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

- Stochastic Spectral Estimation
  - Spectral representation of a stationary process
  - Basic properties of the spectrum
  - Classification of spectra
  - Sampling and Aliasing
  - Comparisons of spectral density function and auto covariance sequence

Stochastic Spectral Estimation

- Spectral Representation:
  - We have discussed representations for deterministic functions or sequences in terms of linear combinations of sinusoids with different frequencies, and defined energy or power from them.
  - Assuming square integrability and in a mean square sense, we have:
    - periodic function → discrete frequencies → infinite energy but finite power over discrete set of frequencies.
    - non-periodic function → continuous frequency → finite energy
  - We now need to find a way to represent a stationary process in terms of a sum of sinusoids so that we can define its spectrum, just as we have done for deterministic functions.
  - Stationary processes have constant variance and, with the exception of harmonic processes, a typical realization has infinite energy.

Spectral Representation

- For stationary stochastic processes, a spectral representation exits
- We motivate the spectral representation by considering a real valued discrete time harmonic process:

$$\boxed{X_t = \sum_{l=1}^{L} D_l \cos(2\pi f_l t + \phi_l) \text{ where } \phi_l \text{ is uniform dist. } [-\pi,\pi]}$$

- D_l and f_l are constants. We assume Δt=1 and hence the frequencies are 0< f_l <1/2.

Spectral Representation

- We can rewrite the equation as

$$X_t = \sum_{l=-L}^{L} C_l e^{i2\pi f_l t} \text{ where } C_l = D_l e^{i\phi_l}/2 \text{ and } C_{-l} = D_l e^{-i\phi_l}/2$$

- With zero expectation for X_t, $C_0=0$ at $f_0=0$. Since the phase are uncorrelated random variables, C_{-l} and C_l are uncorrelated (although $C_{-l} = C^*_l$).
- We have $E\{C_l\}=0$ and $var\{C_l\}=D_l^2/4$.
- We can define the *variance spectrum* by

$$S^{(V)}(f) \equiv \begin{cases} D_l^2/4, \text{ if } f = f_l, l = 0, \pm 1, \ldots, \pm L \\ 0, \quad \text{otherwise} \end{cases}$$

Spectral Representation

- We now define complex stochastic process

$$Z(f) \equiv \sum_{j=0}^{L} C_j, \quad f_l < f \le f_{l+1}, \text{with } l = 0,1,\ldots,L$$

$$Z(f) = \begin{cases} 0, \text{ for } 0 \le f \le f_1 \\ C_1, \text{ for } f_1 < f \le f_2 \\ C_1 + C_2, \text{ for } f_2 < f \le f_3 \\ C_1 + C_2 + C_3, \text{ for } f_3 < f \le f_4 \end{cases}$$

$$dZ(f) = \begin{cases} Z(f+df) - Z(f), \text{ for } 0 \le f < 1/2 \\ 0, \text{ for } f = 1/2 \\ dZ^*(-f), \text{ for } -1/2 \le f < 0 \end{cases}$$

df is small positive increment

Spectral Representation

- We therefore have for l>0

$$\boxed{\begin{array}{l} dZ(f_l) = Z(f_l + df) - Z(f_l) = \sum_{j=0}^{l} C_j - \sum_{j=0}^{l-1} C_j = C_l \\ \text{for any } f \neq f_l \ dZ(f) = 0 \end{array}}$$

- The covariance between dZ(f) and dZ(f') for f≠f' is zero and the {Z(f)} process is said to have *orthogonal increments* and the process is an *orthogonal process*.
- Note that

$$\boxed{E\{|dZ(f_l)|^2\} = E\{|C_l|^2\} = D_l^2/4}$$

Spectral Representation

- Now let g(.) be a continuous function over the interval [-1/2,1/2] and let H(.) be a step function with jumps at $-1/2<a_1<a_2<\ldots<a_n<1/2$ with finite sizes $b_1,b_2,..,b_n$.
- From the definition of the Reimann-Stieltjes integral

$$\boxed{\begin{array}{l} \int_{-1/2}^{1/2} g(f)dH(f) = \sum_{k=1}^{N} g(a_k)b_k \text{ with } g(f) = e^{i2\pi ft} \text{ and } H(f) = Z(f) \\ \text{we have } X_t = \int_{-1/2}^{1/2} e^{i2\pi ft} dZ(f) \end{array}}$$

- This is called the *spectral representation* of a stationary process
- By allowing N (here) and L on slide 5 to go to infinity, the equation above applies to any discrete valued stationary process.

