STAT 153 & 248 - Time Series
Lecture Fifteen

Fall 2025, UC Berkeley

Aditya Guntuboyina

October 16, 2025

1 Smoothing the Periodogram

Consider a time series dataset $y_0, \dots, y_{n-1}$. One of most common tools in time data analysis is the periodogram. It is given by:

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

where b_j is the Discrete Fourier Transform (DFT) defined by

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

Let us assume that n is odd and $m = (n-1)/2$. Then the periodogram ordinate is defined for $j = 1, dots, m$.

The periodogram shows which frequency components contribute most strongly to the behaviour of the observed time series. It is useful for identifying the dominant cyclical patterns or periodicities within a dataset.

For most real datasets, the periodogram appears rough and noisy, especially when looking at its plot on the log-scale (i.e., when looking at the plot of log-periodogram). In such cases, researchers attempt to smooth the periodogram to get a clean summary of the trend in the periodogram without getting distracted by noise. What is a good way of smoothing the periodogram?

A nice way of smoothing is via the use of an appropriate model. Previously, we saw for estimating a smooth trend function, say μ_t, from observed time series data y_t, a natural model is:

$$y_t = \mu_t + \epsilon_t \qquad \text{with } \epsilon_t \overset{\text{i.i.d}}{\sim} N(0, \sigma^2).$$

We can try to use a similar model for the periodogram. Modeling $I(j/n)$ as a smooth trend function plus Gaussian noise does not make much sense for multiple reasons: (a) the periodogram is a positive quantity so using the normal distribution may not be a good idea, (b) the noise in the periodogram is more pronounced on the log-scale so we should probably use an additive noise model for the log-periodogram as opposed to the periodogram directly. In other words, we should use a multiplicative noise model for the periodogram.

The model that we shall use here is:

$$I(j/n) = f(j/n)\eta_j \qquad \text{for } j = 1, \dots, m.$$

where $f(j/n)$ is a smooth function that represents the trend in the periodogram, and

$$\eta_j \overset{\text{i.i.d}}{\sim} \frac{1}{2}\chi_2^2.$$

χ_2^2 is the chi-squared distribution with 2 degrees of freedom; its density is given by:

$$f_{\chi_2^2}(x) := \frac{1}{2}\exp(-x/2)I\{x > 0\}.$$

The density of $\chi_2^2/2$ is then given by:

$$f_{\chi_2^2/2}(x) = 2f_{\chi_2^2}(2x) = \exp(-x)I\{x > 0\}.$$

The above is the density of the Standard Exponential Distribution (the exponential distribution $Exp(\lambda)$ has density $\lambda e^{-\lambda x}I\{x > 0\}$; the above corresponds to $\lambda = 1$). Thus

$$\frac{\chi_2^2}{2} = Exp(1).$$

Our model can therefore also be written as:

$$I(j/n) = f(j/n)\eta_j \qquad \text{with } \eta_j \overset{\text{i.i.d}}{\sim} Exp(1). \tag{1}$$

This is a multiplicative noise model. We can rewrite it in additive form by taking logarithms:

$$\log I(j/n) = \log f(j/n) + \log \eta_j \qquad \text{with } \eta_j \overset{\text{i.i.d}}{\sim} Exp(1). \tag{2}$$

The noise in this additive representation is captured by the terms $\log \eta_j, 1 \le j \le n$. These variables do not have mean zero (from the internet $\mathbb{E} \log \eta_j \approx -0.5772$). Therefore $\log f(j/n)$ does the represent the mean of $\log I(j/n)$ (in fact, it is strictly larger than the mean of $\log I(j/n)$). Rather, $f(j/n)$ represents the mean of $I(j/n)$.

We shall refer to (1) or (2) as the Spectrum Model.

2 Power Spectral Density

Since the mean of the $Exp(1)$ distribution equals 1, the mean of $I(j/n)$ in the Spectrum Model (1) is given by $f(j/n)$. This quantity is known as the **power** of frequency j/n:

$$f(j/n) := \text{power of frequency } j/n = \text{mean of } I(j/n) \text{ in model (1)}.$$

If we plot the points $(j/n, f(j/n))$ for $j = 1, \dots, m$ 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).

Note that the power spectral density is not a probability density function in the sense that it does not integrate to one.

3 Spectrum Model from DFT

In (1) or (2), we formulated the spectrum model in terms of the periodogram or log-periodogram. We can also describe it in terms of the DFT. Specifically, in terms of the DFT $b_0, \dots, b_{n-1}$ of the data $y_0, \dots, y_{n-1}$, the model is given by:

$$\mathrm{Re}(b_j), \mathrm{Im}(b_j) \overset{\text{i.i.d}}{\sim} N(0, \gamma_j^2) \tag{3}$$

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.

