STAT 153 - Introduction to Time Series Lecture Four

Fall 2022, UC Berkeley

Aditya Guntuboyina

September 6, 2022

Last week, we used linear regression to fit trend models to the observed time series data. Today we shall look at some nonlinear regression models. Our motivating example will be the sunspots data.

It is often claimed that (see, for example, `https://en.wikipedia.org/wiki/Sunspot`) the sunspot number varies according to an approximately 11-year cycle. We can verify this by fitting the simple sinusoidal model:

$$Y_i = \mu + \alpha_1 \cos(\omega t_i) + \alpha_2 \sin(\omega t_i) + \epsilon_i \qquad \text{for } i = 1, \dots, n \tag{1}$$

to the observed data $(t_1, y_1), \dots, (t_n, y_n)$. Here t_i refers to year i and y_i denotes the average number of sunspots for year t_i. In the dataset (obtained from `https://wwwbis.sidc.be/silso/infosnytot`), we have data for all years from 1700 to 2021. So we are analyzing the whole data, we can take $n = 322$ and $t_1 = 1700, t_2 = 1701, t_3 = 1702, \dots, t_n = 2021$. In general, it is not necessary to have the observed times t_i to be consecutive (i.e., it is okay for the time series to have some observation gaps).

Today, we shall study the problem of fitting the model (1) and obtaining estimates of the frequency parameter ω from the sunspots data. Note that, if we believe the 11-year cycle for the sunspots data, then we would expect the data to give an estimate of ω (in the model (1)) that is close to $2\pi/11 = 0.5712$.

For the model (1), we shall assume that

$$\epsilon_1, \dots, \epsilon_n \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$$

which is the most standard distributional assumption for errors. The problem then is to estimate the frequency parameter ω. The other four parameters $\mu, \alpha_1, \alpha_2, \sigma$ are unknown but they are not our main focus (these parameters can be termed *nuisance parameters*). For principled estimation of ω in the presence of the nuisance parameters $\mu, \alpha_1, \alpha_2, \sigma$, we shall take the Bayesian approach with the following natural prior:

$$\omega, \mu, \alpha_1, \alpha_2, \log \sigma \overset{\text{i.i.d}}{\sim} \text{Unif}(-C, C)$$

for a large number C (the exact value of C will not matter in the following calculations). Note that as σ is always positive, we have made the uniform assumption on $\log \sigma$ (by the change of variable formula, we would have $f_\sigma(x) = f_{\log \sigma}(\log x)\frac{1}{x} = \frac{I\{-C < \log x < C\}}{2Cx} = \frac{I\{e^{-C} < x < e^C\}}{2Cx}$.

The posterior for all the unknown parameters $\omega, \mu, \alpha_1, \alpha_2, \log \sigma$ is then (below we write the term "data" for $Y_1 = y_1, \dots, Y_n = y_n$):

$$f_{\omega,\mu,\alpha_1,\alpha_2,\sigma|\text{data}}(\omega, \mu, \alpha_1, \alpha_2, \sigma) \propto f_{Y_1,\dots,Y_n|\omega,\mu,\alpha_1,\alpha_2,\sigma}(y_1, \dots, y_n) f_{\omega,\mu,\alpha_1,\alpha_2,\sigma}(\omega, \mu, \alpha_1, \alpha_2, \sigma).$$

The two terms on the right hand side above are

$$\begin{aligned}
f_{Y_1,\dots,Y_n|\omega,\mu,\alpha_1,\alpha_2,\sigma}(y_1, \dots, y_n) &\propto \prod_{i=1}^n f_{Y_i|\omega,\mu,\alpha_1,\alpha_2,\sigma}(y_i) \\
&= \prod_{i=1}^n f_{\epsilon_i|\mu,\sigma,\alpha_1,\alpha_2,\sigma}(y_i - \mu - \alpha_1 \cos(\omega t_i) - \alpha_2 \sin(\omega t_i)) \\
&= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(- \frac{(y_i - \mu - \alpha_1 \cos(\omega t_i) - \alpha_2 \sin(\omega t_i)^2}{2\sigma^2} \right) \\
&\propto \sigma^{-n} \exp\left(- \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu - \alpha_1 \cos(\omega t_i) - \alpha_2 \sin(\omega t_i)^2 \right),
\end{aligned}$$