Properties of Z(f)

- The orthogonal process Z(f) has the following properties:
  - E{dZ(f)}=0 for al |f|<1/2
  - E{| dZ(f)|^2}=dS$^{(I)}$(f) for all |f|≤1/2 where the bounded, non-decreasing function S$^{(I)}$(f) is called the *integrated spectrum* of {X_t}.
  - For any two frequencies f≠f' in the interval [-1/2,1/2] cov{dZ(f),dZ(f')}=E{dZ*(f),dZ(f')}=0
- The equation of {X_t} says any stationary process can be represented by sum of complex exponentials with random amplitudes and phases; and the square modulus of dZ(f) defines the integrated spectrum.

Relationship between acvs and S$^{(I)}$(f)

- We can write the relationship between the autocovariance sequence (s_τ) and the integrated spectrum using:

$$X_t^* X_{t+\tau} = \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} e^{-i2\pi f' t} e^{i2\pi f(t+\tau)} dZ^*(f')dZ(f)$$

$$s_\tau = E\{X_t^* X_{t+\tau}\} = \int_{-1/2}^{1/2}\int_{-1/2}^{1/2} e^{i2\pi(f-f')t} e^{i2\pi f\tau} E\{dZ^*(f')\}E\{dZ(f)\}$$

$$s_\tau = \int_{-1/2}^{1/2} e^{i2\pi f\tau} E\{|dZ(f)|^2\} = \int_{-1/2}^{1/2} e^{i2\pi f\tau} dS^{(I)}(f) \text{ due to orthogonal increments}$$

If $S^{(I)}(f)$ is differentiable then

$$E\{|dZ(f)|^2\} = dS^{(I)}(f) = S(f)df$$

S(f) is the spectral density function (sdf) and

$$s_\tau = \int_{-1/2}^{1/2} S(f)e^{i2\pi f\tau} df \text{ and the inverse relation } S(f) = \sum_{t=-\infty}^{\infty} s_\tau e^{-i2\pi ft}$$

Comments

- From the previous page we can show:

$$\boxed{\mathrm{var}(X_t) = s_0 = \int_{-1/2}^{1/2} S(f)df}$$

- The sdf S(f) is often called the *power spectral density function (psdf)*
- For the continuous time case:

$$\boxed{s(\tau) = \int_{-\infty}^{\infty} S(f)e^{i2\pi f\tau}df \text{ and } S(f) = \int_{-\infty}^{\infty} s(\tau)e^{-i2\pi f\tau}d\tau}$$

Properties of the Spectrum

- From the definition of the integrated spectrum we have

$$\boxed{\begin{array}{l} S^{(I)}(f) = \int_{-1/2}^{f} S(f')df' \quad 0 \le S^{(I)}(f) \le \sigma^2 \\ S^{(I)}(-1/2) = 0 \quad S^{(I)}(1/2) = \sigma^2 \end{array}}$$

- $S^{(I)}$(f) exists for all stationary processes by S(f) may not exist if $S^{(I)}$(f) is not differentiable. (Introduction of Dirac functions can solve this problem).
- S(f) for a white noise process is constant and equal to σ^2 (remember is integration is from [-1/2,1/2].
- In some definitions, the integral is from [0,1/2] in which case S(f) is doubled (3dB on log plot)

Classification of Spectra.

- Integrated spectra are similar to probability distribution functions (with S(f) corresponding to probability density functions).
- We can write $S(I)(f)=S_1^{(I)}(f)+S_2^{(I)}(f)+S_3^{(I)}(f)$ where
  - $S_1^{(I)}(f)$ is absolutely continuous (derivative every)
  - $S_2^{(I)}(f)$ a step function with jumps at specific frequencies
  - $S_3^{(I)}(f)$ is continuous singular function (this latter case is pathological and not normally encountered).
- $S_1^{(I)}(f)$ are *purely continuous spectrum* and their avcs decays to zero as τ goes to infinity. Most ARMA processes are of this type
- $S_2^{(I)}(f)$ are *purely discrete spectrum* and the acvs does not decay to zero. Harmonic processes are of this type.
- Combinations are possible. Discrete with white sdf is a *discrete spectrum* while discrete with non-white is *mixed spectrum*.

