Spring 2015 Statistics 153 (Time Series) : Lectures Twenty Three and Twenty Four

Aditya Guntuboyina

21 and 23 April 2015

1 Periodogram and the Sample Autocovariance Function

Let us start by proving the following relation between the periodogram and the sample autocovariance function of a dataset $x_0, \dots, x_{n-1}$:

$$I(j/n) = \frac{|b_j|^2}{n} = \sum_{|h|<n} \hat{\gamma}(h) \exp\left(-\frac{2\pi ijh}{n}\right) \qquad \text{for } j = 1, \dots, n-1. \tag{1}$$

To see this, observe first, by the formula for the sum of a geometric series, that

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

In other words, if the data is constant i.e., $x_0 = \cdots = x_{n-1}$, then b_0 equals nx_0 and b_j equals 0 for all other j. Because of this, we can write:

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

Therefore, for $j = 1, \dots, n-1$, we write

$$\begin{aligned}
|b_j|^2 = b_j \bar{b}_j &= \sum_{t=0}^{n-1} \sum_{s=0}^{n-1} (x_t - \bar{x})(x_s - \bar{x}) \exp\left(-\frac{2\pi ijt}{n}\right) \exp\left(\frac{2\pi ijs}{n}\right) \\
&= \sum_{t=0}^{n-1} \sum_{s=0}^{n-1} (x_t - \bar{x})(x_s - \bar{x}) \exp\left(-\frac{2\pi ij(t-s)}{n}\right) \\
&= \sum_{h=-(n-1)}^{n-1} \sum_{t,s:t-s=h} (x_t - \bar{x})(x_{t-h} - \bar{x}) \exp\left(-\frac{2\pi ijh}{n}\right) \\
&= n \sum_{|h|<n} \hat{\gamma}(h) \exp\left(-\frac{2\pi ijh}{n}\right)
\end{aligned}$$

which proves (1).

The periodogram gives the strengths of the sinusoids of various frequencies that are present in the data. The formula (1) states that it can be computed from the sample autocovariance function $\hat{\gamma}(h), h = -(n-1), \dots, (n-1)$.

2 Process Representation by Sinusoids

Suppose $\{X_t\}$ is a stationary process with mean zero and variance σ^2. Process Representation by Sinusoids means that $\{X_t\}$ can be approximated by a process of the form

$$Y_t = \sum_{j=1}^{m} (A_j \cos(2\pi\lambda_j t) + B_j \sin(2\pi\lambda_j t)) \tag{2}$$

where $0 \le \lambda_1 < \cdots < \lambda_m \le 1/2$ are fixed constants and $A_1, B_1, A_2, B_2, \ldots, A_m, B_m$ are all uncorrelated random variables with mean zero and

$$\mathrm{var}(A_j) = \mathrm{var}(B_j) = \sigma_j^2.$$

Let $\sum_{j=1}^{m} \sigma_j^2 = \sigma^2$ so that the variance of the process $\{Y_t\}$ also equals σ^2. It turns out that $\{Y_t\}$ can approximate $\{X_t\}$ provided m is large enough and $\lambda_1, \ldots, \lambda_m$ and $\sigma_1^2, \ldots, \sigma_m^2$ are chosen appropriately.

In the process $\{Y_t\}$, the term $A_j \cos(2\pi\lambda_j t) + B_j \sin(2\pi\lambda_j t)$ is a sinusoid of frequency λ_j. Its strength is given by σ_j^2. Indeed, if σ_j^2 is large, then A_j and B_j will be large which means that this term will contribute more towards $\{Y_t\}$, and vice versa.

If $\{X_t\}$ is white noise, then Y_t with the choices

$$\lambda_j = \frac{j}{2m} \text{ and } \sigma_j^2 = \frac{\sigma^2}{m} \qquad \text{for } j = 1, \ldots, m$$

for a large m will be a very good approximation to $\{X_t\}$. This was demonstrated in the last class via simulations.