By definition of the periodogram, we get

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

By the assumption (3), we get

$$\eta_j := \left(\left(\frac{\mathrm{Re}(b_j)}{\gamma_j}\right)^2 + \left(\frac{\mathrm{Im}(b_j)}{\gamma_j}\right)^2\right) \overset{\text{i.i.d}}{\sim} \chi_2^2.$$

Therefore the model (3) implies that

$$I(j/n) = \frac{2\gamma_j^2}{n}\eta_j \qquad \text{with } \eta_j \overset{\text{i.i.d}}{\sim} \frac{\chi_2^2}{2} = Exp(1).$$

The is exactly the same as (1) with the identification:

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

Because of this equivalence, we shall treat (3) as another formulation of the spectrum model.

4 Rewriting the Model in terms of y_t

The model (3) is written in terms of the DFT b_j. Because the original data can be written in terms of the DFT, we can convert (3) into a specification for the original data y_t. This will lead us to another model representation which is directly in terms of the original data y_t.

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

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

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

The right hand side of (4) involves complex numbers (b_j and $\exp(2\pi i j t/n)$). On the other hand, the left hand is the data y_t which is always real. Below we change the right hand side in (4) to make it consist of only real terms.

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

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

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

$$
y_t = \beta_0 + \sum_{j=1}^{m}\left(\beta_{1j}\cos\frac{2\pi jt}{n} + \beta_{2j}\sin\frac{2\pi jt}{n}\right) \tag{5}
$$

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

$$
\beta_0 = \frac{b_0}{n} \quad \beta_{1j} = \frac{2\mathrm{Re}(b_j)}{n} \quad \beta_{2j} = -\frac{2\mathrm{Im}(b_j)}{n} \tag{6}
$$

The formula (5) holds for every dataset $y_0, \dots, y_{n-1}$. As a result, the model (3) is equivalent to (5) with

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

The spectrum model therefore has the following three equivalent definitions:

- **Definition 1**: $\mathrm{Re}(b_1), \mathrm{Im}(b_1), \dots, \mathrm{Re}(b_m), \mathrm{Im}(b_m)$ are all independent with $\mathrm{Re}(b_j), \mathrm{Im}(b_j) \overset{\text{i.i.d}}{\sim} N(0, \gamma_j^2)$ for $j = 1, \dots, m$.
- **Definition 2**:

$$
y_t = \beta_0 + \sum_{j=1}^{m}\left(\beta_{1j}\cos\frac{2\pi jt}{n} + \beta_{2j}\sin\frac{2\pi jt}{n}\right) \tag{7}
$$

with $\beta_{11}, \beta_{21}, \beta_{12}, \beta_{22}, \dots, \beta_{1m}, \beta_{2m}$ all independent with

$$
\beta_{1j}, \beta_{2j} \overset{\text{i.i.d}}{\sim} N(0, \tau_j^2).
$$

- **Definition 3**: $I(j/n) = f(j/n)\eta_j$ with $\eta_j \overset{\text{i.i.d}}{\sim} Exp(1)$.

These two definitions are equivalent because of (5) and (6). The three sets of parameters $(\gamma_1^2, \ldots, \gamma_m^2)$ and $(\tau_1^2, \ldots, \tau_m^2)$ and $(f(1/n), \ldots, f(m/n))$ are related via:

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

which is because

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

and

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

5 Regularized Estimation

Parameter estimation is done by maximizing likelihood with a regularization penalty which ensures smoothness of the estimates (without the regularization penalty, we would overfit in the sense that the estimate of $f(j/n)$ would coincide with the periodogram $I(j/n)$).

To obtain the likelihood, we can use any of the three definitions of the spectrum model. If we use Definition 1 (i.e., (3)), the likelihood will be given by:

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

We can rewrite the above likelihood in terms of the periodogram (because $I(j/n) = |b_j|^2/n$) as follows:

$$\prod_{j=1}^{m} \frac{1}{\gamma_j^2} \exp\left(-\frac{nI(j/n)}{2\gamma_j^2}\right). \tag{8}$$

We can also use the definition (1) and write the likelihood directly using the exponential distribution. Note that the density of η_j is $\exp(-x)I\{x > 0\}$ and the density of $f(j/n)\eta_j$ is $\frac{1}{f(j/n)} \exp(-\frac{x}{f(j/n)})I\{x > 0\}$. Thus the likelihood (joint density of $I(j/n), 1 \leq j \leq m$) is:

$$\prod_{j=1}^{n} \frac{1}{f(j/n)} \exp\left(-\frac{I(j/n)}{f(j/n)}\right). \tag{9}$$

The above is equivalent to (8) (up to proportionality) because $f(j/n) = 2\gamma_j^2/n$.

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 $\{\gamma_j^2\}$.

Once we obtain estimates $\hat{\alpha}_j$ of α_j, we can convert them to estimates of γ_j via $\hat{\gamma}_j = \exp(\hat{\alpha}_j)$ and then to estimates of $f(j/n)$ via $\hat{f}(j/n) = 2\hat{\gamma}_j^2/n$.

6 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:

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

This will be equivalent to (7). 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 $b_{n/2} \sim N(0, \gamma_{n/2}^2)$ but this approach is slightly more complicated.