EART266 Geological Signal Processing
Lecture #7
Power Spectral Density: 2, Blackman-Tukey

We start with a simple idea: white noise–a sequence of uncorrelated, zero-mean, constant variance random variables–has a flat power spectral density such that all frequencies carry equal power. Correlated processes have colored spectra. The latter can be generated by combinations (filtering) of the former. Two end-member schemes for generating colored noise from white noise are the simple MA(q) and AR(n) processes we already examined. AR(n) processes have auto-covariance functions which decay to zero asymptotically: each time step is a weighted sum of all prior time steps and a new white noise term. Since the auto-covariance function is infinite in length, any estimate of it made from sample data must be truncated.

MA(q) processes are weighted sums of the previous n shocks and a new shock. Because the shocks (white noise input) are uncorrelated, the auto-covariance function goes to zero identically for all offsets (lags) greater than q. If our sample is at least this long, we do not have to truncate the auto-covariance function (note that the sample should be much longer such that estimates of the auto-covariance for lags near q are derived from many samples).

The Blackman-Tukey method of estimating the power spectral density assumes a MA(q) process <u>with known q.</u> Once you've made this assumption, the method is fairly straightforward. It does, however, rely on a theorem I haven't shown you yet, so here it is…

The auto-covariance function of a time series is the cross-correlation of the time series with itself. For a continuous time series (of infinite duration), we write this as:

$$c(t) = \int_{-\infty}^{\infty} h(t+\tau)h(\tau)d\tau$$

This integral becomes multiplication in the frequency domain:

$$C(f) = H(f)H^*(f) = |H(f)|^2$$

To within a constant multiplier, this is the power spectral density. Thus the Fourier transform of the auto-covariance function is an estimate of psd. Blackman-Tukey uses this, plus knowledge that the process is MA(q) (such that we can compute the entire auto-covariance function from samples of the time series), to estimate the psd. Because it assumes a particular process, the BT method is parametric.

Okay, make room for BT.

I start with a simple outline of the procedure:

1. Estimate the auto-covariance function $c(t)$ using sample data.

2. Choose the maximum offset, q, and taper the auto-covariance estimate. Commonly used tapers include:

$$\text{Tukey}: \; w_{\text{n}} = \frac{1}{2}\left[1 + \cos\frac{\pi n}{q}\right]$$

and

$$\text{Parzen}: \; w_{\text{n}} = \begin{cases} 1 - 6\left(\dfrac{n}{q}\right)^2 + 6\left(\dfrac{n}{q}\right)^3 & n \le q/2 \\ 2(1 - n/q)^3 & q/2 < n \le q \end{cases}$$

where both are symmetric about the origin.

3. Pad the windowed auto-covariance function with zeroes out to large times. This will increase the number of Fourier frequencies in the DFT but will not alter the power spectrum estimate.

4. DFT (or FFT) the windowed, padded auto-covariance function to obtain the power spectral density estimate.

(5. Do it again for a few more q's to see if the result is stable or not.)

The advantages of BT are:

1. It doesn't arbitrarily extrapolate the auto-covariance function. The periodogram (secretly) does this by assuming it is zero for all lags $> 2P - 1$. Of course you do have to choose q…

2. It produces smoother spectra than the periodogram method. Does smoother equal better? It does if the true power spectrum is smooth…

3. It is consistent (more data equals less variance in the estimate) and is computationally efficient. The latter concern is essentially moot these days given all 'dem new fangled computers.

4. It's right if the process is MA(q). Finally an advantage I'm not mocking.

Before showing how to move directly to the power spectrum of a known MA(q) process, I want to mention that there are two ways in common use to compute the auto-covariance function:

Method 1:

$$c(k)=\frac{1}{P-|k|}\sum_{j=0}^{P-|k|-1}x_j x_{j+|k|}$$

Method 2:

$$c(k)=\frac{1}{P}\sum_{j=0}^{P-|k|-1}x_j x_{j+|k|}$$

The difference is all in the leading term. In Method 1 we divide the sum by the number of terms that went into it. In Method 2 we always divide by the total number of samples in the time series (*P*). Method 2 is biased, but has a lower standard error than Method 1. Most algorithms use Method 2. Provided $q << P$ it hardly matters. If q is at all close to P, you probably shouldn't be trying to estimate the power spectrum. (Note that I have assumed the time series is zero mean in both methods. If it isn't, simply subtract the mean before computing the auto-covariance.)

