Spring 2013 Statistics 153 (Time Series) : Lecture Twenty Three

Aditya Guntuboyina

18 April 2013

1 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

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

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 (1) 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 in the following way.

Suppose that the data x_t are generated from gaussian white noise with variance σ^2 (their mean is zero because they are white noise). What is the distribution of $|b_j|^2/n$ for $j/n \in [0, 1/2]$? Write

$$\begin{aligned}
\frac{|b_j|^2}{n} &= \frac{1}{n} \left| \sum_{t=0}^{n-1} x_t \exp\left(-\frac{2\pi i j t}{n}\right) \right|^2 \\
&= \frac{1}{n} \left| \sum_t x_t \cos(2\pi j t/n) - i \sum x_t \sin(2\pi j t/n) \right|^2 \\
&= \frac{1}{n} \left(A_j^2 + B_j^2 \right),
\end{aligned}$$

where

$$A_j = \sum_t x_t \cos(2\pi j t/n) \text{ and } B_j = \sum_t x_t \sin(2\pi j t/n).$$

If we also assume normality of $x_1, \dots, x_n$, then (A_j, B_j) are jointly normal with

$$\text{var} A_j = \sigma^2 \sum_{t=0}^{n-1} \cos^2(2\pi j t/n) \text{ and } \text{var} B_j = \sigma^2 \sum_{t=0}^{n-1} \sin^2(2\pi j t/n).$$

Also

$$\text{cov}(A_j, B_j) = \sigma^2 \sum_{t=0}^{n-1} \cos(2\pi j t/n) \sin(2\pi j t/n).$$

It can be checked that

$$\begin{aligned}
\sum_{t=0}^{n-1} \cos^2(2\pi j t/n) &= n \qquad \text{when } j \text{ is either 0 or } n/2 \\
&= n/2 \qquad \text{when } j \text{ is neither 0 nor } n/2.
\end{aligned}$$

and

$$\begin{aligned}
\sum_{t=0}^{n-1} \sin^2(2\pi j t/n) &= 0 \qquad \text{when } j \text{ is either 0 or } n/2 \\
&= n/2 \qquad \text{when } j \text{ is neither 0 nor } n/2.
\end{aligned}$$

and

$$\sum_{t=0}^{n-1} \cos(2\pi j t/n) \sin(2\pi j t/n) = 0.$$

Thus when j is neither 0 nor $n/2$ (recall that $0 \le j/n \le 1/2$), we have

$$\frac{\sqrt{2} A_j}{\sigma \sqrt{n}} \sim N(0, 1) \quad \text{and} \quad \frac{\sqrt{2} B_j}{\sigma \sqrt{n}} \sim N(0, 1)$$

which implies that

$$\frac{2}{n\sigma^2}A_j^2 \sim \chi_1^2 \quad \text{and} \quad \frac{2}{n\sigma^2}B_j^2 \sim \chi_1^2.$$

Also because they are independent, we have for $j/n \in (0, 1/2)$

$$\frac{2}{\sigma^2}I(j/n) = \frac{2|b_j|^2}{n\sigma^2} = \frac{2}{n\sigma^2}A_j^2 + \frac{2}{n\sigma^2}B_j^2 \sim \chi_2^2$$

or $I(j/n) \sim (\sigma^2/2)\chi_2^2$.

For $j = 0$ or $n/2$, we have $B_j = 0$ and $A_j \sim N(0, \sigma^2 n)$ which implies that $|b_j|^2/n \sim \sigma^2\chi_1^2$.

It is important to notice that the distribution of $I(j/n)$ does not depend on n. One can also check that (A_j, B_j) is independent of $(A_{j'}, B_{j'})$ for $j \neq j'$.