The important question here is: How does one choose λ_j and σ_j^2 for $j = 1, \ldots, m$ to approximate an arbitrary stationary process $\{X_t\}$? For example, how does one choose λ_j and σ_j^2 for $j = 1, \ldots, m$ to approximate an $AR(p)$ process? or an $MA(q)$ process? or an $ARMA(p, q)$ process? In order to answer these questions, we need to learn the notion of *spectral density*.

3 The Spectral Density

Given a dataset $x_0, \ldots, x_{n-1}$, we saw that the periodogram gives the strengths of the sinusoids of various frequencies that are present in the data. This periodogram was defined through the DFT but we saw that it can also be computed via

$$I(j/n) = \sum_{h:|h|<n} \hat{\gamma}(h) \exp\left(-\frac{2\pi i j h}{n}\right) \qquad \text{for } j = 1, \ldots, n-1.$$

This formula gives a simple way to extend the definition of periodogram to a stationary process $\{X_t\}$: for a stationary process $\{X_t\}$ with autocovariance function $\gamma_X(h)$, define

$$f(\lambda) := \sum_{h=-\infty}^{\infty} \gamma_X(h) \exp(-2\pi i \lambda h) \qquad \text{for } -1/2 \le \lambda \le 1/2. \tag{3}$$

This quantity is called the spectral density of $\{X_t\}$. Note that because $\gamma_X(h) = \gamma_X(-h)$, the spectral density f is symmetric and we really only need to look at its values on $[0, 1/2]$.

In analogy with the periodogram, the spectral density will give the strengths of sinusoids at various frequencies contributing to the stochastic process. It will soon be explained how the spectral density can be used to approximate any stationary process by sinusoids. But prior to that, it will be good to look at the following basic properties of the spectral density.

We have defined the spectral density in terms of the autocovariance function. It turns that the autocovariance function can also be obtained from the spectral density: To see this, just multiply both sides of (3) by $e^{2\pi i\lambda k}$ for a fixed k and integrate from $\lambda = -1/2$ to $\lambda = 1/2$ to get:

$$\gamma_X(k) = \int_{-1/2}^{1/2} e^{2\pi i\lambda k} f(\lambda) d\lambda \tag{4}$$

In other words, the autocovariance function and the spectral density provide equivalent information about the stationary process $\{X_t\}$. The identity (4) actually characterizes the spectral density in the sense that the only function f which satisfies (4) is the spectral density.

The identity (4) for $k = 0$ gives

$$\gamma_X(0) := \sigma^2 = \int_{-1/2}^{1/2} f(\lambda) d\lambda. \tag{5}$$

4 Back to Process Representation

Let us get back to the question of process representation. Let $\{X_t\}$ denote a stationary process with mean zero and variance σ^2. We want to approximate $\{X_t\}$ by a process of the form (2). We will take $\lambda_j = j/(2m)$ for $j = 1, \dots, m$ (i.e., the frequencies $\lambda_1, \dots, \lambda_m$ are equally spaced in $[0, 1/2]$). For choosing $\sigma_j^2, j = 1, \dots, m$, we use the spectral density of f in the following way.

(5) gives

$$\int_{-1/2}^{1/2} f(\lambda) d\lambda = \sigma^2.$$

Because f is symmetric around zero, this implies

$$\int_{0}^{1/2} f(\lambda) d\lambda = \frac{\sigma^2}{2}.$$

Approximating the integral on the left hand side above by a Riemann sum, we obtain

$$\frac{\sigma^2}{2} = \sum_{j=1}^{m} \int_{\lambda_{j-1}}^{\lambda_j} f(\lambda) d\lambda \approx \frac{1}{2m} \sum_{j=1}^{m} f(\lambda_j) \qquad \text{with } \lambda_j = \frac{j}{2m}.$$

We thus take

$$\lambda_j = \frac{j}{2m} \quad \text{and} \quad \sigma_j^2 = \frac{f(\lambda_j)}{m}.$$

With these choices, for large m, the process (2) will approximate the stationary process $\{X_t\}$ whose spectral density is f. For example, it will be shown later that the spectral density of an MA(1) process is given by

$$f(\lambda) = \sigma_Z^2 \left(1 + \theta^2 + 2\theta \cos(2\pi\lambda)\right) \qquad \text{for } -1/2 \le \lambda \le 1/2$$

where σ_Z^2 is the variance of the white noise process. Thus the process (2) with m large and

$$\lambda_j = \frac{j}{2m} \quad \text{and} \quad \sigma_j^2 = \frac{\sigma_Z^2}{m} \left(1 + \theta^2 + 2\theta \cos(2\pi\lambda_j)\right)$$

will approximate an MA(1) process. One can use this, for example, for simulating an MA(1) process.

