STAT 153 & 248 - Time Series
Lecture Fifteen

Spring 2025, UC Berkeley

Aditya Guntuboyina

March 11, 2025

1 Spectrum Model

In the last lecture, we looked at the spectral model for a given time series dataset $y_0, \dots, y_{n-1}$. In terms of the DFT $b_0, \dots, b_{n-1}$, the model is given by:

$$\text{Re}(b_j), \text{Im}(b_j) \overset{\text{i.i.d}}{\sim} N(0, \gamma_j^2) \tag{1}$$

for $j = 1, \dots, m$ where $m = (n-1)/2$ (we are assuming that n is odd). The unknown parameters in this model are $\gamma_1^2, \dots, \gamma_m^2$ and γ_j represents the strength of sinusoids at frequency j/n.

The likelihood corresponding to (1) is proportional to:

$$\prod_{j=1}^{m} \frac{1}{\gamma_j} \exp\left(-\frac{(\text{Re}(b_j))^2}{2\gamma_j^2}\right) \frac{1}{\gamma_j} \exp\left(-\frac{(\text{Im}(b_j))^2}{2\gamma_j^2}\right)$$
$$= \prod_{j=1}^{m} \frac{1}{\gamma_j^2} \exp\left(-\frac{(\text{Re}(b_j))^2 + (\text{Im}(b_j))^2}{2\gamma_j^2}\right) = \prod_{j=1}^{m} \frac{1}{\gamma_j^2} \exp\left(-\frac{|b_j|^2}{2\gamma_j^2}\right).$$

Therefore the likelihood depends on the squared magnitudes $|b_j|^2$ of the DFT coefficients. Recall that the periodogram $I(j/n)$ is defined as

$$I(j/n) := \frac{|b_j|^2}{n}.$$

We can therefore rewrite the likelihood in terms of the periodogram as follows:

$$\prod_{j=1}^{m} \frac{1}{\gamma_j^2} \exp\left(-\frac{nI(j/n)}{2\gamma_j^2}\right). \tag{2}$$

This likelihood depends on the data only through the periodogram ordinates $I(j/n)$ for $j = 1, \dots, m$. Therefore the periodogram forms the sufficient statistic in this model. Under (7), we have

$$I(j/n) = \frac{1}{n}|b_j|^2 = \frac{1}{n}\left((\text{Re}(b_j))^2 + (\text{Im}(b_j))^2\right) \sim \frac{\gamma_j^2}{n}\chi_2^2.$$

The model can therefore be written directly in terms of the periodogram as

$$I(j/n) \overset{\text{ind}}{\sim} \frac{\gamma_j^2}{n}\chi_2^2 \qquad \text{for } j = 1, \dots, m.$$

We can write the likelihood for the above model in terms of the periodogram and this would be proportional to (8). Note also that χ^2_2 distribution with two degrees of freedom actually coincides with the Exponential distribution with λ parameter equal to 1/2.

Model (1) does not care so much about the individual DFT coefficients b_j but only their magnitude.

The negative log-likelihood corresponding to (8) is

$$\sum_{j=1}^{m} \left(2 \log \gamma_j + \frac{nI(j/n)}{2\gamma_j^2} \right).$$

For optimization purposes we work with the logarithms of γ_j. Let $\alpha_j = \log \gamma_j$. The negative log-likelihood in terms of α_j is

$$\sum_{j=1}^{m} \left(2\alpha_j + \frac{nI(j/n)}{2} e^{-2\alpha_j} \right).$$

If we directly minimize the above with respect to α_j (without any additional regularization), we get

$$\alpha_j = \log \sqrt{\frac{nI(j/n)}{2}} \quad \text{and} \quad \gamma_j^2 = e^{2\alpha_j} = \frac{nI(j/n)}{2}.$$

This basically means that the γ_j^2 parameters fully interpolate the periodogram leading to full overfitting. For more meaningful estimation, we need to add regularization. If we assume smoothness of α_j, we can add the penalty $\sum_{j=2}^{m-1}((\alpha_{j+1} - \alpha_j) - (\alpha_j - \alpha_{j-1}))^2$ or $\sum_{j=2}^{m-1} |(\alpha_{j+1} - \alpha_j) - (\alpha_j - \alpha_{j-1})|$ to the negative log-likelihood. This leads to the estimators $\hat{\alpha}_t^{\text{ridge}}(\lambda)$ and $\hat{\alpha}_t^{\text{lasso}}(\lambda)$ which are defined as the minimizers of

$$\sum_{j=1}^{m} \left(2\alpha_j + \frac{nI(j/n)}{2} e^{-2\alpha_j} \right) + \lambda \sum_{j=2}^{m-1} ((\alpha_{j+1} - \alpha_j) - (\alpha_j - \alpha_{j-1}))^2$$

and

$$\sum_{t=1}^{n} \left(2\alpha_j + \frac{nI(j/n)}{2} e^{-2\alpha_j} \right) + \lambda \sum_{j=2}^{m-1} |(\alpha_{j+1} - \alpha_j) - (\alpha_j - \alpha_{j-1})|$$

respectively. The penalties encourage smoothness in $\{\alpha_j\}$, leading to more stable and interpretable estimates for $\{\gamma_j^2\}$.

