Fourier Time Series

Methods of Power Spectral Estimation

1. **Discrete Fourier Transform**

2. **Periodogram**

3. **Fast Fourier Transform**

4. **Tapers**

   Fourier transform is defined for a function on a finite interval and the function needs to be periodic. But with the real data set, this requirment is not met as the data end suddenly at t=0 and t=T and can have discontinuities. This discontinuity introduces distortions (known as Gibbs phenomenon) in fourier transform and generates false high frequency in the spectrum. Tapering (using data window) is used to reduce these artificial presence. The data $y = f(t)$ is multiplied by a taper function $g(t)$ which is a simple, slowly varying function, often going towards zero near the edges. Some of the popular tapers are:

   - Sine taper $g(t) = sin(\pi t/T)$
   - Hanning (offset cosine) taper $g(t) = \frac{1}{2}(1 - cos(2\pi t/T))$
   - Hamming taper $g(t) = 0.54 - 0.46 \ cos(2\pi \ t/T)$
   - Parzen or Bartlett (triangle) window $g(t) = 1 - (t - T/2)/(T/2)$
   - Welch (parabolic) window $g(t) = 1 - (t - T/2)^2/(T/2)^2$
   - Daniell (untapered or rectangular) window $g(t) = 1$

   The frequency resolution in the spectrum of the tapered data is degraded. If the primary interest is the resolution of peaks, then the untapered periodogram is superior. However, tapering significantly reduces the sidelobes and also the bias applied to other nearby peaks by the sidelobes of a strong peak. Because, the taper functions are broad and slowly varying and their fourier transform FT(g) are narrow. The effect of tapering the data is to convolve the fourier transform of the data with the narrow fourier transform of the taper function which amounts to smoothing the spectral values.

   $$FT(fg) = FT(f) * FT(g)$$

```r
# Sine taper
t <- seq(0,1, by=0.01)
T <- 1
g <- sin(pi * t * T)
plot(t, g, t='l', col=1, ylab='g(t)')

# Hanning (offset cosine) taper
g2 <- 1/2 * (1-cos(2*pi*t/T))
lines(t, g2, t='l', col=2)

# Hamming
g3 <- 0.54 - 0.46 * cos(2*pi*t/T)
lines(t, g3, t='l', col=3)

# Parzen or Bartlett (triangle) window
g4 <- ifelse(t>0.5, 1 - (t-T/2)/(T/2), 2*t)
```

```
lines(t, g4, t='l', col=4)

# Welch (parabolic) window
g5 <- 1 - (t-T/2)^2/(T/2)^2
lines(t, g5, t='l', col=5)

# Daniell window
g6 <- rep(0.5, length(t))
g6 <- ifelse(t <= 0.2, 0, g6)
g6 <- ifelse(t >= 0.8, 0, g6)
lines(t, g6, t='l', col=6)

legnd = c('Sine', 'Hanning', 'Hamming', 'Bartlett', 'Welch', 'Daniell(20%)')
legend('topleft', legend=legnd ,col=1:6, lty=1, cex=0.75)
```
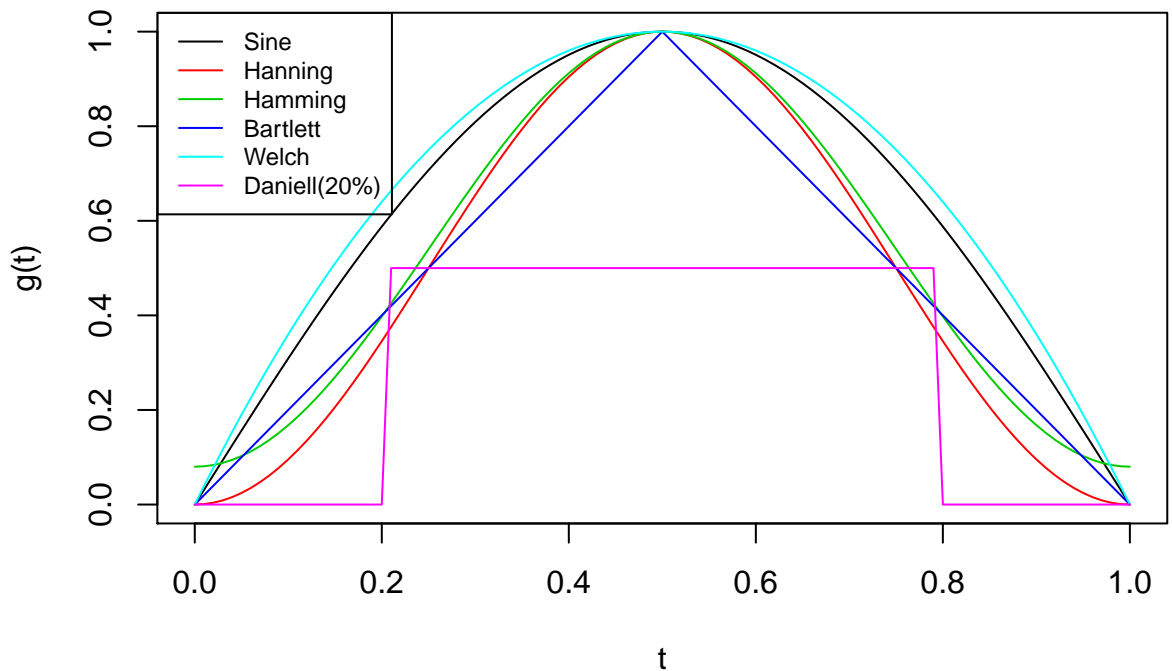

