Spectral Estimation

Overview

1. Representation of the discrete Fourier transform
2. Periodogram
3. Leakage and tapering
4. Consistent estimators
5. Multi-taper estimators

Discrete Fourier transform

Spectral representation The representation for a stationary process (mean zero) $\{X_t\}$ is

$$X_t = \int_{-1/2}^{1/2} e^{2\pi\imath\,\xi t} Z(d\xi)$$

where the random measure Z has orthogonal increments such that (assuming the spectral density $f(\lambda)$ is well-defined)

$$\mathbb{E}\, Z(d\xi)\overline{Z(d\lambda)} = \begin{cases} f(\xi) & \xi = \lambda \\ 0 & \xi \neq \lambda \end{cases}$$

Fourier transform of the process gives for any frequency λ (with $J_{n,j} = J_n(\lambda_j)$ where the Fourier frequency $\lambda_j = j/n$):

$$\begin{aligned} J_n^x(\lambda) &= \frac{1}{n}\sum_{t=0}^{n-1} X_t e^{-2\pi\imath\,\lambda t} \\ &= \frac{1}{n}\sum_{t=0}^{n-1}\left(\int_{-1/2}^{1/2} e^{2\pi\imath\,t\xi} Z(d\xi)\right) e^{-2\pi\imath\,\lambda t} \\ &= \int \left(\frac{1}{n}\sum_{t=0}^{n-1} e^{-2\pi\imath\,t(\lambda-\xi)}\right) Z(d\xi) \\ &= \int \underbrace{e^{2\pi\imath\,(\lambda-\xi)\frac{n}{2}} D_n(\lambda-\xi)} Z(d\xi) \end{aligned}$$

$$= \int_{-1/2}^{1/2} Q_n(\lambda - \xi) Z(d\xi) \qquad (1)$$

The function Q_n is the Dirichlet kernel D_n with a complex multiplier. The norm of Q_n in (1) is $|Q_n(\lambda)| = D_n(\lambda)$. The observed transform is a "blurred" version of the underlying random measure $Z(\lambda)$.

$$D_n(\lambda) = \frac{\sin(n\pi\lambda)}{n \sin(\pi\lambda)} \qquad (2)$$

The maximum value of the Dirichlet kernel $\max_\lambda D_n(\lambda) = 1$ at $\lambda = 0$, and $D_n(\lambda_j) = 0$ at the associated Fourier frequencies $\lambda_j = j/n$ which clearly depend on n.

Periodogram

Definition The periodogram is (up to the choice of a constant scaling factor) the norm of the discrete Fourier transform. In its "raw" state, the periodogram is unbiased for the spectral density, but it is *not* a consistent estimator of the spectral density. The periodogram is defined as

$$\begin{aligned} I_n(\lambda) &= \frac{1}{n} \left| \sum_{t=1}^{n} X_t e^{-2\pi\imath\, t\lambda} \right|^2 \\ &= n\, |J_n(\lambda)|^2. \qquad (3) \end{aligned}$$

All phase (relative location/time origin) information is lost. The periodogram would be the same if all of the data were circularly rotated to a new time origin, as though the observed data series were perfectly periodic with period n. (Take a moment to think about the consequence of this translation invariance.)

Expected value The orthogonal increments process $Z(\lambda)$ defined in the spectral representation introduced in the prior lecture implies

$$\begin{aligned} \mathbb{E}\, I_n(\lambda) &= n\, \mathbb{E}\, |J_n(\lambda)|^2 \\ &= \int_{-1/2}^{1/2} n\, |D_n(\lambda - \xi)|^2 f(\xi) d\xi \\ &= \int_{-1/2}^{1/2} K_n(\lambda - \xi) f(\xi) d\xi\ , \qquad (4) \end{aligned}$$

