Spring 2015 Statistics 153 (Time Series) : Lecture Twenty Five

Aditya Guntuboyina

28 April 2015

1 Spectral Density

Given a stationary process $\{X_t\}$ with autocovariance function $\gamma_X(h)$, its spectral density is defined as

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

The spectral density is symmetric i.e., $f(\lambda) = f(-\lambda)$ so we only need to look at its values in $[0, 1/2]$.

The spectral density gives the strengths of sinusoids at various frequencies contributing to the stochastic process $\{X_t\}$.

The autocovariance function can be recovered from the spectral density via

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

For $h = 0$, this gives $\gamma_X(0) = \int_{-1/2}^{1/2} f(\lambda) d\lambda$.

2 Linear Time-Invariant Filters and how the spectral density changes from input to output

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 $\sum_{j=-\infty}^{\infty} |a_j| < \infty$. The filter coefficients $\{a_j\}$ are collectively known as the *impulse response function*.

We have seen many examples of linear time-invariant filters in this class. Two prominent examples are (1) the moving average filter which has the impulse response function: $a_j = 1/(2q+1)$ for $|j| \leq q$ and $a_j = 0$ otherwise; (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. Also, every stationary ARMA process can be thought of as the output of a linear time invariant filter with white noise input.

Suppose that the input time series $\{X_t\}$ is stationary with spectral density $f_X(\lambda)$. What then is the spectral density $f_Y(\lambda)$ of the output process $\{Y_t\}$? We have seen in the last class that $f_Y(\lambda)$ is given by

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

where

$$A(\lambda) := \sum_j a_j e^{-2\pi i j\lambda}.$$

Therefore, the action of the filter on the spectral density of the input is very easy to explain. It modifies the spectral density 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 2.1 (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*

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

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

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

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 $\lambda = 0, 1/s, 2/s, \ldots$. In other words, it eliminates all components of period s.

Example 2.2. *Now consider the moving average filter which corresponds to the coefficients $a_j = 1/(2q+1)$ for $|j| \le q$. The transfer function is*

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

where it may be recalled (Lecture 19) that

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

. Thus

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

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.

3 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 $1 \leq j \leq p$ and $a_j = 0$ for all other j. Let $A_\phi(\lambda)$ denote the transfer function of this filter. Then we have

$$f_U(\lambda) = |A_\phi(\lambda)|^2 f_X(\lambda). \tag{1}$$

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

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

where $A_\theta(\lambda)$ is the tranfer function of the filter with coefficients $a_0 = 1$ and $a_j = \theta_j$ for $1 \leq j \leq q$ and $a_j = 0$ for all other j. Equating (1) and (2), we obtain

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

Now

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

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

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

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

Example 3.1 (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*

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

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 3.2 (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*

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

Example 3.3 (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*

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

4 Nonparametric Estimation of the Spectral Density

Let $\{X_t\}$ be a stationary process with $\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty$. We have then seen that $\{X_t\}$ has a spectral density that is given by

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

Suppose now that we are given data $x_1, \dots, x_n$ from the process $\{X_t\}$. How then would we estimate $f(\lambda)$ without making any parametric assumptions about the underlying process? This is our next topic.

Why would we want to estimate the spectral density nonparametrically?

When we were fitting ARMA models to the data, we first looked at the sample autocovariance or autocorrelation function and we then tried to find the ARMA model whose theoretical acf matched with the sample acf. Now the sample autocovariance function is a nonparametric estimate of the theoretical autocovariance function of the process. In other words, we first estimated $\gamma(h)$ nonparametrically by $\hat{\gamma}(h)$ and then found an ARMA model whose $\gamma_{ARMA}(h)$ is close to $\hat{\gamma}(h)$.

If we can estimate the spectral density nonparametrically, we can similarly use the estimate for choosing a parametric model. We simply choose the ARMA model whose spectral density is closest to the non-parametric estimate.

Another reason for estimating the spectral density comes from the problem of estimating filter coefficients. Suppose that we know that two processes $\{X_t\}$ and $\{Y_t\}$ are related to each other through a linear time-invariant filter. In other words, $\{Y_t\}$ is the output when $\{X_t\}$ is the input to a filter. Suppose, that we do not know the filter coefficients however but we are given observations from both the input and the output process. The goal is to estimate the filter. In this case, a natural strategy is to estimate the spectral densities of f_X and f_Y from data and then to use $f_Y(\lambda) = f_X(\lambda)|A(\lambda)|^2$ to obtain an estimate of the power transfer function of the filter (to obtain an estimate of the transfer function itself, one needs to use cross-spectra). This is one of the applications of spectral analysis. We might not always be able to make parametric assumptions about $\{X_t\}$ and $\{Y_t\}$ so it makes sense to estimate the spectral densities nonparametrically.

Nonparametric estimation of the spectral density is more complicated than the nonparamtric estimation of the autocovariance function. The main reason is that the natural estimator does not work well.

Because of the formula (3) for the spectral density in terms of the autocovariance function $\gamma_X(h)$, a natural idea to estimate $f(\lambda)$ is to replace $\gamma_X(h)$ by its estimator $\hat{\gamma}(h)$ for $|h| < n$ (it is not possible to estimate $\gamma(h)$ for $|h| > n$). This would result in the estimator:

$$I(\lambda) = \sum_{h:|h|<n} \hat{\gamma}(h) e^{-2\pi i \lambda h} \qquad \text{for } -1/2 \le \lambda \le 1/2.$$

When $\lambda = j/n \in (0, 1/2]$, the above quantity is just the periodogram:

$$I(j/n) = \frac{|b_j|^2}{n} \qquad \text{where } b_j = \sum_t x_t \exp\left(-\frac{2\pi i j t}{n}\right)$$

Unfortunately, $I(\lambda)$ is not a good estimator of f_X. This can be easily seen by simulations. Just generate data from white noise and observe that the periodogram is very wiggly while the true spectral density is constant. The fact that $I(\lambda)$ is a bad estimator can also be verified mathematically. This we will see in the next class when we shall also study improved estimators for f_X.