12.714 Computational Data Analysis

Alan Chave (alan@whoi.edu)
Thomas Herring (tah@mit.edu),
http://geoweb.mit.edu/~tah/12.714

Today's class

- Non-parametric Spectral Estimation
  - Bias reduction: Pre-whitening
  - Statistical Properties of direct spectral estimates
  - Smoothing of direct spectral estimates
  - First moment properties of lag window estimators
  - Second moment properties of lag window estimators

Bias reduction: Pre-whitening

- We already have seen bias reduction through the use of tapers
- The idea of a pre-whitening filter is to pre-filter the time series to reduce the dynamic range. This is done with a *pre-whitening filter* , g_u

$$Y_t = \sum_{u=-L}^{K} g_u X_{t+K-u} \quad 1 \le t \le M \equiv N - (K+L)$$

$$S_Y(f) = \left| \sum_{u=-L}^{K} g_u e^{-i2\pi f u \Delta t} \right|^2 S_X(f) \quad S_X(f) \approx \frac{S_Y^{(d)}(f)}{\left| \sum_{u=-L}^{K} g_u e^{-i2\pi f u \Delta t} \right|^2}$$

$$S_Y^{(d)}(f) = \Delta t \left| \sum_{t=1}^{M} h_t Y_t e^{-i2\pi f t \Delta t} \right|^2$$

Bias reduction: Pre-whitening

- Ideally the spectral density function of Y_t is flat and hence the idea of pre-whitening.
- There are problems and tradeoffs:
  - Since the filter has a finite length, the pre-whitening time series has less data (and lower spectral resolution).
  - "Chicken and Egg" problem: How do you know filter to use before knowing the sdf? Experience and physics can help
  - Estimation of the filter from the data themselves. Discussed in Chapter 9 of PW where an AR(n) process is fit to the data to obtain the pre-whitening filter (still involves assumptions of order to use).

Statistics of Direct Spectral Estimation

- Consider the spectral estimates of white, Gaussian noise, G_t, with variance σ^2, using taper h_t.

$$J(f) = A(f) + iB(f) = (\Delta t)^{1/2} \sum_{t=1}^{N} h_t G_t e^{-i2\pi ft}$$

$$\text{var}\{A(f) \mid B(f\}\} = \sigma^2 \Delta t \sum_{t=1}^{N} h_t \cos^2(2\pi ft\Delta t) \mid \sin^2(2\pi ft\Delta t);$$

$$\text{cov}\{A(f), A(f')\} = \sigma^2 \Delta t \sum_{t=1}^{N} h_t \cos(2\pi ft\Delta t)\cos(2\pi f' t\Delta t);$$

$$\text{cov}\{A(f), B(f')\} = \sigma^2 \Delta t \sum_{t=1}^{N} h_t \cos(2\pi ft\Delta t)\sin(2\pi f' t\Delta t);$$

Statistics: Gaussian White noise

- If we take a rectangular taper and consider the Fourier frequencies (k/(NΔt)) then

$$\text{var}\{A(f_k)\} = \begin{cases} \sigma^2 \Delta t / 2 & \text{for } \mathrm{f_k} \neq 0 \text{ or } \mathrm{f_{(N)}} \\ \sigma^2 \Delta t & \text{for } \mathrm{f_k} = 0 \text{ or } \mathrm{f_{(N)}} \end{cases}$$

$$\text{var}\{B(f_k)\} = \begin{cases} \sigma^2 \Delta t / 2 & \text{for } \mathrm{f_k} \neq 0 \text{ or } \mathrm{f_{(N)}} \\ 0 & \text{for } \mathrm{f_k} = 0 \text{ or } \mathrm{f_{(N)}} \end{cases}$$

