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
  - Estimation of Mean
  - Estimation of autocovariance sequence
  - Naïve spectral estimation: Periodograms
  - Bias reduction - Tapering
- Basic Idea: Determine s_τ and the acvs determine spectrum

$$S(f) = \Delta t \sum_{\tau=-\infty}^{\infty} s_\tau e^{-i2\pi f\tau\Delta t} \text{ for } |f| \le f_{(N)} \equiv 1/(2\Delta t)$$
$$s_\tau = \mathrm{cov}(X_t, X_{t+\tau}) = E\{(X_t - \mu)(X_{t+\tau} - \mu)\}$$

Estimation of mean

- A natural estimator for the mean and its definition as a *consistent estimator* is

$$\overline{X} \equiv \frac{1}{N}\sum_{t=1}^{N} X_t \qquad \lim_{N\to\infty} P[|\overline{X}-\mu| > \varepsilon] = 0, \text{ for } \varepsilon > 0$$

- Chebyshev's inequality shows

$$P[|\overline{X}-\mu| > \varepsilon] \le \frac{E\{|\overline{X}-\mu|^2\}}{\varepsilon^2} = \frac{\mathrm{var}\{\overline{X}\}}{\varepsilon^2}$$

- Variance of the mean:

$$\mathrm{var}\{\overline{X}\} = \frac{1}{N^2}\sum_{t=1}^{N}\sum_{u=1}^{N} E\{(X_t-\mu)(X_u-\mu)\} = \frac{1}{N^2}\sum_{t=1}^{N}\sum_{u=1}^{N} s_{u-t} = \frac{1}{N}\sum_{\tau=-(N-1)}^{N-1}\left(1-\frac{|\tau|}{N}\right)s_\tau$$

Variance of Mean estimate

- We now may consider when N goes to infinity

$$\lim_{N\to\infty} N\,\mathrm{var}\{\overline{X}\} = \lim_{N\to\infty}\sum_{\tau=-(N-1)}^{N-1}\left(1-\frac{|\tau|}{N}\right)s_\tau = \sum_{\tau=-\infty}^{\infty} s_\tau$$

$$\text{and since } S(f) = \Delta\text{t} \sum_{\tau=-\infty}^{\infty} s_\tau e^{-i2\pi f t} \Rightarrow S(0) = \Delta t \sum_{\tau=-\infty}^{\infty} s_\tau$$

$$\mathrm{var}\{\overline{X}\} \approx S(0)/(N\Delta t)$$

- The sample mean is not always a consistent estimator for a stationary process. A process where the sample mean converges to the expectation is an *ergodic process*

Asymptotically efficient estimators

- If we know the covariance matrix, $\boxed{\Gamma_N}$, for the samples from our process then

$$\boxed{\hat{\mu} \equiv \frac{O_N^T \Gamma_N^{-1} \mathbf{X}}{O_N^T \Gamma_N^{-1} O_N} \text{ where } \mathrm{O_N} = [1,1,1,\cdots 1]^T}$$

- Then mean is *asymptotically efficient* if

$$\boxed{e(\overline{X},\hat{\mu}) \equiv \lim_{N \to \infty} \frac{\mathrm{var}\{\hat{\mu}\}}{\mathrm{var}\{\overline{X}\}} = 1}$$

- Example: Fractional difference process has

$$\boxed{\begin{array}{l} S(f) = C|\sin(\pi f \Delta t)|^{\alpha} \text{ where C} > 0,\ \alpha > \text{-}1; \text{ for small } f,\ S(f) \propto |f|^{\alpha} \\ \text{for } \text{-}1 < \alpha < 0, \text{ efficient but } \alpha > 0,\ e(\overline{X},\hat{\mu}) \to 0 \end{array}}$$

Estimation of Autocovariance Sequence

- A natural estimator for s_τ is

$$\boxed{\hat{s}_\tau^{(u)} = \frac{1}{N - |\tau|} \sum_{t=1}^{N-\tau} (X_t - \overline{X})(X_{t+|\tau|} - \overline{X}) \quad \tau = 0, \pm 1, \cdots, \pm(N-1)}$$