and

$$\begin{aligned}
&f_{\omega,\mu,\alpha_1,\alpha_2,\sigma}(\omega, \mu, \alpha_1, \alpha_2, \sigma) = f_\omega(\omega) f_\mu(\mu) f_{\alpha_1}(\alpha_1) f_{\alpha_2}(\alpha_2) f_\sigma(\sigma) \\
&\propto \frac{I\{-C < \omega < C\}}{2C} \frac{I\{-C < \mu < C\}}{2C} \frac{I\{-C < \alpha_1 < C\}}{2C} \frac{I\{-C < \alpha_2 < C\}}{2C} \frac{I\{e^{-C} < \sigma < e^C\}}{2C\sigma} \\
&\propto \frac{1}{\sigma} I\left\{-C < \omega, \mu, \alpha_1, \alpha_2, \log \sigma < C\right\}.
\end{aligned}$$

We thus obtain

$$\begin{aligned}
&f_{\omega,\mu,\alpha_1,\alpha_2,\sigma|\text{data}}(\omega, \mu, \alpha_1, \alpha_2, \sigma) \\
&\propto \sigma^{-n-1} \exp\left(- \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu - \alpha_1 \cos(\omega t_i) - \alpha_2 \sin(\omega t_i)^2 \right) I\left\{-C < \omega, \mu, \alpha_1, \alpha_2, \log \sigma < C\right\}.
\end{aligned}$$

To obtain the posterior density of ω, we simply integrate the above with respect to $\mu, \alpha_1, \alpha_2, \sigma$. Thus for every $\omega \in (-C, C)$,

$$f_{\omega|\text{data}}(\omega) \propto \int_{e^{-C}}^{e^C} \int_{-C}^{C} \int_{-C}^{C} \int_{-C}^{C} \sigma^{-n-1} \exp\left(- \frac{\sum_{i=1}^n (y_i - \mu - \alpha_1 \cos(\omega t_i) - \alpha_2 \sin(\omega t_i))^2}{2\sigma^2} \right) d\mu d\alpha_1 d\alpha_2 d\sigma.$$

When C is large, the above integral is well-approximated by

$$\int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \sigma^{-n-1} \exp\left(- \frac{\sum_{i=1}^n (y_i - \mu - \alpha_1 \cos(\omega t_i) - \alpha_2 \sin(\omega t_i))^2}{2\sigma^2} \right) d\mu d\alpha_1 d\alpha_2 d\sigma. \tag{2}$$

This integral can be evaluated exactly. The calculation is easiest done using matrix notation. Let

$$Y = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix} 1 & \cos(\omega t_1) & \sin(\omega t_1) \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & cos(\omega t_n) & \sin(\omega t_n) \end{pmatrix} \quad \text{and} \quad \beta = \begin{pmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{pmatrix}.$$

With this notation,

$$\sum_{i=1}^{n}(y_i - \mu - \alpha_1 \cos(\omega t_i) - \alpha_2 \sin(\omega t_i)^2 = \|Y - X\beta\|^2.$$

so that (2) is the same as

$$\int_0^\infty \sigma^{-n-1} \int_{\mathbb{R}^3} \exp\left(-\frac{\|Y - X\beta\|^2}{2\sigma^2}\right) d\beta d\sigma \tag{3}$$

Now if $\hat{\beta}$ is the least squares estimator:

$$\hat{\beta} := \underset{\beta}{\operatorname{argmin}} \|Y - X\beta\|^2,$$

then

$$\|Y - X\beta\|^2 = \|Y - X\hat{\beta}\|^2 + \|X\beta - X\hat{\beta}\|^2 = \|Y - X\hat{\beta}\|^2 + \left(\beta - \hat{\beta}\right)^T X^T X \left(\beta - \hat{\beta}\right).$$