All covariances at the Fourier frequencies are zero

- Since $A^2(f)+B^2(f)=S(f)$, it follows (d over = means distributed)

$$\frac{2}{\sigma^2 \Delta t} \hat{S}_G^{(p)}(f_k) \overset{d}{=} \chi_2^2 \quad \hat{S}_G^{(p)}(f_k) \overset{d}{=} \frac{\sigma^2 \Delta t}{2} \chi_2^2 \text{ for } f_k \neq 0, f_{(N)}$$

Statistics: Gaussian White Noise

- For the case of f=0 or the Nyquist rate:

$$\hat{S}_G^{(p)}(f_k) \stackrel{d}{=} \sigma^2 \Delta t \chi_2^2 \text{ for } f_k = 0, f_{(N)}$$

For Gaussian white noise: $\sigma^2 \Delta t = S_G(f)$ and $\therefore$

$$\hat{S}_G^{(p)}(f_k) \stackrel{d}{=} \begin{cases} S_G(f)\chi_2^2/2 & \text{for } f_k \neq 0, f_{(N)} \\ S_G(f)\chi_2^2 & \text{for } f_k = 0, f_{(N)} \end{cases}$$

- So Gaussian white noise, the sdf estimates are Chi-squared distributed with 2 degrees of freedom.
- Remembering $E\{\chi_\nu^2\} = \nu$ and $\mathrm{var}\{\chi_\nu^2\} = 2\nu$ we can write expressions for the variance of our estimates

Expectation and Variance

- Using the chi-squared expectation and variance we have

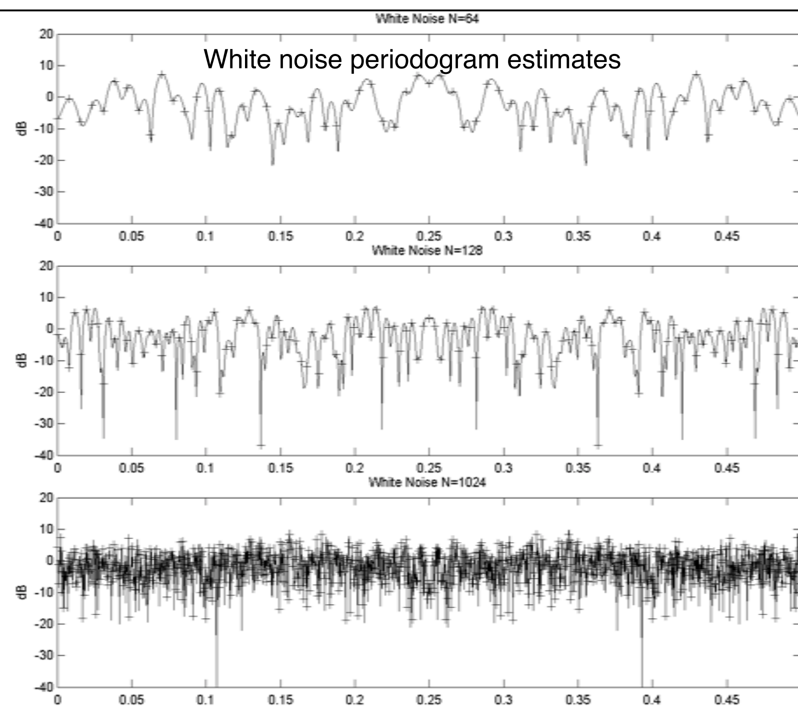
$$E\{\hat{S}_G^{(p)}(f_k)\} = \sigma^2 \Delta t = S_G(f) \text{ for all } \mathrm{f_k}$$

$$\mathrm{var}\{\hat{S}_G^{(p)}(f_k)\} = \begin{cases} \sigma^4 \Delta t^2 = S_G^2(f) & \text{for } f_k \neq 0, f_{(N)} \\ 2\sigma^4 \Delta t^2 = 2S_G^2(f) & \text{for } f_k = 0, f_{(N)} \end{cases}$$

- Samples at the Fourier frequencies are independent.
- The same relationships hold for stationary processes (with some restrictions on the finiteness of higher order moments) as the number of samples used to compute the spectra tends to infinity.

