Note that we did not write the $|X_f^T X_f|^{-1/2}$ term above. This is because, when f is a Fourier frequency in $(0, 0.5)$, we have (2) so that $|X_f^T X_f| = n^3/8$ which does not depend on f (so $|X_f^T X_f|^{-1/2}$ can be absorbed in the constant of proportionality).

It is important to note that restriction to Fourier frequencies is only done for computational reasons. There are no other advantages to doing this, and, in fact, there could be significant loss of information in doing so. This can be easily seen in real datasets such as the sunspots dataset.

2 Sinusoidal Models with more frequencies

Consider the model:
$$y_t = \beta_0 + \beta_1 \cos(2\pi f_1 t) + \beta_2 \sin(2\pi f_1 t) + \beta_3 \cos(2\pi f_2 t) + \beta_4 \sin(2\pi f_2 t) + \epsilon_t. \tag{3}$$
Formal inference for this model proceeds very similarly to inference for model (1). The main difference is that the definition of RSS should now be changed to:

$$RSS(f_1, f_2)$$
$$= \min_{\beta_j, 0 \le j \le 4} \sum_{t=0}^{n-1} (y_t - \beta_0 - \beta_1 \cos(2\pi f_1 t) - \beta_2 \sin(2\pi f_1 t) - \beta_3 \cos(2\pi f_2 t) - \beta_4 \sin(2\pi f_2 t))^2$$

The analysis now proceeds as before with this modified definition of RSS. The least squares estimates of f_1 and f_2 are obtained by minimizing $RSS(f_1, f_2)$ over f_1, f_2, and the Bayesian posterior is given by

$$\propto \left(\frac{1}{RSS(f_1, f_2)}\right)^{(n-5)/2} |X^T_{f_1,f_2} X_{f_1,f_2}|^{-1/2}$$

where X_{f_1,f_2} is now given by

$$X_{f_1,f_2} = \begin{pmatrix} 1 & \cos(2\pi f_1(0)) & \sin(2\pi f_1(0)) & \cos(2\pi f_2(0)) & \sin(2\pi f_2(0)) \\ 1 & \cos(2\pi f_1(1)) & \sin(2\pi f_1(1)) & \cos(2\pi f_2(1)) & \sin(2\pi f_2(1)) \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ 1 & \cos(2\pi f_1(n-1)) & \sin(2\pi f_1(n-1)) & \cos(2\pi f_2(n-1)) & \sin(2\pi f_2(n-1)) \end{pmatrix}$$

Evaluation and minimization of $RSS(f_1, f_2)$ needs to be done on a joint grid for f_1 and f_2.

2.1 Restriction to Fourier Frequencies

Suppose we restrict f_1 and f_2 to be distinct Fourier frequencies lying strictly between 0 and 0.5. Then it can be proved that

$$RSS(f_1, f_2) = \sum_{t=0}^{n-1} (y_t - \bar{y})^2 - 2I(f_1) - 2I(f_2). \tag{4}$$

From the above, it is clear that f_1 and f_2 which minimize $RSS(f_1, f_2)$ equal the top two maximizers of the periodogram (remember that we are assuming f_1 and f_2 to be distinct).

The equality (4) is a consequence of the orthogonality of sinusoidal vectors corresponding to Fourier frequencies. It is proved below.

2.2 Orthogonality of Sinusoidal Vectors at Fourier Frequencies

Orthogonality of Sinusoids at Fourier frequencies is best described through complex sinusoids.

For every $0 \le j \le (n-1)$, let us define the $n \times 1$ vector

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

This vector can be interpreted as the complex sinusoid $e^{2\pi i f t}$ with Fourier frequency $f = j/n$ evaluated at the time points $t = 0, 1, \dots, (n-1)$. It is easy to see that

1. When $j = 0$, we have $u^0 = (1, 1, \dots, 1)$.
2. When $1 \le j \le n-1$, we have $u^j = \overline{u^{n-j}}$. Here $\bar{u}$ denotes complex conjugate of u (the complex conjugate $\bar{u}$ of a vector u is defined as the vector obtained by taking the complex conjugates of each entry of u).

The most important property of these complex valued vectors $u^0, u^1, \dots, u^{n-1}$ is **orthogonality**. Specifically, for $0 \le j \ne k \le n-1$, we have

$$\left\langle u^j, u^k \right\rangle = 0. \tag{5}$$

Recall that the inner product between two complex valued vectors $a = (a_1, \ldots, a_n)^T$ and $b = (b_1, \ldots, b_n)^T$ is given by

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

Note specially the complex conjugate of b_j above.