- If the process expectation is used above (rather than the sample mean, the expectation of this estimate is s_τ and this is an *unbiased* estimate for all $\tau \leq N\text{-}1$.
- When the sample mean is used, the estimate is biased.
- An another (usually preferred) estimator is

$$\boxed{\hat{s}_\tau^{(p)} = \frac{1}{N} \sum_{t=1}^{N-\tau} (X_t - \overline{X})(X_{t+|\tau|} - \overline{X}) \quad \tau = 0, \pm 1, \cdots, \pm(N-1)}$$

ACVS estimators

- The expectation of $s_\tau^{(p)}$ is (and a biased estimator for large τ

$$\boxed{E\{s_\tau^{(p)}\} = \frac{1}{N}\sum_{t=1}^{N-|\tau|} s_\tau = \left(1 - \frac{|\tau|}{N}\right)s_\tau}$$

- The two estimators are called the *unbiased* and *biased* estimates (even though the unbiased is also biased when the sample mean is used).
- In some cases the "unbiased" estimate has larger biases than the biased estimate (eg. AR(2) process)
- Which estimator should be used?

Properties of acvs estimates

- In many cases the mean square error (mse) is smaller for the biased estimator i.e.,

$$\boxed{\mathrm{mse}\{\hat{s}_\tau^{(p)}\} \equiv E\{(\hat{s}_\tau^{(p)} - s_\tau)^2\} < E\{(\hat{s}_\tau^{(u)} - s_\tau)^2\} \equiv \mathrm{mse}\{\hat{s}_\tau^{(u)}\}}$$

- Because the mse = variance + bias^2, the applies when the variability in $s^{(u)}$ is more harmful than the biase in $s^{(p)}$. Specifically, for large τ, $s^{(u)}$ is divided by N-τ where as $s^{(p)}$ is divided by N and therefore will be smaller. This is useful when st approach zero for large τ.
- Since correlation is avcs/variance, the $s^{(u)}$ estimator can generate correlations > 1, whereas $s^{(p)}$ will always be positive definite.
- Figures on page 192 of PW show these effects

Periodogram: Naïve Spectrum estimation

- For a discrete, real valued stationary process with zero mean we have $$S(f) = \Delta t \sum_{\tau=-\infty}^{\infty} s_\tau e^{-i2\pi f\tau\Delta t} \text{ for } |f| \le f_{(N)} \equiv 1/(2\Delta t)$$
- If we s[(p)] in the above equation we have

$$\Delta t \sum_{t=-(N-1)}^{(N-1)} \hat{s}_\tau^{(p)} e^{-i2\pi f\tau\Delta t} = \frac{\Delta t}{N} \sum_{\tau=-(N-1)}^{(N-1)} \sum_{t=1}^{N-|\tau|} X_t X_{t+|\tau|} e^{-i2\pi f\tau\Delta t}$$
$$= \frac{\Delta t}{N} \left| \sum_{t=1}^{N} X_t e^{-i2\pi f\tau\Delta t} \right|^2 \equiv \hat{S}^{(p)}(f)$$

- $S^{(p)}(f)$ is known as the *periodogram* (though depends of f).

Periodogram

- The periodogram is defined in the Nyquist frequency range
- If is related to the Fourier transform with the factor Δt/N instead of $(\Delta t)^2$.
- For discrete frequencies and times we have

$$\hat{s}_\tau^{(p)} = \frac{1}{2N\Delta t} \sum_{k=-(N-1)}^{N-1} \hat{S}^{(p)}(\tilde{f}_k) e^{i\pi k\tau/N}, \tau = 0, \pm 1, \cdots, (N-1), N$$
$$\tilde{f}_k \equiv k/(2N\Delta t)$$

Properties of periodogram

- If $S^{(p)}(f)$ were an ideal estimator we would have:

$$[1]\ E\{\hat{S}^{(p)}(f)\} \approx S(f) \text{ for all } f$$
$$[2]\ \mathrm{var}\{\hat{S}^{(p)}(f)\} \to 0 \text{ as } N \to \infty$$
$$[3]\ \mathrm{cov}\{\hat{S}^{(p)}(f)\hat{S}^{(p)}(f')\} \approx 0 \text{ for } f \neq f'$$

- For some processes [1] is a good approximation but for others it can be very poor;
- [2] is blatantly false when S(f)>0 and
- [3] holds for certain discrete frequencies (namely the Fourier frequencies).