When m equals ∞, the process (2) can be defined to make sense so that it has exactly the same spectral density f. But this requires stochastic integration and this is beyond the scope of this class.

Next we focus on the computation of the spectral density for ARMA processes. For this, we first need to learn about linear time-invariant filters.

5 Linear Time-Invariant Filters

A linear time-invariant filter uses a set of specified coefficients $\{a_j\}$ for $j = \dots, -2, -1, 0, 1, 2, 3, \dots$ to transform an input time series $\{X_t\}$ into an output time series $\{Y_t\}$ according to the formula:

$$Y_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j}.$$

The filter is determined by the coefficients $\{a_j\}$ which are often assumed to satisfy $\sum_{j=-\infty}^{\infty} |a_j| < \infty$.

Suppose that the input series $\{X_t\}$ is given by

$$X_t = \begin{cases} 1 & \text{if } t = 0 \\ 0 & \text{otherwise} \end{cases}$$

Such an $\{X_t\}$ is often called an *impulse function*. The output of the filter $\{Y_t\}$ can then be easily seen to be $Y_t = a_t$. For this reason the filter coefficients $\{a_j\}$ are often collectively known as the *impulse response function*.

The two main examples of linear time-invariant filters that we have seen so far are (1) the moving average filter which has the impulse response function: $a_j = 1/(2q+1)$ for $|j| \le q$ and $a_j = 0$ otherwise; and (2) Differencing which corresponds to the filter $a_0 = 1$ and $a_1 = -1$ and all other a_js equal zero. We have seen that these two filters act very differently; one estimates trend while the other eliminates it.

Suppose that the input time series $\{X_t\}$ is stationary with autocovariance function γ_X. What is the autocovariance function of $\{Y_t\}$? Observe that

$$\gamma_Y(h) := \operatorname{cov}\left(\sum_j a_j X_{t-j}, \sum_k a_k X_{t+h-k}\right) = \sum_{j,k} a_j a_k \operatorname{cov}(X_{t-j}, X_{t+h-k}) = \sum_{j,k} a_j a_k \gamma_X(h-k+j). \quad (6)$$

Note that the above calculation shows also that $\{Y_t\}$ is stationary.

Suppose now that the spectral density of the input stationary series $\{X_t\}$ is f_X. What then is the spectral density f_Y of the output $\{Y_t\}$?

Because the spectral density of $\{X_t\}$ equals f_X, we have

$$\gamma_X(h) = \int_{-1/2}^{1/2} e^{2\pi i h\lambda} f_X(\lambda) d\lambda.$$

We thus have from (6) that

$$\gamma_Y(h) = \sum_j \sum_k a_j a_k \int e^{2\pi i(h-k+j)\lambda} f_X(\lambda) d\lambda = \int e^{2\pi i h\lambda} f_X(\lambda) \left(\sum_j \sum_k a_j a_k e^{-2\pi i k\lambda} e^{2\pi i j\lambda}\right) d\lambda \quad (7)$$

Let us now define the function

$$A(\lambda) := \sum_j a_j e^{-2\pi i j\lambda} \qquad \text{for } -1/2 \le \lambda \le 1/2.$$

Note that this function only depends on the filter coefficients $\{a_j\}$. From (7) it clearly follows that

$$\gamma_Y(h) = \int e^{2\pi i\lambda h} f_X(\lambda) A(\lambda) \overline{A(\lambda)} d\lambda,$$

where, of course, $\overline{A(\lambda)}$ denotes the complex conjugate of $A(\lambda)$. As a result, we have

$$\gamma_Y(h) = \int e^{2\pi i\lambda h} f_X(\lambda) \left|A(\lambda)\right|^2 d\lambda.$$