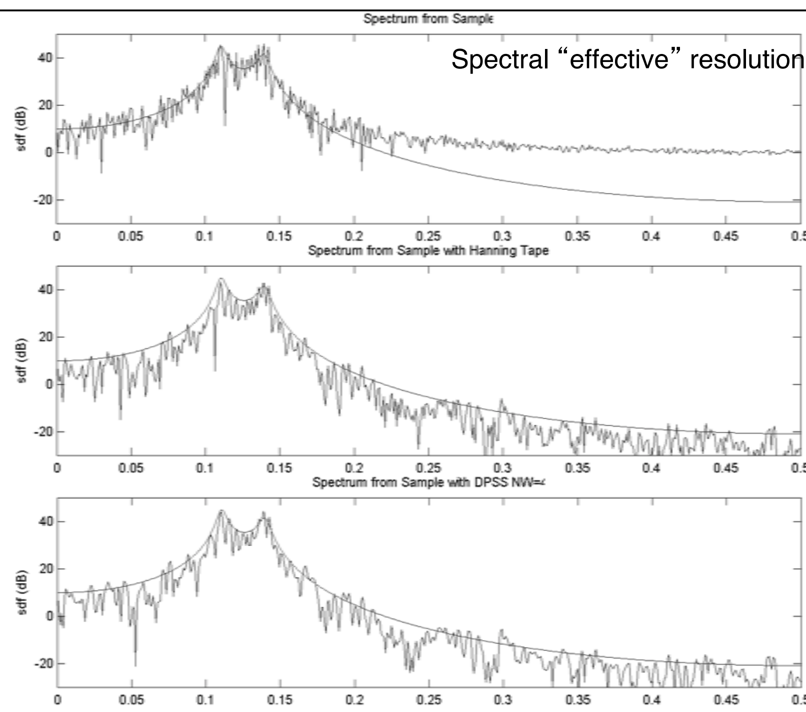
Statistics Direct Spectral Estimates

- The variance properties hold for direct spectral estimates provided the form of $\{h_t\}$ is reasonable again as N tends to infinity.
- However, the grid on which the estimates are uncorrelated is often modified: As we saw the central peak is widened to suppress the side lobes and so the un-correlated estimates occur at the nulls in the central peak.
- Since the variances of the estimates, in all cases, do not depend on N, these estimators are not *consistent estimators* of S(f).
- Because the variance is proportional to S(f), on dB plot, the noise should appear the same at all frequencies (not so on a linear plot). Smoothness of the dB plot implies leakage leading to smooth estimated.



Example

- To show effect on resolution; the next set of figures show the spectral density functions computed for the AR(4)
  - Standard Periodogram
  - Hanning Taper
  - DPSS with NW=4
- For the latter two, note the change in resolution
- For all cases: Specific look will depend on random sequence (try the lecture case with 102 as seed the randn).