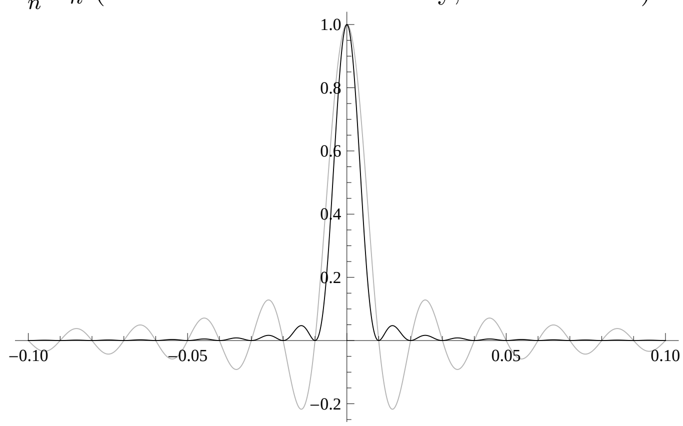
assuming that $dF(\lambda)/d\lambda = f(\lambda)$ so that the s.d.f. exists. The smoothing kernel K_n in this expression is n times the square of the Dirichlet kernel; it is known as *Fejer's kernel*:

$$K_n(\lambda) = n|D_n(\lambda)|^2 = \frac{\sin^2(n\,\pi\,\lambda)}{n\sin^2(\pi\,\lambda)}\ . \tag{5}$$

Unlike the Dirichlet kernel, Fejer's kernel has the delta-function property in the sense that for suitable functions g,

$$\lim_n \int_{-\pi}^{\pi} K_n(\lambda-\xi)g(\xi)d\xi = g(\lambda)\ .$$

Here's a plot of the Dirichlet kernel D_n (gray) and $1/n$ times Fejer's kernel $\frac{1}{n}K_n$ (so both are scaled similarly; with $n = 100$).



Hence, the periodogram $I_n(\lambda)$ shown in equation (5) is asymptotically unbiased. This is of little relevance in practice, however, due to the problems of the size of its variance and the presence of leakage discussed next.

Variance Although asymptotically unbiased, I_n is not a consistent estimator of its mean since its variance does not go to zero as $n \to \infty$. Using the results from harmonic regression ($j \neq 0, \frac{n}{2}$), we know that

$$\text{Regr SS}(\lambda_j) = \frac{n}{2}(A_j^2 + B_j^2) = 2I_n(\lambda_j).$$

with expected value $f(\lambda_j)$ where

$$J_{n,j} = \tfrac{1}{2}\,(A_j - \imath B_j) \Rightarrow n|J_j|^2 = \tfrac{n}{4}(A_j^2 + B_j^2) \tag{6}$$

The squares of the J_j get divided by an extra 2 since there are twice as many of these (the redundancy in the complex values). In the white noise case, it is clear that the coefficients A_j, B_j are independent and asymptotically normal with variance $\text{Var}(A_j) = \text{Var}(B_j) = \sigma^2 \underline{2/n} (j \neq 0, n/2)$. Hence, for $j = 1, \ldots, n/2$,

$$I_n(\lambda_j) = \frac{n}{4\sigma^2}(A_j^2 + B_j^2) \sim \frac{\chi_2^2}{2} f(\lambda_j) = f(\lambda_j) \times \text{ Exponential r.v. }.$$

For $j = 0$ or $n/2$, the r.v. is χ_1^2 with no factor of 1/2. Hence, for white noise, the periodogram coordinates are independent with equal variance. The same is asymptotically true in general for stationary processes.

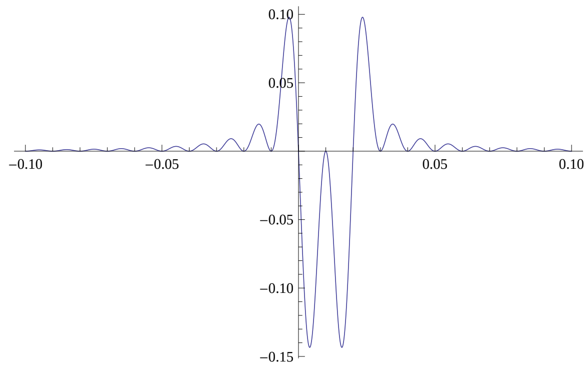
Covariances of the Fourier transform The spectral representation yields the covariances of the terms in the Fourier transform:

$$n\mathbb{E}\, J_{n,j}\overline{J}_{n,k} = \int_{-1/2}^{1/2} nQ_n(\lambda_j - \xi)\overline{Q_n(\lambda_k - \xi)} f(\xi) d\xi$$