This is clearly of the form $\gamma_Y(h) = \int e^{2\pi i\lambda h} f_Y(\lambda) d\lambda$. We therefore have

$$f_Y(\lambda) = f_X(\lambda) \left|A(\lambda)\right|^2 \qquad \text{for } -1/2 \le \lambda \le 1/2. \tag{8}$$

In other words, the action of the filter on the spectrum of the input is very easy to explain. It modifies the spectrum by multiplying it with the function $|A(\lambda)|^2$. Depending on the value of $|A(\lambda)|^2$, some frequencies may be enhanced in the output while other frequencies will be diminished.

This function $\lambda \mapsto |A(\lambda)|^2$ is called the *power transfer function* of the filter. The function $\lambda \mapsto A(\lambda)$ is called the *transfer function* or the *frequency response function* of the filter.

The spectral density is very useful while studying the properties of a filter. While the autocovariance function of the output series γ_Y depends in a complicated way on that of the input series γ_X, the dependence between the two spectral densities is very simple.

Example 5.1 (Power Transfer Function of the Differencing Filter). *Consider the Lag s differencing filter: $Y_t = X_t - X_{t-s}$ which corresponds to the weights $a_0 = 1$ and $a_s = -1$ and $a_j = 0$ for all other j. Then the transfer function is clearly given by*

$$A(\lambda) = \sum_j a_j e^{-2\pi ij\lambda} = 1 - e^{-2\pi is\lambda} = 2i\sin(\pi s\lambda)e^{-\pi is\lambda},$$

where, for the last equality, the formula $1 - e^{i\theta} = -2i\sin(\theta/2)e^{i\theta/2}$ is used. Therefore the power transfer function equals

$$\left|A(\lambda)\right|^2 = 4\sin^2(\pi s\lambda) \qquad \text{for } -1/2 \le \lambda \le 1/2.$$

To understand this function, we only need to consider the interval $[0, 1/2]$ because it is symmetric on $[-1/2, 1/2]$.

When $s = 1$, the function $\lambda \mapsto |A(\lambda)|^2$ is increasing on $[0, 1/2]$. This means that first order differencing enhances the higher frequencies in the data and diminishes the lower frequencies. Therefore, it will make the data more wiggly.

For higher values of s, the function $A(\lambda)$ goes up and down and takes the value zero for $\lambda = 0, 1/s, 2/s, \dots$. In other words, it eliminates all components of period s.

Example 5.2. *Now consider the moving average filter which corresponds to the coefficients $a_j = 1/(2q + 1)$ for $|j| \le q$. The transfer function is*

$$\frac{1}{2q+1}\sum_{j=-q}^{q} e^{-2\pi ij\lambda} = \frac{S_{q+1}(\lambda) + S_{q+1}(-\lambda) - 1}{2q+1},$$

where it may be recalled (Lecture 19) that

$$S_n(g) := \sum_{t=0}^{n-1} \exp(2\pi igt) = \frac{\sin(\pi ng)}{\sin(\pi g)} e^{i\pi g(n-1)}$$

. Thus

$$S_n(g) + S_n(-g) = 2\frac{\sin(\pi ng)}{\sin(\pi g)}\cos(\pi g(n-1)),$$

which implies that the transfer function is given by

$$A(\lambda) = \frac{1}{2q+1}\left(2\frac{\sin(\pi(q+1)\lambda)}{\sin(\pi\lambda)}\cos(\pi q\lambda) - 1\right),$$

This function only depends on q and can be plotted for various values of q. For q large, it drops to zero very quickly. The interpretation is that the filter kills the high frequency components in the input process.

6 Spectral Densities of ARMA Processes

Suppose $\{X_t\}$ is a stationary ARMA process: $\phi(B)X_t = \theta(B)Z_t$ where the polynomials ϕ and θ have no common zeroes on the unit circle. Because of stationarity, the polynomial ϕ has no roots on the unit circle.