2 Power Spectral Density

The sufficient statistic for Model 2 is the periodogram $I(j/n)$. The mean of the periodogram (according to the model) is given by $2\gamma_j^2/n$. This quantity is known as the **power** of frequency j/n:

$$f(j/n) = \text{power of frequency } j/n = \frac{2\gamma_j^2}{n}.$$

If we plot the points $(j/n, f(j/n))$ for $j = 1, \ldots, m$ and join the neighboring points by lines, we get a continuous function plot. This function is known as the **power spectral density** and is defined on $[0, 0.5]$.

This definition of the power spectral density is not rigorous. For a rigorous treatment, see any book on time series (e.g., Chapter 4 of Shumway and Stoffer; or the book "Spectral Analysis for Univariate Time Series" by Percival and Walden).

After estimating the parameters $\gamma_1^2, \dots, \gamma_m^2$, it is customary to look at a plot of the estimated power spectral density $f(j/n) = 2\gamma_j^2/n$ (also known as the power spectrum).

3 Rewriting the Model in terms of y_t

The model (1) is written in terms of the DFT b_j. Because the original data can be written in terms of the DFT, we can convert (1) into a specification for the original data y_t. This will lead us to the model representation that we already saw in Lecture 13.

The key to this is the following formula which writes the data in terms of the DFT.

$$y_t = \frac{1}{n}\sum_{j=0}^{n-1} b_j \exp\left(\frac{2\pi ijt}{n}\right) \qquad \text{for } t = 0, 1, \dots, n-1. \tag{3}$$

We saw this formula previously in Lecture 8. It is known as the **Inverse DFT** formula.

The right hand side of (3) involves complex numbers (b_j and $\exp(2\pi ijt/n)$). On the other hand, the left hand is the data y_t which is always real. Below we change the right hand side in (3) to make it consist of only real terms.

We can rewrite the inverse DFT formula in the following way.

$$\begin{aligned} y_t &= \frac{1}{n}\sum_{j=0}^{n-1} b_j \exp\left(\frac{2\pi ijt}{n}\right) \\ &= \frac{1}{n}\sum_{j=0}^{n-1} (\mathrm{Re}(b_j) + i\ \mathrm{Im}(b_j))\left(\cos\left(\frac{2\pi jt}{n}\right) + i\sin\left(\frac{2\pi jt}{n}\right)\right) \\ &= \frac{1}{n}\sum_{j=0}^{n-1}\left(\mathrm{Re}(b_j)\cos\left(\frac{2\pi jt}{n}\right) - \mathrm{Im}(b_j)\sin\left(\frac{2\pi jt}{n}\right)\right) + i\frac{1}{n}\sum_{j=0}^{n-1}\left(\mathrm{Re}(b_j)\sin\left(\frac{2\pi jt}{n}\right) + \mathrm{Im}(b_j)\cos\left(\frac{2\pi jt}{n}\right)\right) \end{aligned}$$

We can ignore the imaginary part above as the dataset consists of real numbers, and this leads to

$$\begin{aligned} y_t &= \frac{1}{n}\sum_{j=0}^{n-1}\left(\mathrm{Re}(b_j)\cos\left(\frac{2\pi jt}{n}\right) - \mathrm{Im}(b_j)\sin\left(\frac{2\pi jt}{n}\right)\right) \\ &= \frac{b_0}{n} + \frac{1}{n}\sum_{j=1}^{n-1}\left(\mathrm{Re}(b_j)\cos\left(\frac{2\pi jt}{n}\right) - \mathrm{Im}(b_j)\sin\left(\frac{2\pi jt}{n}\right)\right) \end{aligned}$$

We are assuming that n is odd and that $m = (n-1)/2$. We split the sum above into $j = 1, \dots, m$ and then $j = m+1, \dots, n-1$, and then use $b_{n-j} = \bar{b}_j$ or, equivalently, $\mathrm{Re}(b_{n-j}) = \mathrm{Re}(b_j)$ and $\mathrm{Im}(b_{n-j}) = -\mathrm{Im}(b_j)$. This gives

$$y_t = \frac{b_0}{n} + \sum_{j=1}^{m}\left(\frac{2\mathrm{Re}(b_j)}{n}\cos\left(\frac{2\pi jt}{n}\right) + \frac{-2\mathrm{Im}(b_j)}{n}\sin\left(\frac{2\pi jt}{n}\right)\right).$$

where we also used $\cos(2\pi(n-j)t/n) = \cos(2\pi jt/n)$ and $\sin(2\pi(n-j)t/n) = -\sin(2\pi jt/n)$.