Expectation of $S^{(p)}(f)$

- The expectation of S(p)(f) is given by

$$E\{\hat{S}^{(p)}(f)\} = \Delta t \sum_{\tau=-(N-1)}^{N-1} \left(1 - \frac{|\tau|}{N}\right) s_\tau e^{-i2\pi f\tau\Delta t}$$

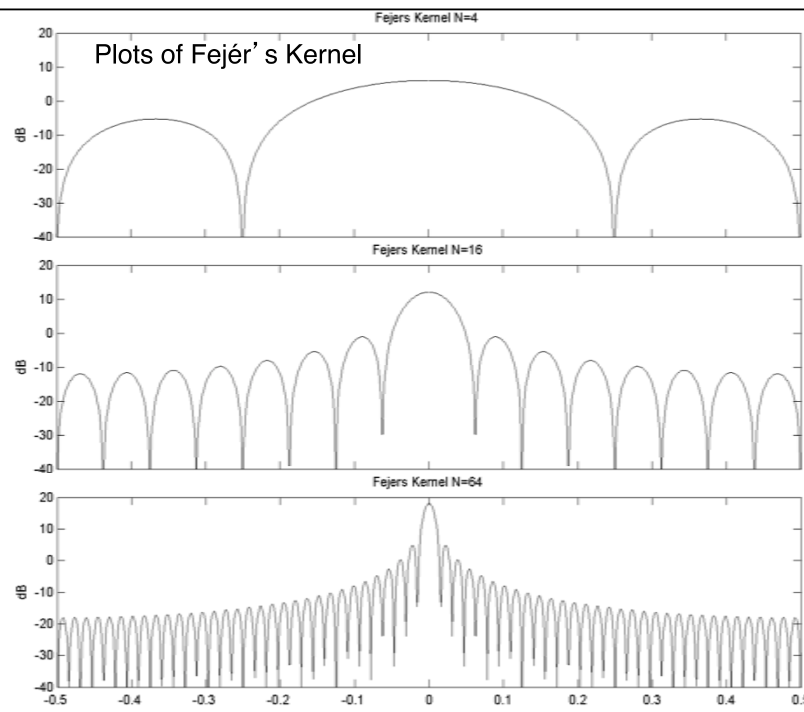
$$E\{\hat{S}^{(p)}(f)\} = N\Delta t \int_{-f_{(N)}}^{f_{(N)}} D_N^2([f - f']\Delta t) S(f')df' = \int_{-f_{(N)}}^{f_{(N)}} F(f - f')S(f')df'$$

$$F(f) \equiv N\Delta t D_N^2(f\Delta t) = \frac{\Delta t \sin^2(N\pi f\Delta t)}{N \sin^2(\pi f\Delta t)}$$

- D_N(.) is Dirichlet' s Kernel and F(.) is Fejér' s Kernel.

Properties of Fejér's Kernel

- For all $N \geq 1$ $F(f) \rightarrow N\Delta t$ as $f \rightarrow 0$.
- For $N>1$, $f \in [-f_{(N)}, f_{(N)}]$ and $f \neq 0$, $F(f) < F(0)$
- For $f \in [-f_{(N)}, f_{(N)}]$ and $f \neq 0$, $F(f) \rightarrow 0$ as $N \rightarrow \infty$
- For any $k \neq 0$ and $f_k = k/(N\Delta t)$, $F(f_k)=0$ (happens because $\sin(k\pi)=0$)
- The integral of F(f) over the Nyquist frequency range is 1.
- As N goes to infinity, F(f) acts like a Dirac delta function and $S^{(p)}(f)$ approaches S(f).
- Following figures show F(f) for N=4, 16 and 64

