Spectral Estimation

Overview

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

Discrete Fourier transform

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

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

The observed transform is a "blurred" version of the underlying random measure $Z(\lambda)$. The norm of Q_n in (1) is $|Q_n(\lambda)| = D_n(\lambda)$, the Dirichlet kernel, (Note: I put the π into the kernel in keeping with the use of frequencies from -1/2 to 1/2.)

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

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

Periodogram

Norm The periodogram is (up to optional scaling) the norm of the discrete Fourier transform, and in its "raw" state, is not a consistent estimator of the spectral density. The periodogram is defined as

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

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 rotation invariance.)

Expected value The orthogonal increments process $Z(\lambda)$ implies

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

assuming that $dF(\lambda)/d\lambda = f(\lambda)$. The smoothing kernel in this expression is the square of the Dirichlet kernel, a multiple of *Fejer's kernel*. Fejer's kernel is

$$K_n(\lambda) = n|D_n(\lambda)|^2 = \frac{\sin^2(n\,\pi\,\lambda)}{n\sin^2(\pi\,\lambda)}\ . \tag{4}$$

Fejer's kernel acts like a delta function in that for suitable functions g,

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

Hence, the periodogram $I_n(\lambda)$ is asymptotically unbiased. However, this is of little value in practice due to the problems of leakage and variance discussed next.

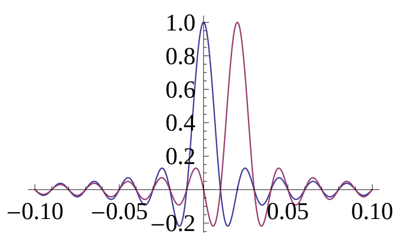
Covariances The spectral representation also yields the covariances of the terms in the Fourier transform:

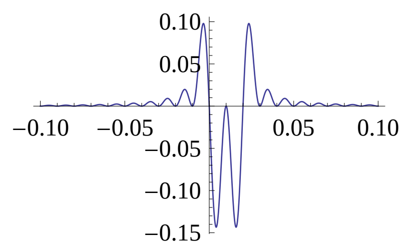
$$n\mathbb{E}\, J_{n,j}\overline{J}_{n,k} \;=\; \int_{-1/2}^{1/2} nQ_n(\lambda_j-\xi)\overline{Q_n(\lambda_k-\xi)} f(\xi) d\xi$$

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:

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

So, as long as f changes slowly relative to the oscillations of the kernel function, the covariances between the frequencies is approximately zero. Here's a picture of two Dirichlet kernels and their product (with $n=100$):





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 You need not rely upon the spectral representation to obtain the expected value of I_n. 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 converges to the same limit as $\sum_{|r|<n}\gamma_r e^{2\pi\imath\,\lambda r}$ as $n\to\infty$, which means that $\mathbb{E}\,I_n(\lambda_j)\to f(\lambda_j)$.

Variance 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 $\lambda = 0, \frac{1}{2}$). The exponential multiplier, combined with the fact that adjacent values of the periodogram are uncorrelated, explains why the periodogram shows so much irregular fluctuations. Though unbiased, it is not consistent 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

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

with expected value $f(\lambda_j)$. Since the coefficients A_j, B_j are asymptotically independent (the regressors are orthogonal and asymptotically normal)

$$I_n(\lambda_j) = \frac{n}{2}(A_j^2+B_j^2) \sim \frac{\chi^2_2}{2}f(\lambda_j) = f(\lambda_j)\times \text{ Exponential r.v. }.$$

For $j=0$ or $n/2$, the r.v. is χ^2_1 with no factor of 1/2. The ordinates of the periodogram are asymptotically independent.

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 Lecture 17 shows that a spike and a constant form a Fourier transform pair. Sudden changes on the time/frequency scale become flat features on the frequency/time scale. The default sampling weights — the "boxcar" — have sudden transitions, such as the jump from 0 at $t=-1$ to 1 at $t=0$ (first observation). This sudden change produces 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 bring power (*i.e.*, "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.)

Data windows and tapers Rather than analyze $\{X_t\}$ directly, consider the product $Y_t = w_t X_t$ with a set of weights $w_t = W(t/n)$. The bounded variation function W defined on $[0,1]$ and zero elsewhere is known as a data window or *taper*. Typically, the weight function is normalized so that $\|W\| = \int_0^1 W(t)^2 dt = 1$ to keep things properly scaled.

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. 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 the weighted data $Y_t = w_t X_t$ form a product, the transform becomes the convolution,

$$J_n^y(\lambda) = \int_{-1/2}^{1/2} H_n(\lambda - \xi)dZ(\xi) \tag{5}$$

where the new kernel is the transform of the weights,

$$H_n(\lambda) = \frac{1}{n}\sum_t w_t e^{-it\lambda} \ . \tag{6}$$

Hence, w_t is chosen so as to produce smaller sidelobes than the Dirichlet kernel in (1), though typically one gets more broad peaks (less resolution).

Consistent estimators

Local averaging An early type of spectral estimators smoothes the periodogram by local averaging. 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 $4\pi d/n$), smoothing produces some bias. On the other hand, smoothing reduces variance. Since the periodogram ordinates are approximately independent,

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

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.

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.