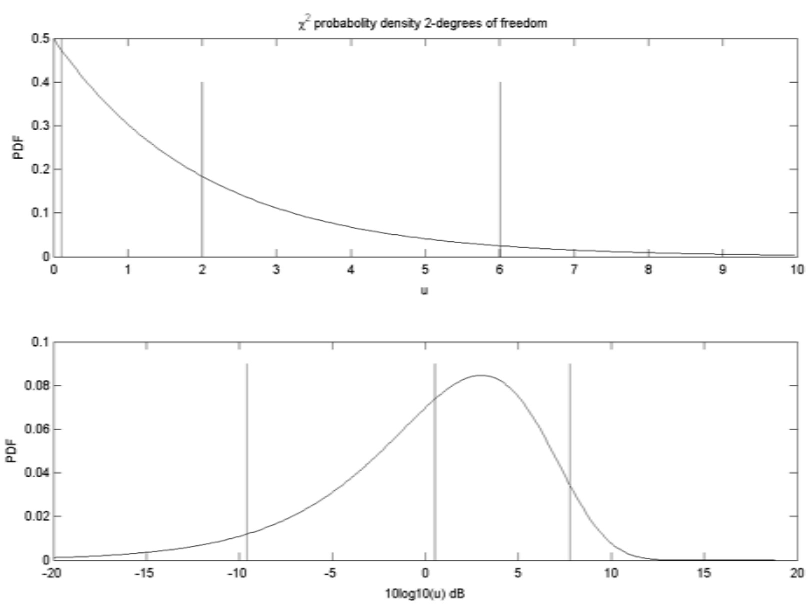
Statistical properties

- Since the sdf estimates are chi-squared distributed with 2-degrees of freedom, the asymmetry in this distribution leads to interesting visual effect.
- The PDF for chi-squared in linear and log space is given by

$$\boxed{\begin{array}{l} f_u(u) = \begin{cases} e^{-u/2}/2 & \text{for } u \geq 0 \\ 0 & \text{for } u < 0 \end{cases} \\ f_v(v) = \dfrac{\log(10)}{20} 10^{v/10} e^{-(10^{v/10})/2} \text{ for } v = 10\log(\mathrm{u}) \end{array}}$$

- On a linear plot, the "upshots" appear more frequent while on a dB (log) scale the "down shots" appear more prominent.

Chi-squared 2-degrees of freedom



Smoothing direct estimates

- The periodogram and direct spectral estimates have problems because of large variability and possibly weaken statistical tests because of high noise levels (bias can also be a problem).
- The traditional approach is to smooth the estimates of S(f). If N is large enough then we can generate an average:

$$\boxed{\begin{aligned}&\bar{S}(f_k) \equiv \frac{1}{2M+1}\sum_{j=-M}^{M}\hat{S}^{(p)}(f_{k-j})\\&\mathrm{var}\{\bar{S}(f_k)\} \approx S^2(f_k)/(2M+1)\end{aligned}}$$

- As N and M increase (keeping fk the same), the variance decreases so this is a *consistent estimator*,

Smoothing direct estimates

- Rather than just taking an average of the spectral estimates we can use smoothing sequence

$$\boxed{\bar{S}^{(ds)}(f_k^{'}) \equiv \sum_{j=-M}^{M} g_j \hat{S}^{(d)}(f_{k-j}^{'}) \;\; \text{with } f_k^{'} \equiv \frac{k}{N'\Delta t}}$$

- Where N' is chosen to control the frequency spacing. Normally N' is greater than or equal the sample size
- This estimate is called the *discretely smoothed direct spectral estimator*.
- The coefficients $\{g_j\}$ are a LTI digital filter.

Lag Window estimator

- The previous case was discrete smoothing but we can define the spectrum continuously in frequency and use a continuous convolution

$$\hat{S}^{(lw)}(f) = \int_{-f_{(N)}}^{f_{(N)}} V_m(f-\phi)\hat{S}^{(d)}(\phi)d\phi$$

$$\hat{S}^{(lw)}(f) = \int_{-f_{(N)}}^{f_{(N)}} V_m(f-\phi)\left(\Delta t \sum_{\tau=-(N-1)}^{N-1} \hat{s}_\tau^{(d)} e^{-i2\pi\phi\tau\Delta t}\right)d\phi$$

$$\hat{S}^{(lw)}(f) = \Delta t \sum_{\tau=-(N-1)}^{N-1} v_{\tau,m}\hat{s}_\tau^{(d)} e^{-i2\pi\phi\tau\Delta t} \quad v_{\tau,m} = \int_{-f_{(N)}}^{f_{(N)}} V_m(\phi)e^{i2\pi\phi\tau\Delta t}d\phi$$

Lag Window Estimator

- To be precise, the window is written as

$$w_{\tau,m} = \begin{cases} v_{\tau,m}, & |\tau| < N \\ 0, & |\tau| \geq N \end{cases} \quad W_m(f) \equiv \Delta t \sum_{t=-(N-1)}^{N-1} w_{\tau,m}e^{-i2\pi f\tau}$$

- $W_m(.)$ is called a smoothing window (some authors use spectral window) and $\{w_{\tau,m}\}$ is called a lag window (other names include quadratic window, quadratic taper)
- $S^{(lw)}(f)$ is called a lag window spectral estimator.
- The directly smoothed spectral estimator can be expressed as a lag window estimator.