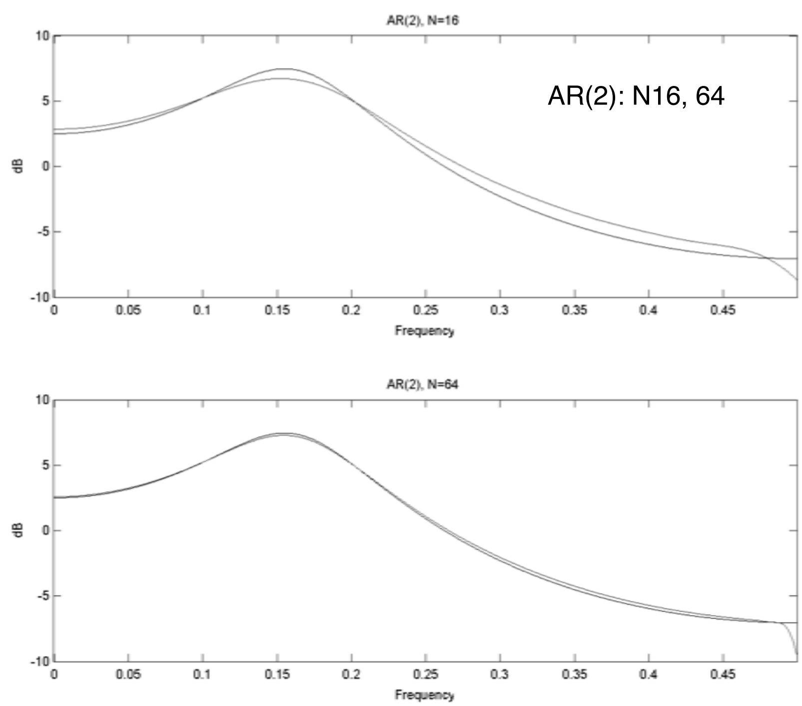


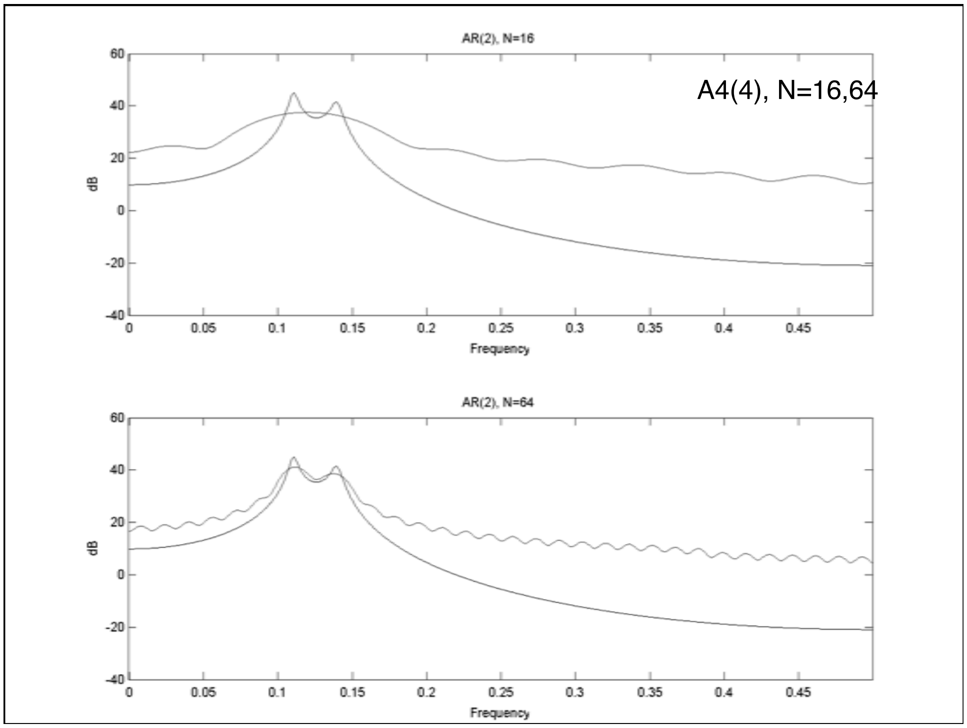
Bias in periodogram

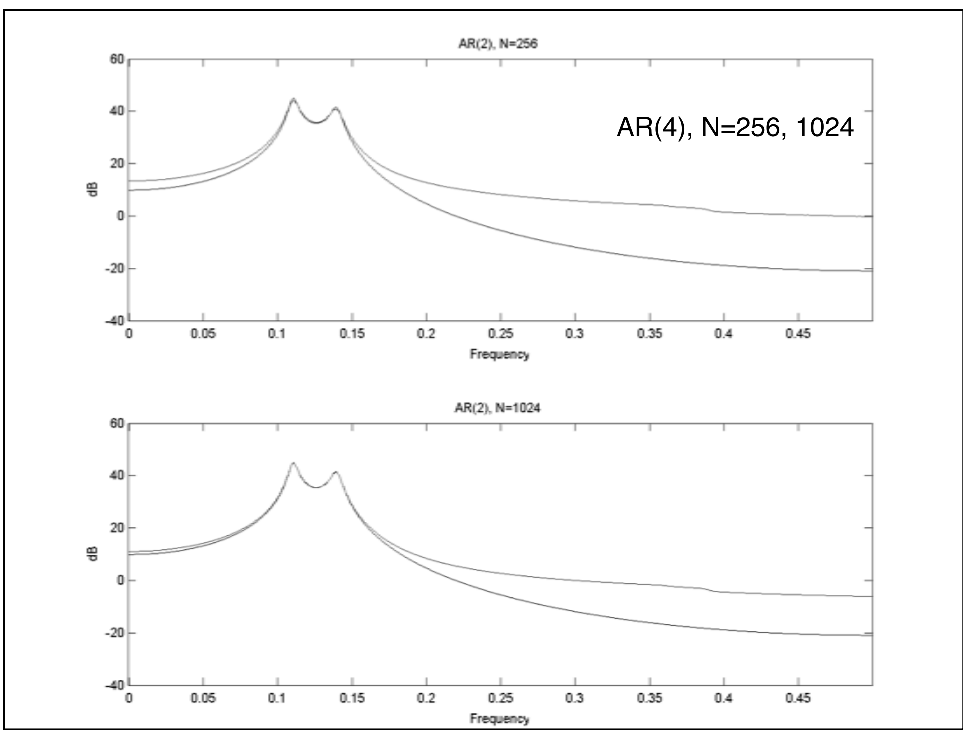
- The amount bias in the periodogram depends of the nature of the spectrum (particularly its dynamic range).
- If the acvs is such that:

$$\boxed{\sum_{\tau=-\infty}^{\infty} \left|\tau s_\tau\right| < \infty \text{ then } E\{\hat{S}^{(p)}(f)\} = S(f) + O(\frac{1}{N})}$$

- This only tells us the rate of change of bias, not the magnitude of the bias.
- Biases in periodograms are shown in next two slides for an AR(2) and AR(4) process




AR(2), N=16
A4(4), N=16,64
dB
Frequency
AR(2), N=64
dB
Frequency


AR(2), N=256
AR(4), N=256, 1024
dB
Frequency
AR(2), N=1024
dB
Frequency

Bias in Periodograms

- The bias in the periodograms is arising in the spectra that have high dynamic range.
- The biases arise from the sidelobes in the Fejer's kernel. This transfer of power is called leakage.
- There are two common techniques for lessening the biases:
  - Tapering : Modifies the kernel to reduce the sidelobes (at the expense of something else).
  - Pre-whitening : Pre-processes the time series to reduce the dynamic range.