Here is the proof of (5). Fix $0 \le j \ne k \le n-1$ and write

$$\begin{aligned}
\left\langle u^j, u^k \right\rangle &= \sum_{t=0}^{n-1} \exp\left(2\pi i \frac{j}{n} t\right) \overline{\exp\left(2\pi i \frac{k}{n} t\right)} \\
&= \sum_{t=0}^{n-1} \exp\left(2\pi i \frac{j}{n} t\right) \exp\left(-2\pi i \frac{k}{n} t\right) \\
&= \sum_{t=0}^{n-1} \exp\left(2\pi i \frac{j-k}{n} t\right) \\
&= \sum_{t=0}^{n-1} \left[\exp\left(2\pi i \frac{j-k}{n}\right)\right]^t \\
&= \frac{1 - \left(\exp\left(2\pi i \frac{j-k}{n}\right)\right)^n}{1 - \exp\left(2\pi i \frac{j-k}{n}\right)} \\
&= \frac{1 - \exp\left(2\pi i (j-k)\right)}{1 - \exp\left(2\pi i \frac{j-k}{n}\right)} \\
&= \frac{1 - \cos\left(2\pi (j-k)\right) - i \sin\left(2\pi (j-k)\right)}{1 - \exp\left(2\pi i \frac{j-k}{n}\right)} = \frac{1 - 1 - 0}{1 - \exp\left(2\pi i \frac{j-k}{n}\right)} = 0
\end{aligned}$$

This proves (5). It is also easy to see that (just take $j = k$ in the above calculation and the answer can be found in the third line)

$$\left\langle u^j, u^j \right\rangle = \|u^j\|^2 = n.$$

This orthogonality of complex sinusoids also extends to real sinusoids made of sines and cosines. For a Fourier frequency j/n strictly lying between 0 and 1 (i.e., $0 < j/n < 1/2$), define

$$\mathbf{c}^j := (1, \cos(2\pi j/n), \cos(2\pi 2j/n), \ldots, \cos(2\pi(n-1)j/n))^T$$

and

$$\mathbf{s}^j := (0, \sin(2\pi j/n), \sin(2\pi 2j/n), \ldots, \sin(2\pi(n-1)j/n))^T.$$

These are the vectors obtained by evaulating $\cos(2\pi f t)$ and $\sin(2\pi f t)$ with Fourier Frequency $f = j/n$ at time points $t = 0, 1, \ldots, (n-1)$. So far we have assumed that $0 < f = j/n < 1/2$ i.e., $0 < j < n/2$. The definition can be extended to the boundary points 0 and 1/2 as well.

When $j = 0$, the vector $\mathbf{c}^0$ is the vector of all ones, while $\mathbf{s}^0$ equals the zero vector. When $f = 1/2$ (this is only when n is even other 1/2 will not be a Fourier frequency) or $j = n/2$, we have

$$\mathbf{c}^{n/2} = (1, -1, 1, -1, \ldots, (-1)^{n-1}) \quad \text{and} \quad \mathbf{s}^{n/2} = (0, \ldots, 0).$$

Thus when n is even, the non-zero vectors among these are:

$$\mathbf{c}^0, \mathbf{c}^1, \mathbf{s}^1, \ldots, \mathbf{c}^{\frac{n}{2}-1}, \mathbf{s}^{\frac{n}{2}-1}, \mathbf{c}^{\frac{n}{2}}.$$

When n is odd, we are looking at

$$\mathbf{c}^0, \mathbf{c}^1, \mathbf{s}^1, \ldots, \mathbf{c}^{\frac{n-1}{2}}, \mathbf{s}^{\frac{n-1}{2}}.$$

In either case, the total number of these vectors equals n.