Therefore, when the data $x_1, \ldots, x_n$ are generated from the Gaussian White Noise model, the periodogram ordinates $I(j/n)$ for $0 < j \leq n/2$ are independent random variables having the distribution $(\sigma^2/2)\chi_2^2$ for $0 < j < n/2$ and $\sigma^2\chi_1^2$ for $j = n/2$. Because of this independence and the fact that the distribution does not depend on n, it should be clear that $I(\lambda)$ is not a good estimate of $f(\lambda)$.

We have done the above calculations for data from the gaussian white noise. For general ARMA processes, under some regularity conditions, it can be shown that when n is large, the random variables:

$$\frac{2I(j/n)}{f(j/n)}, \qquad \text{for } 0 < j < n/2$$

are approximately independently distributed according to the χ_2^2 distribution.

Note that because the χ_2^2 distribution has mean 2, the expected value of $I(j/n)$ is approximately $f(j/n)$. In other words, the periodogram is approximately unbiased. On the other hand, the variance of $I(j/n)$ is approximately $f^2(j/n)$. So, in the gaussian white noise case, for example, the variance of the periodogram ordinates is σ^4 which does not decrease with n. This and the approximate independence of the neighboring periodogram ordinates makes the periodogram very noisy and a bad estimator of the true spectral density.

2 Modifying the Periodogram for good estimates of the spectral density

2.1 Method One

The approximate distribution result allows us to write:

$$\frac{2I(j/n)}{f(j/n)} \approx 2 + 2U_j \qquad \text{for } 0 < j < n/2,$$

where $U_1, U_2, \ldots$ are independent, have mean zero and variance 1. In other words $\{U_j\}$ is white noise. Thus

$$I(j/n) = f(j/n) + U_j f(j/n) \qquad \text{for } 0 < j < n/2.$$

Therefore, we can think of $I(j/n)$ as an uncorrelated time series with a trend $f(j/n)$ that we wish to estimate. Our previous experience with trend estimation suggests that we do this by smoothing $I(j/n)$ with a filter, say the simple moving average filter:

$$\frac{1}{2m+1}\sum_{k=-m}^{m} I\left(\frac{j+k}{n}\right).$$

More generally, we can consider using unequal weights as well to yield estimators of the form:

$$\hat{f}(j/n) = \sum_{k=-m_n}^{m_n} W_n(k) I\left(\frac{j+k}{n}\right).$$

Note that if we take $m_n = 0$, we get back the periodogram. We can extend this definition of $\hat{f}$ to the entire interval $[0, 1/2]$ in the following way: For each $\lambda \in [0, 1/2]$, let $g(\lambda, n)$ denote the multiple of $1/n$ that is closest to λ. Define

$$\hat{f}(\lambda) = \hat{f}(g(\lambda, n)).$$

It can be shown that this estimator is consistent (i.e., it gets closer and closer to $f(j/n)$ as n becomes larger) provided:

$$m_n \to \infty \quad \text{and} \quad \frac{m_n}{n} \to 0 \tag{2}$$

as $n \to \infty$. One also needs the weights $W_n(k)$ to be symmetric: $W_n(k) = W_n(-k)$, nonnegative $W_n(k) \geq 0$, add up to 1: $\sum_{k=-m_n}^{m_n} W_n(k) = 1$ and their sum of squares to go to zero: $\sum_{k=-m_n}^{m_n} W_n^2(k) \to 0$ as $n \to \infty$. Note that all these conditions are satisfied for the simple moving average filter with m chosen as in (2).

If the above conditions are met, then, for $0 < \lambda < 1/2$, we have

$$\mathbb{E}\hat{f}(\lambda) \approx f(\lambda) \quad \text{and} \quad \mathrm{var}(\hat{f}(\lambda)) \approx \left(\sum_{k=-m_n}^{m_n} W_n^2(k)\right) f^2(\lambda).$$

When λ equals 0 or 1/2, the variance is twice the one given by the equation above. The expectation is still the same.

Also the covariance between $\hat{f}(\lambda_1)$ and $\hat{f}(\lambda_2)$ is approximately zero.