4.1 Multitaper Analysis

We apply taper or data window to reduce the side lobes of the spectral lines. Basically we want to minimize the leakage of power from the strong peaks to other frequencies. In multitaper method, several different tapers are applied to the data and the resulting powers then averaged. Each data taper is multiplied element-wise by the signal to provide a windowed trial from which one estimates the power at each component frequency. As each taper is pairwise orthogonal to all other tapers, the windowed signals provide statistically independent estimates of the underlying spectrum. The final spectrum is obtained by averaging over all the tapered spectra. D. Thomson chose the Slepian or discrete prolate spheroidal sequences as tapers since these vectors are mutually orthogonal and possess desirable spectral concentration properties. Multitaper method can suppress sidelobes but have higher resolution. If we use few tapers, the resolution won't be degraded, but then sidelobe reduction won't happen much. So, there is a trade-off which is often misunderstood.

5. Blackman-Tuckey Method

Blackman and Tuckey prescribed some techniques to analyze a continuous spectrum that was biased by the presence of sidelobes of strong peaks in the ordinary periodogram. Blackman-Tuckey(BT) method was developed before 1958, prior to the FFT(Fast Fourier Transform) method. A discrete fourier transform of N points would require the calculation of N^2 sines and cosines. With the slower computer in the pre-FFT days, the calculation of fourier transform was thus expensive. BT method has reduced the time by reducing the size of the dataset by a factor of the lag in the autocorrelation calculation. BT method is based on a fundamental theorem of Fourier transform that the Fourier transform of a correlation is equal to the product of the Fourier transforms. The correlation of two functions $g(t)$ and $h(t)$ is given by the first equation below.

$$C(\tau) = g \otimes h = \int_{-\infty}^{\infty} g(t)h(t+\tau)d\tau$$
$$FT(g \otimes h) = FT(g)FT(h)$$

When g = h, it is called Wiener-Khintchine theorem. Here, P is the spectral power.

$$FT(g \otimes g) = |FT(g)|^2 = P$$

The algorithm in BT method calculates partial autocorrelation function, defined by

$$A_{BT}(\tau) = \int_0^{N/l} f(t+\tau)f(t)dt$$

Here, N is the length of the data set but we integrate only upto N/l. l is associated with the lag. When $l = 3$ (recommended by Blackman and Tuckey) is used, we say that "a lag of 1/3" is used. Now the fourier transform of partial autocorrelation function A_{BT} gives us the spectral power. Moreover, the symmetric property of the partial autocorrelation function ($A(-\tau) = A(\tau)$) saves half of the computation time.

$$FT(A_{BT}) = \int_{-\infty}^{\infty} e^{2\pi i f t} A_{BT}(\tau)d\tau = P_{BT}(f)$$
$$P_{BT}(f) = 2\int_0^{\infty} cos(2\pi f)A_{BT}(\tau)$$

If $N = 1$, then it is basically the full autocorrelation function $A(\tau)$ and gives the same answer as the ordinary periodogram.