Lag Window Conditions

- For a lag window to be have smaller variance than the direct estimator we require
  - $W_m(.)$ should be an even $2f_{(N)}$ periodic function
  - The integral of $W_m(.)$ over the Nyquist range should be 1 (or equivalently $w_{0,m}$=1)
  - For any $\varepsilon>0$ and for $|f|> \varepsilon$, $W_m(f)\to 0$ and $m \to\infty$
  - $W_m(f)\geq 0$ for all m and f (desirable to ensure the lag window sdf is always positive but not sufficient or necessary).

Bandwidth of LW estimator

- If $W_m(f)$ is always positive we can define

$$\beta_W \equiv \left(12\int_{-f_{(N)}}^{f_{(N)}} f^2 W_m(f)df\right)^{1/2} = \left(\frac{1}{(\Delta t)^2}\left(1+\frac{12}{\pi^2}\sum_{t=1}^{N-1}\frac{(-1)^\tau}{\tau^2}w_{\tau,m}\right)\right)^{1/2}$$

- This form can have computational problems because of alternating signs.
- Another definition that matches the earlier definition of the autocorrelation width is

$$B_w \equiv \frac{1}{\int_{-f_{(N)}}^{f_{(N)}} W_m^2(f)df} = \frac{1}{\Delta t \sum_{\tau=-(N-1)}^{N-1} w_{\tau,m}^2}$$

First moment properties of LW estimators

- Since the lag window estimators are effectively a convolution with a convolution we have

$$E\{\hat{S}^{(lw)}(f)\} = \int_{-f_{(N)}}^{f_{(N)}} U_m(f-\phi)S(\phi)d\phi$$

$$U_m(f) \equiv \int_{-f_{(N)}}^{f_{(N)}} W_m(f-f')\mathrm{H}(f')df'$$

- $U_m(.)$ is called the *spectral window* (by PW)
- The bias between the $S^{(lw)}f$ and S(f) will depend on two things: The curvature of S(f) (depends on the second derivative) and on the bandwidth of the smoothing window (too wide a bandwidth will smear peaks).

Second Moment Properties

- With some assumptions we have

$$\mathrm{var}\{\hat{S}^{(lw)}(f_k^{'})\} \approx \frac{S^2(f_k^{'})}{N'\Delta t} \int_{-f_{(N)}}^{f_{(N)}} W_m^2(\phi)d\phi$$

- Assumptions:
  - Pair-wise uncorrelated estimates at f_k
  - Large sample variance for $S^{(d)}$
  - Smoothness assumptions
  - Wm is approximately zero for frequencies greater than specified value
  - A summation can be replaced with a Riemann integral

Second Moment Properties

- The spacing between uncorrelated estimates in the direct spectrum estimate can be quantified with N' =N/C_h. Details on computing Ch are given p 250-253 PW.
- Results of different tapers

Data Taper	C_h
Rectangle	1.00
20% cosine	1.12
50% cosine	1.35
100% cosine	1.94
NW 1 dpss	1.34
NW 2 dpss	1.96
NW 3 dpss	2.80
NW 4 dpss	3.94

Summary of today' class

- Non-parametric Spectral Estimation
  - Bias reduction: Pre-whitening
  - Statistical Properties of direct spectral estimates: Allows to assess the variance of the spectral estimates
  - Smoothing of direct spectral estimates: Two methods direct smoothing and lag-window estimates
  - First moment properties of lag window estimators: Bias in estimates (especially leakage)
  - Second moment properties of lag window estimators: Variance of estimates and the effects of bandwidth and the effective number of samples available.