Bias Reduction: Tapering

- Tapering a method reducing the side lobes in the Fejer's kernel.
- A products is formed $h_t X_t$ where h_t is a sequence of real values called a *data taper* (also known as a *data window, linear taper, linear window, fader and shading sequence*).
- With a data taper we have

$$\hat{S}^{(d)}(f) = \Delta t \left| \sum_{t=1}^{N} h_t X_t e^{-i2\pi f t \Delta t} \right|^2 \qquad H(f) = \Delta t \sum_{h=-1}^{N} h_t e^{-i2\pi f t \Delta t}$$

$$E\{\hat{S}^{(d)}(f)\} = \int_{-f_{(N)}}^{f_{(N)}} \mathrm{H}(f - f')S(f')df' \qquad \mathrm{H}(f) = \frac{1}{\Delta t}\left|H(f)\right|^2$$

Types of Tapers

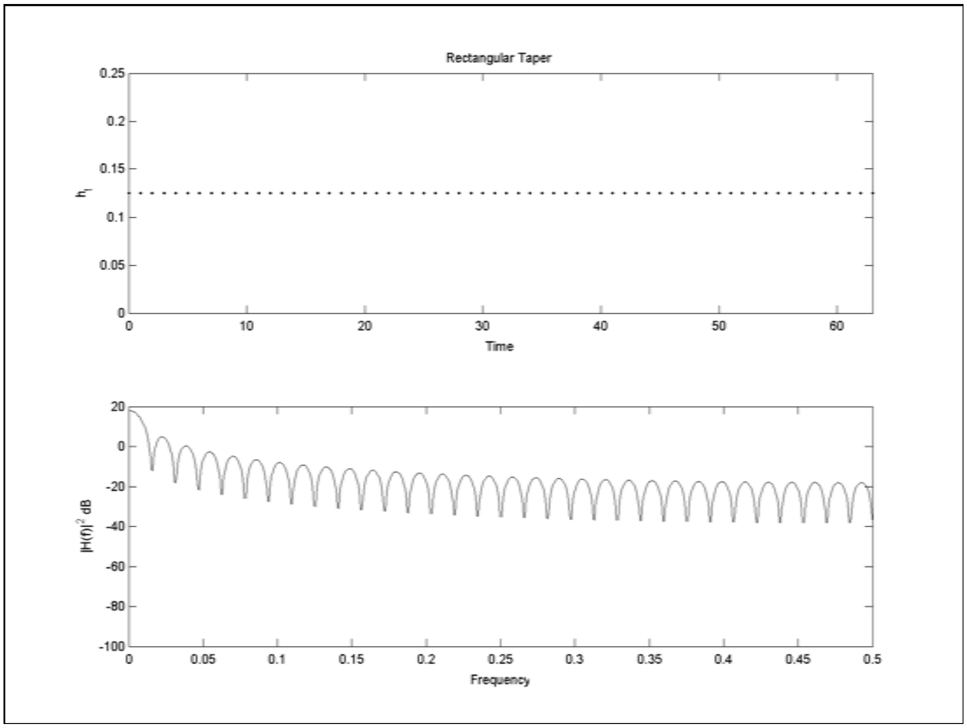
- Simplest: $h_t=1/\sqrt{N}$ for $1 \le t \le N$ called the *rectangle* or *default taper*. This case is the standard periodogram.
- By setting the normalization such that sum over N of $h_t^2=1$, the expectation integrated power under S(d) over the Nyquist frequency range will equal that under the true spectrum.
- The integral of $\mathcal{H}(f)$ over the Nyquist range will also equal 1.
- The idea is select h_t to minimize sidelobes. Since sharp edges generate ripples, one method is to smooth the edges of rectangular taper.
- One common method is a cosine taper.

