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

Aditya Guntuboyina

30 April 2015

1 Nonparametric Estimation of the Spectral Density

Let $\{X_t\}$ be a stationary process with $\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty$. The spectral density of $\{X_t\}$ is given by

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

Given data $x_1, \dots, x_n$ from the process $\{X_t\}$, we want to estimate the spectral density $f_X(\lambda)$ nonparametrically.

The most natural estimator is:

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

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. We saw this via simulations in the last class. 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

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

When $0 < j/n < 1/2$, it can be checked that

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

and

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

Thus for $j/n \in (0, 1/2)$, 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 $I(j/n) \sim (\sigma^2/2)\chi_2^2$.

It is important to notice here 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 $x_1, \dots, x_n$ are generated from the Gaussian White Noise model, the periodogram ordinates $I(j/n)$ for $0 < j < n/2$ are independent random variables having the distribution $(\sigma^2/2)\chi_2^2$ for $0 < j < n/2$. 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

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:

$$\hat{f}(j/n) := \sum_{k=-m}^{m} W_m(k) I\left(\frac{j+k}{n}\right)$$

The set of weights $\{W_m(k)\}$ is often referred to as a kernel or a spectral window. Simplest choice of $W_m(k)$ is

$$W_m(k) = \frac{1}{2m+1} \qquad \text{for } -m \le k \le m.$$

This window is called the *Daniell Spectral Window*.

3 Bias and Variance of the spectral density estimates

What are the bias and variance of $\hat{f}(j/n)$ as an estimator for $f(j/n)$? Let us start with the bias first. Because the periodogram is unbiased, we have

$$\mathbb{E}\hat{f}(j/n) = \sum_{k=-m}^{m} W_m(k) \mathbb{E} I\left(\frac{j+k}{n}\right) \approx \sum_{k=-m}^{m} W_m(k) f\left(\frac{j+k}{n}\right)$$

Let $\lambda = j/n$ for ease of notation. By a second order Taylor expansion around λ, we get

$$\mathbb{E}\hat{f}(\lambda) \approx \sum_{k=-m}^{m} W_m(k) \left(f(\lambda) + \frac{k}{n} f'(\lambda) + \frac{k^2}{2n^2} f''(\lambda) \right).$$

If the weights are such that $\sum_k W_m(k) = 1$ and $\sum_k k W_m(k) = 0$ (satisfied for the Daniell kernel for example), then we have

$$\mathbb{E}\hat{f}(\lambda) - f(\lambda) \approx \frac{f''(\lambda)}{2} \sum_{k=-m}^{m} \left(\frac{k}{n}\right)^2 W_m(k)$$

The quantity

$$\sqrt{\sum_{k=-m}^{m} \left(\frac{k}{n}\right)^2 W_m(k).}$$

is called the bandwidth of the kernel. Therefore the bias of $\hat{f}(\lambda)$ depends on the kernel only through the bandwidth. The smaller the bandwidth, the lower the bias. For the Daniell kernel, the bandwidth is given by the standard deviation of the discrete uniform distribution on $\{-m/n, -(m-1)/n, \ldots, (m-1)/n, m/n\}$ which is very close to the standard deviation of the continuous uniform distribution on $[-m/n, m/n]$ which equals:

$$\sqrt{\frac{(2m)^2}{12n^2}} \approx \sqrt{\frac{L^2}{12n^2}} = \frac{L}{n\sqrt{12}} \qquad \text{where } L = 2m+1.$$

Thus for the Daniell kernel, the bandwidth increases as m increases. This means that the bias of $\hat{f}(\lambda)$ increases with m.

Let us now consider the variance of $\hat{f}(\lambda)$. Because the variance of $I(k/n)$ is approximately $f^2(k/n)$ and because of the approximate independence of the periodogram ordinates, we have

$$var(\hat{f}(\lambda)) \approx \sum_{k=-m}^{m} W_n^2(k) f^2\left(\frac{j+k}{n}\right) \approx f^2(\lambda) \sum_{k=-m}^{m} W_n^2(k).$$

Thus the variance of $\hat{f}(\lambda)$ depends on the kernel through the quantity $\sum_{k=-m}^{m} W_n^2(k)$. For the Daniell kernel, this equals $1/L = 1/(2m+1)$. Thus as m increases, the variance of $\hat{f}(\lambda)$ decreases.

Observe the bias-variance tradeoff as m changes. For large m, we have low variance and high bias while for small m, we have low bias and high variance.

4 Approximate Confidence Intervals for $f(j/n)$

Recall that 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.

Therefore, approximately

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

This would allow us to approximate the distribution of $\hat{f}(j/n)$ in the following way:

$$2(2m+1)\frac{\hat{f}(j/n)}{f(j/n)} \sim \chi^2_{2(2m+1)}.$$

If $\chi^2_{2(2m+1)}(\alpha/2)$ and $\chi^2_{2(2m+1)}(1-\alpha/2)$ satisfy

$$\mathbb{P}\left\{\chi^2_{2(2m+1)}(\alpha/2) \leq \chi^2_{2(2m+1)} \leq \chi^2_{2(2m+1)}(1-\alpha/2)\right\} = 1-\alpha,$$

then we conclude that approximately

$$\mathbb{P}\left\{\chi^2_{2(2m+1)}(\alpha/2) \leq 2(2m+1)\frac{\hat{f}(j/n)}{f(j/n)} \leq \chi^2_{2(2m+1)}(1-\alpha/2)\right\} \approx 1-\alpha.$$

This would lead to the following confidence interval for $f(j/n)$ of level approximately $1-\alpha$:

$$2(2m+1)\frac{\hat{f}(j/n)}{\chi^2_{2(2m+1)}(1-\alpha/2)} \leq f(j/n) \leq 2(2m+1)\frac{\hat{f}(j/n)}{\chi^2_{2(2m+1)}(\alpha/2)}.$$

Note that the width of this confidence interval for $f(j/n)$ depends on j. But the corresponding confidence interval for $\log f(j/n)$ has a width that does not depend on j.