The integral (3) then becomes

$$\int_0^\infty \int_{\mathbb{R}^3} \sigma^{-n-1} \exp\left(-\frac{\|Y - X\hat{\beta}\|^2}{2\sigma^2}\right) \exp\left(-\frac{(\beta - \hat{\beta})X'X(\beta - \hat{\beta})}{2\sigma^2}\right) d\beta d\sigma$$
$$= \int_0^\infty \sigma^{-n-1} \exp\left(-\frac{\|Y - X\hat{\beta}\|^2}{2\sigma^2}\right) \int_{\mathbb{R}^3} \exp\left(-\frac{(\beta - \hat{\beta})X'X(\beta - \hat{\beta})}{2\sigma^2}\right) d\beta d\sigma.$$

We shall now use the formula:

$$\int_{\mathbb{R}^p} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right) dx_1 \dots dx_p = (2\pi)^{p/2}\sqrt{\det(\Sigma)}$$

where Σ is a $p \times p$ positive definite matrix and the integral is over $x = (x_1, \dots, x_p)$. This is basically the formula for the normalizing constant for the multivariate normal distribution.

This formula with $p = 3$ and $\Sigma^{-1} = X'X/(\sigma^2)$ (or equivalently $\Sigma = \sigma^2(X'X)^{-1}$) gives

$$\int_{\mathbb{R}^3} \exp\left(-\frac{(\beta - \hat{\beta})^T X^T X(\beta - \hat{\beta})}{2\sigma^2}\right) d\beta = (2\pi)^{p/2}\sqrt{\det\left(\sigma^2(X'X)^{-1}\right)} = (2\pi)^{p/2}\sigma^p \left(\det(X'X)\right)^{-1/2}.$$

The integral (2) thus equals

$$(2\pi)^{p/2}(\det(X'X))^{-1/2} \int_0^\infty \sigma^{-n+p-1} \exp\left(-\frac{\|Y - X\hat{\beta}\|^2}{2\sigma^2}\right) d\sigma.$$

The change of variable

$$t = \frac{\sigma}{\|Y - X\hat{\beta}\|}$$

then gives

$$(2\pi)^{p/2}(\det(X'X))^{-1/2} \int_0^\infty \sigma^{-n+p-1} \exp\left(-\frac{\|Y - X\hat{\beta}\|^2}{2\sigma^2}\right) d\sigma$$
$$= (2\pi)^{p/2}(\det(X'X))^{-1/2}\|Y - X\hat{\beta}\|^{-n+p} \int_0^\infty t^{-n+p-1} \exp\left(-\frac{1}{2t^2}\right) dt$$
$$\propto (\det(X'X))^{-1/2}\|Y - X\hat{\beta}\|^{-n+p}.$$

Putting everything together, we have proved that

$$f_{\omega|\text{data}}(\omega) \propto (\det(X'X))^{-1/2} \|Y - X\hat{\beta}\|^{-n+p}.$$

Note that the right hand side depends crucially on ω because X depends on ω. Also $\hat{\beta}$ depends on X as $\hat{\beta} = (X'X)^{-1}X'Y$. To make this explicit, let us write $X(\omega)$ for X and $\hat{\beta}(\omega)$ for $\hat{\beta}$:

$$f_{\omega|\text{data}}(\omega) \propto \left(\det(X(\omega)'X(\omega))\right)^{-1/2} \|Y - X(\omega)\hat{\beta}(\omega)\|^{-(n-p)}. \tag{4}$$

This function of ω can be plotted on the computer (and normalized so the density integrates to one). Note that $p = 3$. This allows inference on ω based on the data.