So long as the spectral density $f(\xi)$ is smooth, the product of the kernels Q_n at the distinct Fourier frequencies is approximately zero. In fact, the integral is *exactly* zero if the spectral density is constant. To see that, use the definition of Q_n as a sum of complex exponentials:

$$\begin{aligned} n\int_{-1/2}^{1/2} Q_n(\lambda_j - \xi)\overline{Q_n(\lambda_k - \xi)} d\xi &= \frac{1}{n}\sum_{t,s} e^{2\pi\imath\,(\lambda_j t - \lambda_k s)} \int e^{2\pi\imath\,\xi(t-s)} d\xi \\ &= \frac{1}{n}\sum_{t} e^{2\pi\imath\,(\lambda_j - \lambda_k)t} \\ &= \begin{cases} 1 & \text{if } j = k \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

As long as f changes slowly relative to the oscillations of the Dirichlet kernel function, the covariances between the Fourier transform at the Fourier frequencies are approximately zero. Here's a picture of the product of two Dirichlet kernels $nD_n(\lambda)$ and $nD_n(\lambda - \frac{2}{n})$ (with $n = 100$):



Hence, under normality, the lack of correlation among the $J_{n,j}$s means that these are asymptotically independent (*i.e.*, the covariance in the asymptotic distribution is zero).The exponential multiplier, combined with the fact that adjacent values of the periodogram are uncorrelated, explains why the periodogram shows so much irregular fluctuations

Harmonic regression The connection to harmonic regression shows that the asymptotic distribution of the periodogram at Fourier frequencies $I_n(\lambda_j)$ is a multiple of a χ^2_2 random variable — an exponential random variable (except at the extremes $\lambda = 0, \frac{1}{2}$).

Summary: Distribution of periodogram The periodogram ordinates are thus roughly distributed as chi-squared with 2 degrees of freedom, independently of each other, times the spectral density:

$$I(\lambda_j) \sim f(\lambda_j)\left(\tfrac{1}{2}\chi^2_2\right) \quad j \neq 0, n/2.$$

Alternative derivation The S&S text takes a different approach to getting the properties of the periodogram. You need not rely upon the spectral representation to obtain the expected value of I_n. You can work directly from the properties of the covariances. This approach also has the advantage of reminding you that the covariances and spectral density (either in a sample or the population) form a Fourier transform pair. Proceeding directly,

$$\begin{aligned} \mathbb{E}\, I_n(\lambda) &= \mathbb{E}\,\frac{1}{n}\left|\sum_{t=1}^{n} X_t e^{2\pi\imath\,\lambda t}\right|^2 \\ &= \frac{1}{n}\sum_{t,s} \left(\mathbb{E}\, X_t X_s\right) e^{2\pi\imath\,\lambda(t-s)} \end{aligned}$$

$$= \frac{1}{n}\sum_{t,s}\gamma_{t-s}e^{2\pi\imath\,\lambda(t-s)}$$
$$= \sum_{|r|<n}\left(1-\frac{|r|}{n}\right)\gamma_r e^{2\pi\imath\,\lambda r}.$$

If the covariances are absolutely summable $\sum|\gamma(j)| < \infty$, this last expression has the same limit as $\sum_{|r|<n}\gamma_r e^{2\pi\imath\,\lambda r}$ as $n \to \infty$, which means that $\mathbb{E}\,I_n(\lambda_j) \to f(\lambda_j)$.

Consistency The periodogram is thus not a consistent estimator of the spectral density function. Some averaging *must take place* in order for one to obtain consistency. As n increases, we observe estimates that are more tightly spaced in frequency, but nonethess roughly independent.

To obtain a consistent estimator of the spectral density, we begin with n observations and convert them into n values J_j (noting $J_{n,j} = \overline{J}_{n,n-j}$ for real data), obtaining $n/2+1$ variance components (for n even), one at each Fourier frequency. As data are added, the number of estimates increases (the spacing $1/n$ between Fourier frequences diminishes). Since the estimates are approximately uncorrelated, the periodogram looks very "rough."

