STAT 153 & 248 - Time Series
Lecture Five

Spring 2025, UC Berkeley

Aditya Guntuboyina

February 4, 2025

1 Nonlinear Regression

Today, we shall start our discussion on models in which certain parameters appear in non-linear fashion. The simplest example is the sinusoidal model which we attempt to fit to the sunspots data. Before we start discussing the sinusoidal model, let us review the basic sinusoid functions.

2 The Sinusoid

When we say sinusoid, we refer to the following function of time (t):

$$s(t) := R\cos(2\pi ft + \phi) \tag{1}$$

R is called the *amplitude*, f is called the *frequency* and Φ is called the *phase*. The quantity $1/f$ is called the *period* and $2\pi f$ is termed the *angular frequency*. Sometimes, we shall use the notation $\omega = 2\pi f$ for the angular frequency.

Using the formula $\cos(\alpha+\beta) = (\cos\alpha)(\cos\beta) - (\sin\alpha)(\sin\beta)$, we can represent the sinusoid (1) in the following equivalent alternative form:

$$s(t) = A\cos 2\pi ft + B\sin 2\pi ft. \tag{2}$$

The parameters A, B in (2) are related to R, ϕ in (1) via $A = R\cos\phi$ and $B = R\sin\phi$. While working with models involving sinuosoids, we use the representation (2) because the parameters A and B appear linearly in (2).

2.1 Discrete sampling and restricting f to $[0, 1/2]$

Often in time series analysis, we work with equally spaced time points and assume that the time variable t takes the values $1, \dots, n$ (where n is the sample size). It turns out that if we consider the sinusoid (1) and restrict the time t to $1, \dots, n$, then we can always constrain the frequency parameter f to $[0, 1/2]$. This is a consequence of the following result.

Fact 2.1. *For every $f \in (-\infty, \infty)$ and $\phi \in (-\infty, \infty)$, there exists $f_0 \in [0, 1/2]$ and $\phi_0 \in (-\infty, \infty)$ such that*

$$s(t) = R\cos(2\pi ft + \phi) = R\cos(2\pi f_0 t + \phi_0) \qquad \textit{for all } t = 1, \dots, n.$$

Proof. Consider the following three cases.

1. If $f < 0$, then we can write $\cos(2\pi f t + \phi) = \cos(2\pi(-f)t - \phi)$. Clearly, $-f \geq 0$.
2. If $f \geq 1$, then we write (below $[f]$ is the largest integer less than or equal to f):
$$\cos(2\pi f t + \phi) = \cos(2\pi[f]t + 2\pi(f - [f])t + \phi) = \cos(2\pi(f - [f])t + \phi),$$
because $\cos(\cdot)$ is periodic with period 2π. Clearly $0 \leq f - [f] < 1$.
3. If $f \in [1/2, 1)$, then
$$\cos(2\pi f t + \phi) = \cos(2\pi t - 2\pi(1-f)t + \phi) = \cos(2\pi(1-f)t - \phi)$$
because $\cos(2\pi t - x) = \cos x$ for all integers t. Clearly $0 < 1 - f \leq 1/2$.

Thus the sinusoid $R\cos(2\pi f t + \phi)$ equals $R\cos(2\pi f_0 t + \phi_0)$ at all integers t for some $0 \leq f_0 \leq 1/2$ and a phase ϕ_0 that is possibly different from ϕ. □

From now on, when we discuss sinusoids $s(t) = R\cos(2\pi f t + \phi)$ in the context of $t = 1, \dots, n$, we shall assume that the frequency parameter f is restricted to $[0, 1/2]$. Note also the behavior of the sinusoid for the two frequency extremes $f = 0$ and $f = 1/2$. When $f = 0$, the sinusoid $s(t)$ is simply a constant function equal to $R\cos(\phi)$. When $f = 1/2$, we have
$$s(t) = R\cos(\pi t + \phi) = R(\cos\phi)\cos(\pi t) = R(-1)^t \cos\phi.$$
This sinusoid exhibits the maximum possible oscillation going back and forth between $R\cos\phi$ and $-R\cos\phi$.

3 The sinusoidal model

The simplest sinusoidal model for a time series $y_1, \dots .y_n$ is
$$y_t = \beta_0 + \beta_1 \cos(2\pi f t) + \beta_2 \sin(2\pi f t) + \epsilon_t \qquad \text{where } \epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2). \tag{3}$$
The unknown parameters in this model are $\beta_0, \beta_1, \beta_2, \sigma$ as well as the frequency parameter f. As discussed above, we assume that the frequency f lies between 0 and 0.5. If f is assumed to be known, then clearly (3) is a multiple linear regression model and we can use the techniques of the past few lectures to do inference on $\beta_0, \beta_1, \beta_2, \sigma$. But if f is unknown (as will be the case for the sunspots dataset for example), then this is a nonlinear regression model.