These vectors are orthogonal in $\mathbb{R}^n$. This can be proved as a consequence of the orthogonality of $u^0, \ldots, u^{n-1}$, and the facts:

$$\mathbf{c}^j = \frac{u^j + \overline{u^j}}{2} = \frac{u^j + u^{n-j}}{2} \quad \text{and} \quad \mathbf{s}^j = \frac{u^j - \overline{u^j}}{2i} = \frac{u^j - u^{n-j}}{2i}$$

For example, fix two distinct Fourier frequencies j/n and k/n which are both strictly in $(0, 1/2)$. Then

$$\begin{aligned}\left\langle \mathbf{c}^j, \mathbf{c}^k \right\rangle &= \left\langle \frac{u^j + u^{n-j}}{2}, \frac{u^k + u^{n-k}}{2} \right\rangle \\ &= \frac{1}{4} \left(\left\langle u^j, u^k \right\rangle + \left\langle u^j, u^{n-k} \right\rangle + \left\langle u^{n-j}, u^k \right\rangle + \left\langle u^{n-j}, u^{n-k} \right\rangle \right).\end{aligned}$$

Each inner product above equals zero because of orthogonality of $u^0, \ldots, u^{n-1}$. Indeed, $\langle u^j, u^k \rangle$ and $\langle u^{n-j}, u^{n-k} \rangle$ are zero because $j \neq k$. Also we assumed that $j/n < 0.5$ and $k/n < 0.5$ so that $j + k < n$ which means that $n - j \neq k$ and also $n - k \neq j$. Thus $\langle u^j, u^{n-k} \rangle$ and $\langle u^{n-j}, u^k \rangle$ are also zero. Therefore: $\langle \mathbf{c}^j, \mathbf{c}^k \rangle = 0$ for $j \neq k$ (it can be easily checked that this will be true even when j/n or k/n equal 0 or $1/2$).

One can similarly prove that other inner products (such as those between sines) also equal zero. As another example

$$\langle \mathbf{c}^j, \mathbf{s}^j \rangle = \left\langle \frac{u^j + u^{n-j}}{2}, \frac{u^j - u^{n-j}}{2i} \right\rangle = \frac{1}{4i} \left[\langle u^j, u^j \rangle - \langle u^{n-j}, u^{n-j} \rangle \right] = \frac{1}{4i} (n - n) = 0$$

where we used $\langle u^j, u^j \rangle = n$ for every j.

2.3 Proof of (4)

Suppose $f_1 = j/n$ and $f_2 = k/n$ with $j \neq k$ and both $0 < j/n, k/n < 0.5$. Then note that X_{f_1,f_2} is a $n \times 5$ matrix with columns $\mathbf{c}^0, \mathbf{c}^j, \mathbf{s}^j, \mathbf{c}^k, \mathbf{s}^k$. As a result,

$$\begin{aligned} X^T_{f_1,f_2} X_{f_1,f_2} &= \begin{pmatrix} \langle \mathbf{c}^0, \mathbf{c}^0 \rangle & \langle \mathbf{c}^0, \mathbf{c}^j \rangle & \langle \mathbf{c}^0, \mathbf{s}^j \rangle & \langle \mathbf{c}^0, \mathbf{c}^k \rangle & \langle \mathbf{c}^0, \mathbf{s}^k \rangle \\ \langle \mathbf{c}^j, \mathbf{c}^0 \rangle & \langle \mathbf{c}^j, \mathbf{c}^j \rangle & \langle \mathbf{c}^j, \mathbf{s}^j \rangle & \langle \mathbf{c}^j, \mathbf{c}^k \rangle & \langle \mathbf{c}^j, \mathbf{s}^k \rangle \\ \langle \mathbf{s}^j, \mathbf{c}^0 \rangle & \langle \mathbf{s}^j, \mathbf{c}^j \rangle & \langle \mathbf{s}^j, \mathbf{s}^j \rangle & \langle \mathbf{s}^j, \mathbf{c}^k \rangle & \langle \mathbf{s}^j, \mathbf{s}^k \rangle \\ \langle \mathbf{c}^k, \mathbf{c}^0 \rangle & \langle \mathbf{c}^k, \mathbf{c}^j \rangle & \langle \mathbf{c}^k, \mathbf{s}^j \rangle & \langle \mathbf{c}^k, \mathbf{c}^k \rangle & \langle \mathbf{c}^k, \mathbf{s}^k \rangle \end{pmatrix} \\ &= \begin{pmatrix} \langle \mathbf{c}^0, \mathbf{c}^0 \rangle & 0 & 0 & 0 & 0 \\ 0 & \langle \mathbf{c}^j, \mathbf{c}^j \rangle & 0 & 0 & 0 \\ 0 & 0 & \langle \mathbf{s}^j, \mathbf{s}^j \rangle & 0 & 0 \\ 0 & 0 & 0 & \langle \mathbf{c}^k, \mathbf{c}^k \rangle & 0 \\ 0 & 0 & 0 & 0 & \langle \mathbf{s}^k, \mathbf{s}^k \rangle \end{pmatrix} \end{aligned}$$