Logs and plots Since I_n is a random multiple of the spectral density, its variance depends on $f(\lambda)$. By taking logs, one breaks this tie of level and variance. Plots of logs of spectral estimates have roughly constant variance regardless of the level of the spectral density $f(\lambda)$. Hence, by default, **R** graphs spectral estimates on a log scale (some prefer the *decibel* scale, which is $10\log_{10}$).

Leakage and Tapering

Heuristic Think of the observed data X_t as a segment of an infinitely long stochastic process denoted $\tilde{X}_t$,

$$X_t = \begin{cases} \tilde{X}_t, & t = 1,2,\ldots,n \\ 0 & \text{otherwise}\ . \end{cases}$$

We can also write this as using a “data window” W_t times the complete series,

$$X_t = W_t\,\tilde{X}_t \quad \text{where} \quad W_t = \begin{cases} 1 & t = 1, 2, \ldots, n \\ 0 & \text{otherwise} . \end{cases} \tag{7}$$

The previous lecture shows that a spike and a constant form a Fourier transform pair. Spikes on the time/frequency scale become flat features on the frequency/time scale. The default data window W_t — the “boxcar” — have sudden transitions, such as the jump from 0 at $t = 1$ to 1 at $t = n$. These sudden changes produce the Dirichlet kernel in $J(\lambda)$ which has a very slow decay — a relatively flat function. When convolved with the underlying spectrum, the sidelobes of the Dirichlet kernel allow power to “leak” from other frequencies.

To obtain a more rapid decay, and thus less leakage, the data weights need to rise more slowly. One example is the cosine bell in which the weights are proportional (essentially) to the values of the cosine function on $-\pi/2$ to $\pi/2$. (This is the default R taper; see the software documentation. Note: an analogous heuristic argument suggests how one passes from continuous to discrete time using a device called a “Dirichlet comb.”)

Data windows and tapers In general, spectral analysis works with the product $Y_t = W_t X_t$ with a data window $W_t = W(t/n)$. The function W defined on $[0, 1]$ and zero elsewhere is known as a data window or taper. Typically, $W(t)$ is normalized so that $\|2\|W = \int_0^1 W(t)^2 dt = 1$ to keep things properly scaled. The boxcar function is the simplest data window: so simple that we often forget about it.

Leakage The observed transform $J_n(\lambda) = \int Q_n(\lambda - \xi) Z(d\xi)$ is a convolution of Q_n with the unobserved random measure Z. Since $Q_n(\lambda)$ is large for λ away from 0, variance from other frequencies affects the value of $J(\lambda)$. This undesired property is known as leakage. For an example of leakage, refer back to the R code for the previous lecture. The FT of a sinusoid at a Fourier frequency λ_j produces a single spike at index j in the transform. For frequency $\lambda \neq \lambda_j$, however, the transform of the sinusiod is non-zero at numerous frequencies near λ. Leakage does not occur with a sinusoid in the data at a Fourier frequency since the zeros of the associated Dirichlet kernel are located at just the right locations to cancel out. The cosine/sine terms at Fourier frequencies are uncorrelated; if the harmonic component in the data lies at some other frequency, it has correlation with many of the sinusoids at the Fourier frequencies.

Reducing leakage Because $Y_t = w_t X_t$ is a product, we can treat the Fourier transform as the convolution of the data window with the FT of the process,

$$J_n^y(\lambda) = \int_{-1/2}^{1/2} H_n(\lambda - \xi) dZ(\xi) \tag{8}$$

where the new kernel H_n is the transform of the weights,

$$H_n(\lambda) = \frac{1}{n} \sum_t w_t e^{-it\lambda} \ . \tag{9}$$

Hence, W_t is chosen so as to produce smaller sidelobes than the Dirichlet kernel in (1), though typically one must get more broad peaks (lower resolution) due to the Heisenberg uncertainty principle.

Consistent estimators

Local averaging An early type of spectral estimators smoothes the periodogram by local averaging. (Aside: Smoothing methods so common in regression modeling originated from the analogous methods in smoothing spectra density estimates.) Known as Daniell estimators, these are defined as

$$\hat{f}(\lambda_k) = \sum_{j=-d}^{d} \frac{I_n(\lambda_{k-d})}{2d+1}.$$

(The endpoints are clearly a problem.) One obtains this estimator in R by setting the option `spans` in `spec.pgram` to the length of the moving average, $2d + 1$.