We discuss the problem of parameter estimation and inference particularly focussing on the parameter f.

3.1 MLE

Let us first discuss the MLE. The likelihood is given by:
$$\begin{aligned} f_{\text{data}|f,\beta_0,\beta_1,\beta_2,\sigma}(y_1, \dots, y_n) &= \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_t - \beta_0 - \beta_1\cos(2\pi f t) - \beta_2 \sin(2\pi f t))^2}{2\sigma^2}\right) \\ &\propto \sigma^{-n} \exp\left(-\frac{\sum_{t=1}^{n}(y_t - \beta_0 - \beta_1\cos(2\pi f t) - \beta_2 \sin(2\pi f t))^2}{2\sigma^2}\right). \end{aligned} \tag{4}$$

It is clear that the MLEs $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{f}$ will be given by minimizing the least squares criterion:

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{f}) = \underset{\beta_0,\beta_1,\beta_2,f}{\operatorname{argmin}} S(\beta_0, \beta_1, \beta_2, f) \tag{5}$$

where

$$S(\beta_0, \beta_1, \beta_2, f) := \sum_{t=1}^{n} (y_t - \beta_0 - \beta_1 \cos 2\pi f t - \beta_2 \sin 2\pi f t)^2 \tag{6}$$

Let us use here the following notation (previously similar notation was used in the context of linear models):

$$y = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{pmatrix} \quad \text{and} \quad X_f = \begin{pmatrix} 1 & \cos(2\pi f(1)) & \sin(2\pi f(1)) \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & \cos(2\pi f(n)) & \sin(2\pi f(n)) \end{pmatrix} \quad \text{and} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}.$$

The X_f matrix is the same as the X-matrix in the linear model when f is assumed known (its first column is all ones, second column is $\cos(2\pi f t)$ evaluated at $t = 1, \dots, n$, and the third column is $\sin(2\pi f t)$ evaluated at $t = 1, \dots, n$).

With this notation, (6) becomes

$$S(\beta_0, \beta_1, \beta_2, f) = S(\beta, f) = \|y - X_f \beta\|^2. \tag{7}$$

For each fixed value of f, the quantity $S(\beta, f)$ is minimized (over β) at $\beta = \hat{\beta}(f)$ where

$$\hat{\beta}(f) := (X_f^T X_f)^{-1} X_f^T y.$$

Further

$$S(\hat{\beta}(f), f) = \min_{\beta} S(\beta, f) = RSS(f)$$

where $RSS(f)$ is the Residual Sum of Squares in the regression problem of y over X_f for fixed f. Because

$$\min_{\beta, f} S(\beta, f) = \min_{f} \left(\min_{\beta} S(\beta, f) \right) = \min_{f} S(\hat{\beta}(f), f).$$

These observations can be put together to get the following algorithm for computing the MLEs:

1. Take a grid of all possible values of f in the range $[0, 1/2]$.
2. For each frequency value f in the grid,
   a) Form the matrix X_f
   b) Do a regression of y on X_f and compute the Residual Sum of Squares $RSS(f)$
3. Take $\hat{f}$ to be the grid value which minimizes $RSS(f)$ over all the grid values.
4. Take $\hat{\beta}$ and $\hat{\sigma}$ to be the usual regression estimates (of β and σ) in the linear regression of y on $X_{\hat{f}}$.

After finding the MLE, the next step is uncertainty quantification (which involves getting confidence intervals of the parameters etc.). However, this is tricky to do in this problem. We will instead use Bayesian analysis for uncertainty quantification.

3.2 Bayesian Inference

We shall work with the following prior. We assume that $\beta_0, \beta_1, \beta_2, \sigma, f$ are independent with

$$\beta_0, \beta_1, \beta_2, \log\sigma \overset{\text{i.i.d}}{\sim} \text{unif}(-C, C) \text{ and } f \sim \text{unif}[0, 1/2].$$

The priors on $\beta_0, \beta_1, \beta_2, \sigma$ are the same as before in linear regression. The prior on f is confined to $[0, 1/2]$ because, as already discussed, we are restricting the frequency parameter to $[0, 1/2]$.