$$P(f) = 2\int_0^{\infty} cos(2\pi f)A(\tau) = FT(A)$$

Because we are using partial correlation function instead of the full correlation, the spectral power function gets smoother. Therefore, we lose resolution in the BT method. However, it averages the sidelobes into the main peak, and thereby gives a better estimate of the true power. The smoothing in BT method is different from the smoothing when we use a taper. With a taper, the fourier transform is smoothed, where as with Blackman-Tukey, it is the spectral power which gets smoothed. A spectral amplitude that is rapidly varying will be averaged to zero with a taper. But in BT method, a rapidly varying amplitude does not necessarily average to zero, since the process of squaring can make the function positive over the region of smoothing. The tapering does not average the sidelobes into the main peak. Because, shift in the time scale behaves like phase modulation. The sidelobes, when tapering is applied, will not have the same phase, and if averaged in amplitude, they can reduce the strength of the peaks. A major challenge in the BT method is that we will have to estimate the proper lag to use before doing all the calculations. Blackman and Tukey recommended starting with the value 1/3 for the lag.

6. Lomb-Scargle Periodogram

The classic periodogram requires evenly spaced data, but we frequently encounter with unevenly spaced data in paleoclimatic research. Lomb and Scargle showed that if the cosine and sine coefficients are normalized separately, then the classic periodogram can be used with unevenly spaced data. If we have a data set (t_k, y_k), we first calculate the mean and variance:

$$\bar{y} = \frac{1}{N} \sum_{k=1}^{N} y_k$$
$$\sigma^2 = \frac{1}{N-1} \sum_{k=1}^{N} [y_k - \bar{y}]^2$$

For every frequency f, a time constant τ is defined by

$$\tau = \frac{\sum_{k=1}^{N} sin(4\pi f t_k)}{\sum_{k=1}^{N} cos(4\pi f t_k)}$$

Then the Lomb-Scargle periodogram of the spectral power $P(f)$ at frequency f is given by

$$P(f) = \frac{1}{2\sigma^2} \frac{\sum_{k=1}^{N} (y_k - \bar{y})[cos(2\pi f(t_k - \tau))]^2}{\sum_{k=1}^{N} cos^2(2\pi f(t_k - \tau))} + \frac{\sum_{k=1}^{N} (y_k - \bar{y})[sin(2\pi f(t_k - \tau))]^2}{\sum_{k=1}^{N} cos^2(2\pi f(t_k - \tau))}$$

With evenly spaced data, two signals of different frequencies can have identical values which is known as Aliasing. That is why the classic periodogram is usually shown with the frequency range from 0 to 0.5, as the rest is a mirrored version. But with Lomb-Scargle periodogram, the aliasing effect can be significantly reduced.

7. Maximum Likelihood Analysis

In maximum likelihood method, we adjust the parameter of the model and ultimately find the parameters with which our model have the maximum probability/likelihood of generating the data. To estimate the spectral power, we first select a false alarm probability and calculate the normalized periodogram. We identify the maximum peak and test it against the false alarm probability. If the maximum peak meets the false alarm test, we determine the amplitude and phase of the sinusoid representing the peak. Then we subtract the sinusoidal curve from the data which also removes the annoying sidelobes associated with that peak. After peak removal, the variance in the total record is also reduced. Now, with the new subtracted data, we continue finding the other stronger peaks following the same procedure. We stop when a peak does not meet the false alarm test. We need to carefully choose the false alarm probability, as if it is too low, we can miss some significant peaks; it is too low, we can mislavel noise as peaks. ## 8. Maximum Entropy Method It is assumed that the true power spectrum can be approximated by an equation which has a power series. This method finds the spectrum which is closest to white noise (has the maximum randomness or "entropy") while still having an autocorrelation function that agrees with the measured values - in the range for which there are measured values. It yields narrower spectral lines. This method is suitable for relatively smooth spectra. With noisy input functions, if very high order is chosen, there may occur spurious peaks. This method should be used in conjuction with other conservative methods, like periodograms, to choose the correct model order and to avoid getting false peaks. ## 9. Cross Spectrum and Coherency

10. Bispectra

11. Monte Carlo Simulation

Climate Proxies

1. Oxygen Isotopes

2. Deuterium - Temperature Proxy

3. Carbon-13

4. Vostok

5. Atmospheric $\delta^{18}O$ and Dole Effect

6. $\delta^{18}O$ / CO_2 Mystery

7. Sea Floor Records

8. Terrestrial Records

Paper Review