It is also easy to see that

$$\langle \mathbf{c}^0, \mathbf{c}^0 \rangle = n \text{ and } \langle \mathbf{c}^j, \mathbf{c}^j \rangle = \langle \mathbf{s}^j, \mathbf{s}^j \rangle = \langle \mathbf{c}^k, \mathbf{c}^k \rangle = \langle \mathbf{s}^k, \mathbf{s}^k \rangle = n/2.$$

We noted these in Lecture 7. They also follow from properties of the complex sinusoidal vectors u^j. For example,

$$\begin{aligned}\langle \mathbf{c}^j, \mathbf{c}^j \rangle &= \left\langle \frac{u^j + u^{n-j}}{2}, \frac{u^j + u^{n-j}}{2} \right\rangle \\ &= \frac{1}{4}\left(\langle u^j, u^j\rangle + \langle u^j, u^{n-j}\rangle + \langle u^{n-j}, u^j\rangle + \langle u^{n-j}, u^{n-j}\rangle\right) \\ &= \frac{1}{4}(n + 0 + 0 + n) = \frac{n}{2}\end{aligned}$$

Thus

$$X_{f_1,f_2}^T X_{f_1,f_2} = \begin{pmatrix} n & 0 & 0 & 0 & 0 \\ 0 & n/2 & 0 & 0 & 0 \\ 0 & 0 & n/2 & 0 & 0 \\ 0 & 0 & 0 & n/2 & 0 \\ 0 & 0 & 0 & 0 & n/2 \end{pmatrix}$$

From here the proof proceeds in the same way as in the case of a single Fourier frequency in Lecture 7 to yield (4).

2.4 More than two Fourier frequencies

When there are three or more Fourier frequencies, the formula (4) generalizes in the same way. For example, if f_1, f_2, f_3 are all distinct Fourier frequencies strictly lying between 0 and 1/2, then

$$RSS(f_1, f_2, f_3) = \sum_{t=0}^{n-1}(y_t - \bar{y})^2 - 2I(f_1) - 2I(f_2) - 2I(f_3).$$

So if we are trying to find the three best Fourier frequencies which best fit the data, we simply pick the top three maximizers of the periodogram. If we do not restrict to Fourier frequencies however, we have to do a harder (say grid based) minimization of $RSS(f_1, f_2, f_3)$. Depending on the application, there could be significant gains in RSS if we go beyond Fourier frequencies.

3 Other Nonlinear Models

Our methodology for inference in these sinusoid models also extends to other similar nonlinear regression models. For example, consider the following model which is applicable when we want to introduce two break points for the regression line:

$$y_t = \beta_0 + \beta_1 t + \beta_2 \mathrm{ReLU}(t - c_1) + \beta_3 \mathrm{ReLU}(t - c_2) + \epsilon_t$$

with $\epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. This model can also be written as

$$y = X_c \beta + \epsilon$$

where $c = (c_1, c_2)$ and

$$X_c := \begin{pmatrix} 1 & 1 & \mathrm{ReLU}(1 - c_1) & \mathrm{ReLU}(1 - c_2) \\ 1 & 2 & \mathrm{ReLU}(2 - c_1) & \mathrm{ReLU}(2 - c_2) \\ 1 & 3 & \mathrm{ReLU}(3 - c_1) & \mathrm{ReLU}(3 - c_2) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & n & \mathrm{ReLU}(n - c_1) & \mathrm{ReLU}(n - c_2) \end{pmatrix}$$

We estimate $c = (c_1, c_2)$ by minimizing $RSS(c)$ over all $c = (c_1, c_2)$ with $c_1, c_2 \in [1, n]$, and

$$RSS(c) = \min_{\beta} \|y - X_c \beta\|^2.$$

One can numerically minimize $RSS(c)$ over all $c_1, c_2 \in [1, n]$. A natural grid one can use here is $\{1, \ldots, n\}$.

The posterior of c becomes:

$$\pi(c \mid \text{data}) \propto \left(\frac{1}{RSS(c)}\right)^{(n-4)/2} |X_c^T X_c|^{-1/2}.$$

For more break points, one can consider:

$$y_t = \beta_0 + \beta_1 t + \sum_{j=1}^{k} \beta_{j+1} \text{ReLU}(t - c_j) + \epsilon_t.$$

Conceptually estimation and inference here proceed just as before with X_c changed appropriately. But the method can become computationally expensive if $k \geq 4$.