0.1 Connection to the Periodogram

It turns out the Bayesian posterior (4) can be related to the periodogram which is a standard object in time series analysis. The periodogram corresponding to the time series data (t_i, y_i) is defined as

$$I(\omega) := \frac{1}{n}\left[\left(\sum_j y_j \cos(\omega t_j)\right)^2 + \left(\sum_j y_j \sin(\omega t_j)\right)^2\right]. \tag{5}$$

This is a function of $\omega \in \mathbb{R}$. Usually, the periodogram is computed for uniformly spaced data (where the time points t_j can be taken to be consecutive integers such as $0, \dots, n-1$) and when ω is of the form $\frac{2\pi k}{n}$ for some integer $k \in \{1, \dots, n-1\}$. These values of ω are known as *Fourier Frequencies.* Observe that $I(\omega)$ can also be written as

$$I(\omega) = \frac{1}{n}\left|\sum_j y_j e^{i\omega t_j}\right|^2$$

where $i = \sqrt{-1}$, $e^{i\omega t_j}$ is the complex number $\cos(\omega t_j) + i\sin(\omega t_j)$ and $|z|$ for a complex number z denotes its modulus. The complex number

$$b(\omega) := \sum_j y_j e^{i\omega t_j}$$

is termed the Discrete Fourier Transform of the data when $t_j = j - 1$ and ω ranges over the Fourier frequencies. Thus, the periodogram is basically the squared modulus of the DFT (scaled by n).

It is a standard procedure to look at the periodogram of an observed time series in order to determine periodic components present in the data. It turns out that the Bayesian posterior (4) is related to the periodogram as we shall argue below. To see this, first note that the posterior (4) is described in terms of the matrix $X(\omega)$. For this matrix, it is easy to see that

$$X'(\omega)X(\omega) = \begin{pmatrix} n & \sum_{j=1}^n \cos(\omega t_j) & \sum_{j=1}^n \sin(\omega t_j) \\ \sum_{j=1}^n \cos(\omega t_j) & \sum_{j=1}^n \cos^2(\omega t_j) & \sum_{j=1}^n \cos(\omega t_j)\sin(\omega t_j) \\ \sum_{j=1}^n \sin(\omega t_j) & \sum_{j=1}^n \cos(\omega t_j)\sin(\omega t_j) & \sum_{j=1}^n \sin^2(\omega t_j). \end{pmatrix}$$

Quite often, this $X'(\omega)X(\omega)$ matrix can be well-approximated as

$$n\begin{pmatrix} 1 & \frac{1}{n}\sum_{j=1}^n \cos(\omega t_j) & \frac{1}{n}\sum_{j=1}^n \sin(\omega t_j) \\ \frac{1}{n}\sum_{j=1}^n \cos(\omega t_j) & \frac{1}{n}\sum_{j=1}^n \cos^2(\omega t_j) & \frac{1}{n}\sum_{j=1}^n \cos(\omega t_j)\sin(\omega t_j) \\ \frac{1}{n}\sum_{j=1}^n \sin(\omega t_j) & \frac{1}{n}\sum_{j=1}^n \cos(\omega t_j)\sin(\omega t_j) & \frac{1}{n}\sum_{j=1}^n \sin^2(\omega t_j). \end{pmatrix} = n\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{pmatrix}$$

and we shall make this assumption in the rest of this section:

$$X'(\omega)X(\omega) \approx \begin{pmatrix} n & 0 & 0 \\ 0 & n/2 & 0 \\ 0 & 0 & n/2 \end{pmatrix}$$

We will see some rigorous justification for this assumption in the next class.