In other words, when n is odd and $m = (n-1)/2$, we have

$$y_t = \beta_0 + \sum_{j=1}^{m} \left(\beta_{1j} \cos \frac{2\pi jt}{n} + \beta_{2j} \sin \frac{2\pi jt}{n} \right) \tag{4}$$

where, for $j = 1, \ldots, m$,

$$\beta_0 = \frac{b_0}{n} \quad \beta_{1j} = \frac{2\text{Re}(b_j)}{n} \quad \beta_{2j} = -\frac{2\text{Im}(b_j)}{n} \tag{5}$$

The formula (4) holds for every dataset $y_0, \ldots, y_{n-1}$. As a result, the model (1) is equivalent to (4) with

$$\beta_{1j} = \frac{2\text{Re}(b_j)}{n} \sim N\left(0, \frac{4}{n^2}\gamma_j^2\right) \quad \text{and} \quad \beta_{2j} = \frac{-2\text{Re}(b_j)}{n} \sim N\left(0, \frac{4}{n^2}\gamma_j^2\right).$$

The spectrum model therefore has the following two equivalent definitions:

- **Definition 1**: $\text{Re}(b_1), \text{Im}(b_1), \ldots, \text{Re}(b_m), \text{Im}(b_m)$ are all independent with $\text{Re}(b_j), \text{Im}(b_j) \overset{\text{i.i.d}}{\sim} N(0, \gamma_j^2)$ for $j = 1, \ldots, m$.
- **Definition 2**:
$$y_t = \beta_0 + \sum_{j=1}^{m} \left(\beta_{1j} \cos \frac{2\pi jt}{n} + \beta_{2j} \sin \frac{2\pi jt}{n} \right) \tag{6}$$
with $\beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}, \ldots, \beta_{1m}, \beta_{2m}$ all independent with
$$\beta_{1j}, \beta_{2j} \overset{\text{i.i.d}}{\sim} N(0, \tau_j^2).$$

These two definitions are equivalent because of (4) and (5). The two sets of parameters $\gamma_1^2, \ldots, \gamma_m^2$ and $\tau_1^2, \ldots, \tau_m^2$ are related via:

$$\frac{4\gamma_j^2}{n^2} = \tau_j^2 \iff \frac{2\gamma_j}{n} = \tau_j.$$

This is because

$$\text{Re}(b_j) \sim N(0, \gamma_j^2) \implies \frac{2\text{Re}(b_j)}{n} \sim N\left(0, \frac{4\gamma_j^2}{n^2}\right).$$

The power spectrum is given by:

$$f(j/n) = \frac{2\gamma_j^2}{n} = \frac{n\tau_j^2}{2} \qquad \text{for } 0 < \frac{j}{n} < \frac{1}{2}.$$

4 Two key properties of the Spectrum Model

Consider Definition 2 of the spectrum model. We focus on two key properties of $\{y_t\}$. First, as noted in Lecture 13, the variance of y_t can be written as

$$\text{var}(y_t) = \sum_{j=1}^{m} \tau_j^2 = \frac{2}{n} \sum_{j=1}^{m} f(j/n) \approx 2 \int_0^{1/2} f(\omega) d\omega.$$

In other words, the variance of y_t is closely approximated by twice the integral of the spectral density $f(\omega)$ over the interval $[0, 0.5]$.

Next consider the covariance between y_t and y_{t+h}:

$$\text{cov}(y_t, y_{t+h}) = \sum_{j=1}^{m} \tau_j^2 \cos\left(\frac{2\pi jh}{n}\right) = \frac{2}{n}\sum_{j=1}^{m} f(j/n)\cos\left(\frac{2\pi jh}{n}\right) \approx 2\int_0^{1/2} f(\omega)\cos(2\pi\omega h)d\omega.$$

This shows that $\text{cov}(y_t, y_{t+h})$ can be nonzero, implying that y_t and y_{t+h} may be correlated. Consequently, the spectrum model naturally allows for dependence in y_t across different time points, illustrating its ability to capture and represent temporal correlations.

5 The case of even n

We assumed that n is odd (and $m = (n-1)/2$). If n is even, then $1/2$ becomes a Fourier frequency and $b_{n/2}$ becomes real (because $\sin(\pi t) = 0$ for all t). In this case, we can simply avoid working with $1/2$ by taking $m = (n-2)/2$ and using the model:

$$\text{Re}(b_j), \text{Im}(b_j) \overset{\text{i.i.d}}{\sim} N(0, \gamma_j^2) \qquad \text{for } j = 1, \ldots, m.$$

This will be equivalent to (6). Basically everything will stay the same as before (only difference is that $m = (n-2)/2$). Here we are essentially forcing $\gamma_{n/2} = 0$. One can try to also try to estimate $\gamma_{n/2}$ using $b_{n/2} \sim N(0, \gamma_{n/2}^2)$ (as was done in Lab 7) but this approach is more complicated.