The posterior joint density of all the parameters β, f, σ (here β denotes the vector consisting of $\beta_0, \beta_1\beta_2$) is given by

$$\text{posterior}(\beta, f, \sigma) \propto \text{likelihood} \times \text{prior}.$$

The likelihood is given in (4) and the prior is

$$\begin{aligned} \text{prior density} &= \frac{I\{-C < \beta_0, \beta_1, \beta_2, \log\sigma < C\}}{(2C)^4\sigma} \frac{I\{0 \le f \le 1/2\}}{1/2} \\ &\propto \frac{I\{-C < \beta_0, \beta_1, \beta_2, \log\sigma < C, 0 \le f \le 1/2\}}{\sigma} \end{aligned}$$

We shall drop the indicator terms involving C while writing the posterior because, as we have seen previously in the case of linear regression, they will have essentially no impact on the posterior. The posterior is thus given by

$$\text{posterior}(\beta, f, \sigma) \propto \sigma^{-n-1} \exp\left(-\frac{S(\beta, f)}{2\sigma^2}\right) I\{\sigma > 0\} I\{0 \le f \le 1/2\}.$$

To get the posterior density of f alone (f is the most important parameter in the sinusoidal model), we need to integrate the joint posterior density above with respect to β and σ:

$$\text{posterior}(f) \propto I\{0 \le f \le 1/2\} \int_0^\infty \sigma^{-n-1} \int \exp\left(-\frac{S(\beta, f)}{2\sigma^2}\right) d\beta d\sigma.$$

Let us first calculate the inner integral. $S(\beta, f)$ is a quadratic function in β so that $\int \exp(-S(\beta, f)/(2\sigma^2)) d\beta$ should be related to the normalizing constants in the multivariate normal density. To figure the integral precisely, we first use the Pythagorean identity (discussed last lecture):

$$S(\beta, f) = S(\hat{\beta}(f), f) + \left(\beta - \hat{\beta}(f)\right)^T X_f^T X_f \left(\beta - \hat{\beta}(f)\right).$$

Thus

$$\begin{aligned} \int \exp\left(-\frac{S(\beta, f)}{2\sigma^2}\right) d\beta &= \int \exp\left(-\frac{S(\hat{\beta}(f), f)}{2\sigma^2}\right) \exp\left(-\frac{\left(\beta - \hat{\beta}(f)\right)^T X_f^T X_f \left(\beta - \hat{\beta}(f)\right)}{2\sigma^2}\right) d\beta \\ &= \exp\left(-\frac{S(\hat{\beta}(f), f)}{2\sigma^2}\right) \int \exp\left(-\frac{\left(\beta - \hat{\beta}(f)\right)^T X_f^T X_f \left(\beta - \hat{\beta}(f)\right)}{2\sigma^2}\right) d\beta \\ &= \exp\left(-\frac{S(\hat{\beta}(f), f)}{2\sigma^2}\right) (\sqrt{2\pi})^p \sqrt{\det\left(\sigma^2 (X_f^T X_f)^{-1}\right)} \\ &= \exp\left(-\frac{S(\hat{\beta}(f), f)}{2\sigma^2}\right) (\sqrt{2\pi})^p \sigma^p \left|X_f^T X_f\right|^{-1/2} \end{aligned}$$

where $|X_f^T X_f| = \det(X_f^T X_f)$. Here $p = 3$ because there are three components inside β. Therefore

$$\text{posterior}(f) \propto I\{0 \le f \le 1/2\} \int_0^\infty \sigma^{-n-1} \exp\left(-\frac{S(\hat{\beta}(f), f)}{2\sigma^2}\right) (\sqrt{2\pi})^p \sigma^p \left|X_f^T X_f\right|^{-1/2} d\sigma$$

$$\propto I\{0 \le f \le 1/2\} \left|X_f^T X_f\right|^{-1/2} \int_0^\infty \sigma^{-n+p-1} \exp\left(-\frac{S(\hat{\beta}(f), f)}{2\sigma^2}\right) d\sigma.$$

The change of variable $\sigma = s\sqrt{S(\hat{\beta}(f), f)}$, gives

$$\text{posterior}(f) \propto I\{0 \le f \le 1/2\} \left|X_f^T X_f\right|^{-1/2} \left(\frac{1}{S(\hat{\beta}(f), f)}\right)^{(n-p)/2} \int_0^\infty t^{-n+p-1} \exp\left(-\frac{1}{2t^2}\right) dt$$

$$\propto I\{0 \le f \le 1/2\} \left|X_f^T X_f\right|^{-1/2} \left(\frac{1}{S(\hat{\beta}(f), f)}\right)^{(n-p)/2}.$$

The main term in this posterior is $(S(\hat{\beta}(f), f))^{-(n-p)/2}$ (the other term $|X_f^T X_f|^{-1/2}$ does not vary significantly with f). It takes its largest value when f equals the MLE $\hat{f}$ which minimizes $S(\hat{\beta}(f), f)$. The size of the power $n - p$ determines the amount of concentration of the posterior around the MLE $\hat{f}$. When n is large, this posterior is concentrated very tightly around $\hat{f}$.

This posterior is evaluated numerically over a grid of values of f in the range $[0, 0.5]$. The term $|X_f^T X_f|^{-1/2}$ becomes infinite when $|X_f^T X_f| = 0$ i.e., when X_f does not have full column rank. This is the case when $f = 0$ or when $f = 1/2$. We need to exclude these edge cases while computing this posterior.