Under this condition, the integral (3) can be evaluated in the following alternative way. We start with

$$\begin{aligned}
&\|Y - X\beta\|^2 = Y'Y - 2Y'X\beta + \beta'X'X\beta \\
&= \sum_i y_i^2 - 2\sum_{i=1}^n y_i(\mu + \alpha_1\cos(\omega t_i) + \alpha_2\sin(\omega t_i)) + n\mu^2 + \frac{n}{2}\alpha_1^2 + \frac{n}{2}\alpha_2^2 \\
&= \sum_i y_i^2 - 2\mu\sum_i y_i + n\mu^2 - 2\alpha_1\sum_i y_i\cos(\omega t_i) + \frac{n}{2}\alpha_1^2 - 2\alpha_2\sum_i y_i\sin(\omega t_i) + \frac{n}{2}\alpha_2^2
\end{aligned}$$

Thus the inner integral over $\mathbb{R}^3$ in (3) can be broken down into 3 one dimensional integrals (as opposed to one three-dimensional integral) as

$$\begin{aligned}
&\int_{\mathbb{R}^3} \exp\left(-\frac{\|Y - X\beta\|^2}{2\sigma^2}\right) d\beta \\
&= \exp\left(-\frac{\sum_i y_i^2}{2\sigma^2}\right)\left[\int \exp\left(\frac{\mu\sum_i y_i}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}\right) d\mu\right]\left[\int \exp\left(\frac{\alpha_1\sum_i y_i\cos(\omega t_i)}{\sigma^2} - \frac{n\alpha_1^2}{4\sigma^2}\right) d\alpha_1\right] \\
&\left[\int \exp\left(\frac{\alpha_2\sum_i y_i\sin(\omega t_i)}{\sigma^2} - \frac{n\alpha_2^2}{4\sigma^2}\right) d\alpha_2\right]
\end{aligned}$$

can be evaluated in the following alternative way. Each of the above three integrals can be evaluated explicitly using the one-dimensional integration formula:

$$\int_{-\infty}^{\infty} \exp\left(xC_1 - \frac{C_2}{2}x^2\right) dx = \sqrt{\frac{2\pi}{C_2}}\exp\left(\frac{C_1^2}{2C_2}\right).$$

We thus deduce

$$\begin{aligned}
&\int_{\mathbb{R}^3} \exp\left(-\frac{\|Y - X\beta\|^2}{2\sigma^2}\right) d\beta \\
&\propto \exp\left(-\frac{\sum_i y_i^2}{2\sigma^2}\right)\sigma^3\exp\left(\frac{(\sum_i y_i)^2}{2n\sigma^2}\right)\exp\left(\frac{(\sum_i y_i\cos(\omega t_i))^2}{2\sigma^2}\right)\exp\left(\frac{(\sum_i y_i\sin(\omega t_i))^2}{2\sigma^2}\right).
\end{aligned}$$

Finally the integration over σ can be done as before to obtain

$$\begin{aligned}
f_{\omega|\text{data}}(\omega) &\propto \left[\frac{\sum_i y_i^2}{2} - \frac{(\sum_i y_i)^2}{2n} - \frac{1}{n}\left(\sum_i y_i\cos(\omega t_i)\right)^2 - \frac{1}{n}\left(\sum_i y_i\sin(\omega t_i)\right)^2\right]^{-(n-p)/2} \\
&= \left[\frac{\sum_i (y_i - \bar{y})^2}{2} - \frac{1}{n}\left(\sum_i y_i\cos(\omega t_i)\right)^2 - \frac{1}{n}\left(\sum_i y_i\sin(\omega t_i)\right)^2\right]^{-(n-p)/2}
\end{aligned}$$

Using the periodogram formula (5), we can write the above as

$$\begin{aligned}
f_{\omega|\text{data}}(\omega) &\propto \left[\frac{\sum_i (y_i - \bar{y})^2}{2} - I(\omega)\right]^{-(n-p)/2} \\
&\propto \left[1 - \frac{2I(\omega)}{\sum_{i=1}^n (y_i - \bar{y})^2}\right]^{-(n-p)/2}.
\end{aligned}$$

Thus the Bayesian posterior for ω is essentially a function of the periodogram (and the sample variance of the data). But it is important to note that it is a very specific function which can look quite different from the raw periodogram. For example, for the sunspots dataset, the periodogram has several peaks but the Bayesian posterior is typically quite strongly concentrated. Thus if we are trying to find a single frequency in a time series dataset, the Bayesian posterior will provide that information much more precisely compared to the periodogram.