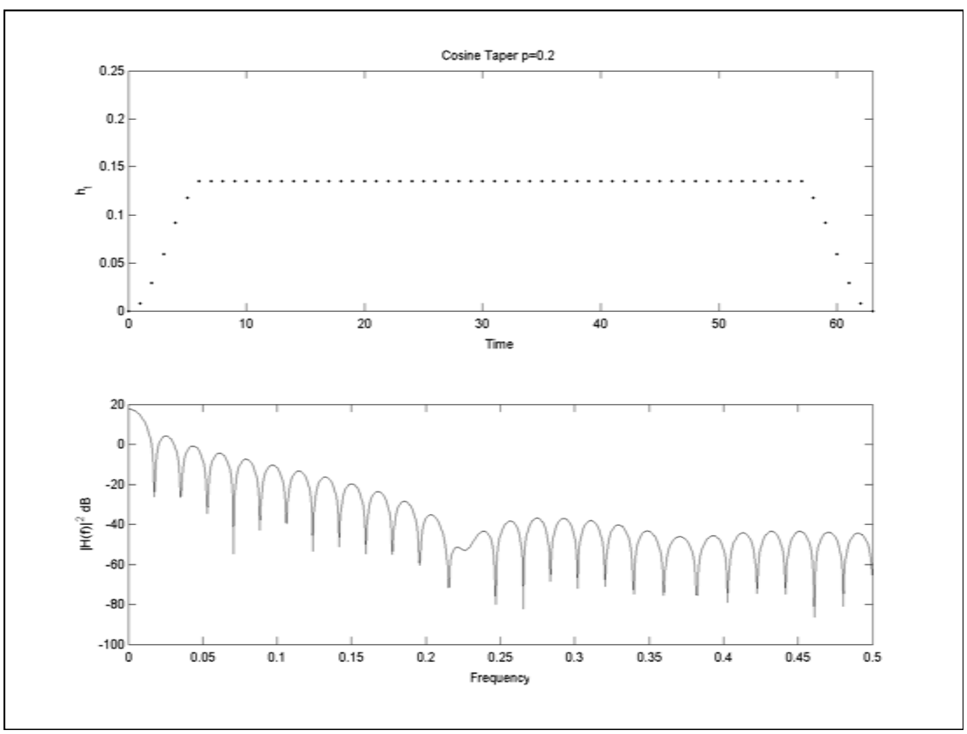
Cosine tapers

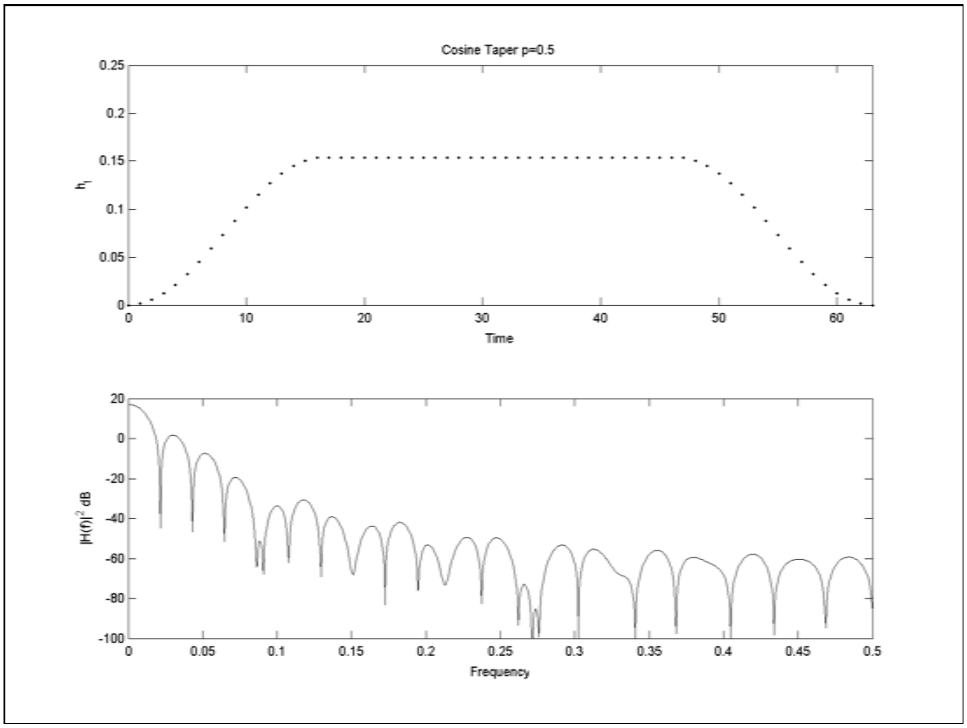
- The cosine taper called the *p% cosine taper* is defined by