The Power Spectral Density of a Known MA(q) Process

This section contains some material that is relevant to all Fourier analysis. I'm developing it here because this is the first time I've needed it.

The DFT of a sampled time series is linear in the sense that:

$$DFT(z_n = x_n + y_n) = DFT(x_n) + DFT(y_n)$$

The DFT of a time series consisting of a single isolated spike at time step *j*:

$$h_n = a\delta_{nj}$$

is easily computed from the definition of the DFT:

$$H_k = \sum_{n=0}^{P-1} h_n e^{-2\pi ink/P} = \sum_{n=0}^{P-1} a\delta_{nj} e^{-2\pi ink/P} = ae^{-2\pi ijk/P}$$

Combining these two results, we can compute the DFT of a sampled time series by considering it as a sum of many spikes (one at each time step) and summing the DFTs of each spike. I'd write this out mathematically, but I already have: the definition of the DFT is the sum of transforms of spikes at each time step. This equivalence (sum of DFTs of spikes and the DFT of the sum of spikes) is especially useful for analytically transforms of time series (or auto-covariance functions) that have few non-zero terms. The auto-covariance of a MA(q) is such a function.

Let our MA(q) process be given by:

$$x_n = \varepsilon_n + \sum_{i=1}^{q} \alpha_i \varepsilon_{n-i}$$

The ε are uncorrelated, Gaussian random variables with zero mean and variance σ^2. We wish to compute the auto-covariance function of this process, i.e.:

$$E[x_n x_{n+k}] = C(k)$$

This isn't too hard (just messy):

$$C(k) = \begin{cases} \sigma^2\left(1 + \sum_{i=1}^{q} \alpha_i^2\right) & k = 0 \\ \sigma^2\left(\alpha_k + \sum_{i=1}^{q-k} \alpha_i \alpha_{i+k}\right) & 0 < k \le q \\ 0 & k > q \end{cases}$$

For a MA(1) process, this simplifies considerable:

$$C(k) = \begin{cases} \sigma^2\left(1 + \alpha_1^2\right) & k = 0 \\ \sigma^2 \alpha_1 & k = 1 \\ 0 & else \end{cases}$$

This auto-covariance function is non-zero for only three times ($k = -1$, 0 and 1), hence its DFT is simply the sum of three terms:

$$C(f_k) = \sigma^2 \alpha_1 e^{2\pi i k/P} + \sigma^2\left(1 + \alpha_1^2\right) + \sigma^2 \alpha_1 e^{-2\pi i k/P}$$
$$C(f_k) = \sigma^2\left(1 + \alpha_1^2\right) + 2\sigma^2 \alpha_1 \cos(2\pi k/P)$$

As a simple check on this result, let $\alpha_1 = 0$. This results in white noise. Is the transform white? Also check that this result can never be negative (the power spectrum is a non-negative quantity).

The MA(1) spectrum is red if $\alpha_1 > 0$ because the cosine term goes from 1 at $k = 0$ (zero frequency) to -1 at $k = P / 2$ (Nyquist frequency) causing C to decrease with increasing k. The opposite is true if $\alpha_1 < 0$, resulting in a blue spectrum.

The process for higher-order MA processes is a straight forward extension of the MA(1) process. The result, included because I'm sitting here typing while waiting for someone who is very late, is:

$$C(f_k) = \sigma^2\left(1+\sum_{i=1}^{q}\alpha_i^2\right)+2\sigma^2\sum_{j=1}^{q}\left(\alpha_j+\sum_{i=1}^{q-j}\alpha_i\alpha_{i+j}\right)\cos(2\pi jk/P)$$

This is a specific example of an important result: the DFT of an auto-covariance function (which is an estimate of the power spectrum) is the sum of a constant plus cosine functions of frequency. The result must be smooth provided *q* isn't very large.

The following examples apply the Blackman Tukey power spectrum estimate to a variety of MA and AR processes. It does quite well with the MA processes (it better–it assumes them). It kinda sucks with AR processes. This makes sense: AR processes have infinitely long auto-covariance functions: truncating them after *q* throws away a lot of the covariance structure that the spectrum estimate is based on.