Let $U_t = \phi(B)X_t = \theta(B)Z_t$. Let us first write down the spectral density of $U_t = \phi(B)X_t$ in terms of that of $\{X_t\}$. Clearly, U_t can be viewed as the output of a filter applied to X_t. The filter is given by $a_0 = 1$ and $a_j = -\phi_j$ for $1 \le j \le p$ and $a_j = 0$ for all other j. Let $A_\phi(\lambda)$ denote the transfer function of this filter. Then we have

$$f_U(\lambda) = |A_\phi(\lambda)|^2 f_X(\lambda). \tag{9}$$

Similarly, using the fact that $U_t = \theta(B)Z_t$, we can write

$$f_U(\lambda) = |A_\theta(\lambda)|^2 f_Z(\lambda) = \sigma_Z^2 |A_\theta(\lambda)|^2 \tag{10}$$

where $A_\theta(\lambda)$ is the tranfer function of the filter with coefficients $a_0 = 1$ and $a_j = \theta_j$ for $1 \le j \le q$ and $a_j = 0$ for all other j. Equating (9) and (10), we obtain

$$f_X(\lambda) = \frac{|A_\theta(\lambda)|^2}{|A_\phi(\lambda)|^2}\sigma_Z^2 \qquad \text{for } -1/2 \le \lambda \le 1/2.$$

Now

$$A_\phi(\lambda) = 1 - \phi_1 e^{-2\pi i\lambda} - \phi_2 e^{-2\pi i(2\lambda)} - \dots - \phi_p e^{-2\pi i(p\lambda)} = \phi(e^{-2\pi i\lambda}).$$

Similarly $A_\theta(\lambda) = \theta(e^{-2\pi i\lambda})$. As a result, we have

$$f_X(\lambda) = \sigma_Z^2 \frac{|\theta(e^{-2\pi i\lambda})|^2}{|\phi(e^{-2\pi i\lambda})|^2} \qquad \text{for } -1/2 \le \lambda \le 1/2.$$

Note that the denominator on the right hand side above is non-zero for all λ because of stationarity.

Example 6.1 (MA(1)). *For the MA(1) process: $X_t = Z_t + \theta Z_{t-1}$, we have $\phi(z) = 1$ and $\theta(z) = 1 + \theta z$. Therefore*

$$\begin{aligned} f_X(\lambda) &= \sigma_Z^2 \left|1 + \theta e^{2\pi i\lambda}\right|^2 \\ &= \sigma_Z^2 \left|1 + \theta\cos 2\pi\lambda + i\theta\sin 2\pi\lambda\right|^2 \\ &= \sigma_Z^2 \left[(1 + \theta\cos 2\pi\lambda)^2 + \theta^2 \sin^2 2\pi\lambda\right] \\ &= \sigma_Z^2 \left[1 + \theta^2 + 2\theta\cos 2\pi\lambda\right] \qquad \text{for } -1/2 \le \lambda \le 1/2. \end{aligned}$$

Check that for $\theta = -1$, the quantity $1 + \theta^2 + 2\theta\cos(2\pi\lambda)$ equals the power transfer function of the first differencing filter.

Example 6.2 (AR(1)). *For AR(1): $X_t - \phi X_{t-1} = Z_t$, we have $\phi(z) = 1 - \phi z$ and $\theta(z) = 1$. Thus*

$$f_X(\lambda) = \sigma_Z^2 \frac{1}{|1 - \phi e^{2\pi i\lambda}|^2} = \frac{\sigma_Z^2}{1 + \phi^2 - 2\phi\cos 2\pi\lambda} \qquad \text{for } -1/2 \le \lambda \le 1/2.$$

Example 6.3 (AR(2)). *For the AR(2) model: $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = Z_t$, we have $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$ and $\theta(z) = 1$. Here it can be shown that*

$$f_X(\lambda) = \frac{\sigma_Z^2}{1 + \phi_1^2 + \phi_2^2 - 2\phi_1(1 - \phi_2)\cos 2\pi\lambda - 2\phi_2\cos 4\pi\lambda} \qquad \text{for } -1/2 \le \lambda \le 1/2.$$