Properties of Daniell estimators Smoothing trades bias for variance. Unless the spectral density f is constant or linear over the interval $[\lambda_{k-d}, \lambda_{k+d}]$ (which has length $4\,d/n$), smoothing produces some bias. On the other hand, smoothing reduces variance. Since the periodogram ordinates are approximately independent,

$$\text{Var}\hat{f}(\lambda_k) \approx \text{Var}\left(\sum_{j=-d}^{d} \frac{\chi_2^2}{2} f(\lambda_{k-j})\right) \approx \frac{f(\lambda_k)^2}{2d+1} \, ,$$

if we pretend that f is constant over the relevant interval. Thus as long as the window width d grows with the sample size n (though at a slower rate), one gets a consistent estimator. (For example, let d grow at the same rate as n so that a fixed proportion of the data fall in the interval.) Also note that smoothing introduces correlation into the spectral estimator. Thus, the estimator is smoother but peaks become more blurred.

Multi-taper estimates These more recent estimates

1. Focus most of variance in a narrow band and avoid leakage.
2. Avoid smoothing periodograms (smoothing using a moving average is "slow" compared to the Fourier transform).

The averaging is done in a different way using prolate spheroidal functions rather than cosine bell data tapers. One of the first applications (in climate) is D.J. Thompson (1990), "Time series analysis of Holocene climate data," *Phil. Trans. Royal Soc. London A* 330, 601-616, or the text of Percival and Walden (1993), *Spectral Analysis for Physical Applications.*

Prolate spheroidal functions How do we get the least amount of leakage under the constraint that the sum of squared weights is 1 (to avoid the trivial case of all weights zero)? Let H denote the transform of the weights w_t. The goal is to maximize the concentration of variance coming from frequencies near zero, say the interval $[-\delta, \delta]$. Specifically, maximize the ratio

$$R(\delta) = \frac{\int_{-\delta}^{\delta} |H(\lambda)|^2 d\lambda}{\int |H(\lambda)|^2 d\lambda}$$

over choices of H. The functions that maximize this ratio are known as the prolate spheroids.

Eigenvectors With the chosen definitions, the denominator of the ratio $R(\delta)$ reduces to our initial constraint on the weights,

$$\begin{aligned}\int_{-1/2}^{1/2} |H(\lambda)|^2 d\lambda &= \int_{-1/2}^{1/2} \left| \frac{1}{n} \sum_t w_t e^{-2\pi\imath\, t\lambda} \right|^2 d\lambda \\ &= n^{-2} \sum_{s,t} w_t w_s \int e^{-2\pi\imath\,(t-s)\lambda} d\lambda \\ &= n^{-2} \sum_t w_t^2 = \frac{1}{n^2} \;,\end{aligned}$$

and the numerator becomes the quadratic form

$$\begin{aligned}\int_{-\delta}^{\delta} |H(\lambda)|^2 d\lambda &= n^{-2} \int_{-\delta}^{\delta} |\sum_t w_t e^{-2\pi\imath\, t\lambda}|^2 d\lambda \\ &= n^{-2} \sum_{s,t} w_t w_s \int_{-\delta}^{\delta} e^{-2\pi\imath\,(t-s)\lambda} d\lambda \\ &= n^{-2} \sum_{s,t} w_t w_s \int_{-\delta}^{\delta} \cos 2\pi\lambda(t-s) + i \sin 2\pi\lambda(t-s) d\lambda \\ &= n^{-2} \sum_{s,t} w_t w_s \frac{2 \sin \delta(t-s)}{t-s} \\ &= n^{-2} w' M w \;,\end{aligned}$$

where M is the matrix with elements $2 \sin d(t-s)/(t-s)$ in position (t, s). Thus maximizing R is equivalent to

$$\max w' M\, w \qquad w'w = 1 \;.$$

This is the classical eigenvector problem. Using the resulting data weights gives the least leakage possible under this constraint.

Averaging Rather than smooth the periodogram, compute the transforms using the second, third, etc eigenvectors of the matrix M. Since these vectors are orthogonal, one can average the differently tapered estimates to obtain consistency rather than by smoothing the periodogram directly. One obtains very high resolution without sacrificing leakage protection or consistency.