2.2 Method Two

Here is a slightly different way of coming up with estimators for the spectral density that are different from the periodogram. The periodogram is defined by:

$$I(\lambda) = \sum_{h:|h|<n} \hat{\gamma}(h) \exp(-2\pi i \lambda h) \qquad \text{for } -1/2 \leq \lambda \leq 1/2. \tag{3}$$

Note that the above formula involves the estimates of all the autocovariances $\gamma_X(h)$ for $|h| < n$. Now we know that from a sample of size n, it is impossible to come up with good estimates of $\gamma_X(h)$ for h close to n. This is often cited as a reason why the periodogram is not a good estimator. In light of this reason, a reasonable way to obtain better estimators is to truncate the sum on the right hand side of (3) by omitting $\hat{\gamma}(h)$ for h near n. In other words, we consider

$$\tilde{f}(\lambda) = \sum_{h:|h|\leq r} \hat{\gamma}(h) \exp(-2\pi i \lambda h).$$

If we assume that $r = r_n$ is a function of n such that $r_n \to \infty$ and $r_n/n \to 0$ as $n \to \infty$, then $\tilde{f}$ is a sum of $(2r+1)$ terms, each with a variance of about $1/n$. In this case, under regularity conditions, it can be shown that $\tilde{f}$ is a consistent estimator of f.

More generally, we can take

$$\tilde{f}(\lambda) = \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) \hat{\gamma}(h) \exp(-2\pi i \lambda h),$$

where $w(x)$ is a symmetric $w(x) = w(-x)$ function satisfying $w(0) = 1$, $|w(x)| \leq 1$ and $w(x) = 0$ for $|x| > 1$. This is sometimes called a *lag window spectral density estimator*.

2.3 Equivalence of the Two Methods

We shall now show that these two ways of improving the periodogram: by smoothing it and the lag window spectral density estimator are essentially the same. To see this, we first need an *inverse* relationship between $I(\lambda)$ and $\hat{\gamma}(h)$. We have defined $I(\lambda)$ as

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

It is possible to invert this formula to write $\hat{\gamma}(k)$ in terms of $I(\lambda)$. Fix an integer k with $|k| < n$ and multiply both sides of the above formula by $e^{2\pi i \lambda k}$. Integrating the resulting expression with respect to λ from $-1/2$ to $1/2$, we get

$$\int_{-1/2}^{1/2} e^{2\pi i \lambda k} I(\lambda) d\lambda = \sum_{h:|h|<n} \hat{\gamma}(h) \int_{-1/2}^{1/2} e^{2\pi i \lambda (k-h)} d\lambda = \hat{\gamma}(k).$$

This therefore implies

$$\hat{\gamma}(k) = \int_{-1/2}^{1/2} e^{2\pi i \lambda k} I(\lambda) d\lambda. \tag{4}$$

In other words, the function $I(\lambda)$ is the spectral density corresponding to the sample autocorrelation function. Using the formula (4), we can write the lag window spectral density estimator as

$$\begin{aligned}
\tilde{f}(\lambda) &= \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) \hat{\gamma}(h) e^{-2\pi i \lambda h} \\
&= \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) \int_{-1/2}^{1/2} e^{2\pi i \rho h} I(\rho) d\rho \; e^{-2\pi i \lambda h} \\
&= \int_{-1/2}^{1/2} I(\rho) \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) e^{2\pi i (\rho-\lambda) h} d\rho.
\end{aligned}$$

By the change of variable $\rho = \lambda + u$, we get

$$\tilde{f}(\lambda) = \int I(\lambda + u) \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) e^{2\pi i u h} du.$$

Letting

$$W(u) = \sum_{h:|h|\leq r} w\left(\frac{h}{r}\right) e^{2\pi i u h},$$

we get that

$$\tilde{f}(\lambda) = \int I(\lambda + u) W(u) du.$$

Thus the lag window spectral density estimator $\tilde{f}$ can also be thought of as obtained by smoothing the periodogram.