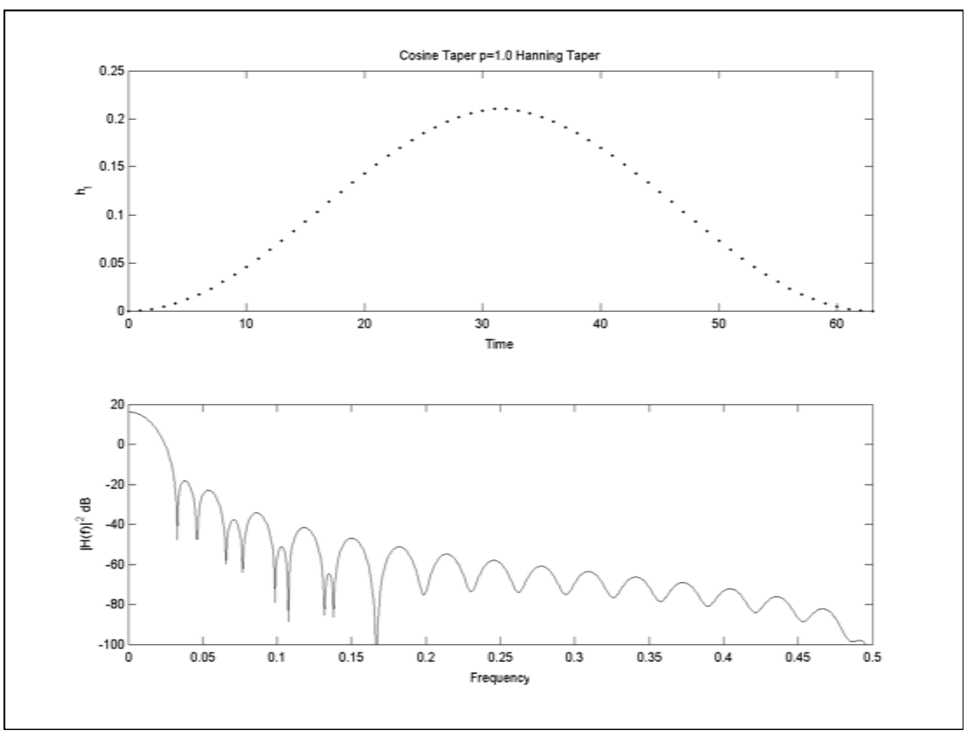
$$h_t = \begin{cases} \dfrac{C}{2}\left[1-\cos\left(\dfrac{2\pi t}{\lfloor pN \rfloor + 1}\right)\right] & 1 \le t \le \dfrac{\lfloor pN \rfloor}{2} \\ C & \dfrac{\lfloor pN \rfloor}{2} < t < N+1-\dfrac{\lfloor pN \rfloor}{2} \\ \dfrac{C}{2}\left[1-\cos\left(\dfrac{2\pi t}{\lfloor pN \rfloor + 1}\right)\right] & N+1-\dfrac{\lfloor pN \rfloor}{2} \le t \le N \end{cases}$$

- C is a constant such that the sum of h_t is 1.
- When p is 100%, the taper called a *Hanning Taper*.


Rectangular Taper
h_t
Time
|H(f)|^2 dB
Frequency


Cosine Taper p=0.2
h_t
Time
|H(f)|^2 dB
Frequency


Cosine Taper p=0.5
h_t
Time
|H(f)|^2 dB
Frequency


Cosine Taper p=1.0 Hanning Taper
h_t
Time
|H(f)|^2 dB
Frequency

Summary of Today's class

- Non-parametric Spectral Estimation
  - Estimation of Mean: Possible problems with some processes
  - Estimation of autocovariance sequence: Concepts of biased and unbiased estimators
  - Naïve spectral estimation: Periodograms
    - Leakage and bias in this method
    - Depends on dynamic range.
  - Bias reduction - Tapering: One method for reducing bias.