Sampling and Aliasing

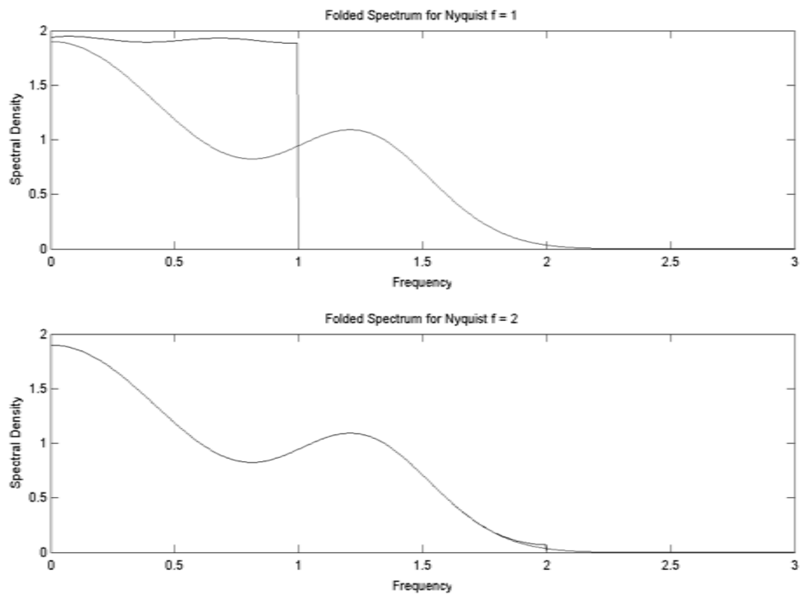
- We can define a discrete sampling of a continuous time series as $X_t=X(t_0+t\Delta t)$, $t=0,\pm1,\pm2,\ldots$
- If $\{X(t)\}$ is stationary then $\{X_t\}$ is also stationary.
- Based on the sampling (notice we still have an infinite number of samples)

$$s_\tau = \text{cov}\{X_t, X_{t+\tau}\} = \text{cov}\{X(t_0+t\Delta t), X(t_0+[t+\tau]\Delta t)\} = s(\tau\Delta t)$$

and hence s_τ is s(.) sampled at intervals ofΔt

$$S_{X_t}(f) = \sum_{k=-\infty}^{\infty} S_{X(t)}(f + k/\Delta t), \quad |f| \le \frac{1}{2\Delta t} \equiv f_{(N)}$$

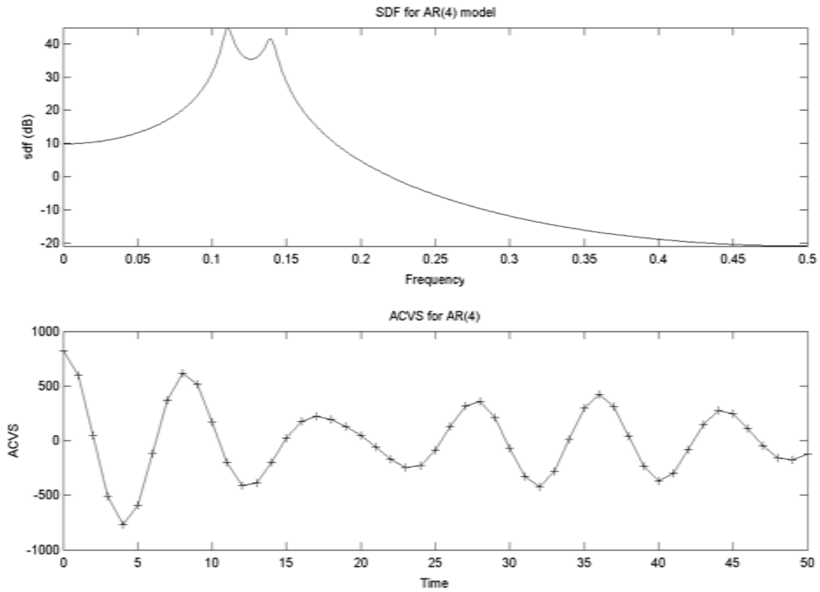
Example of aliasing



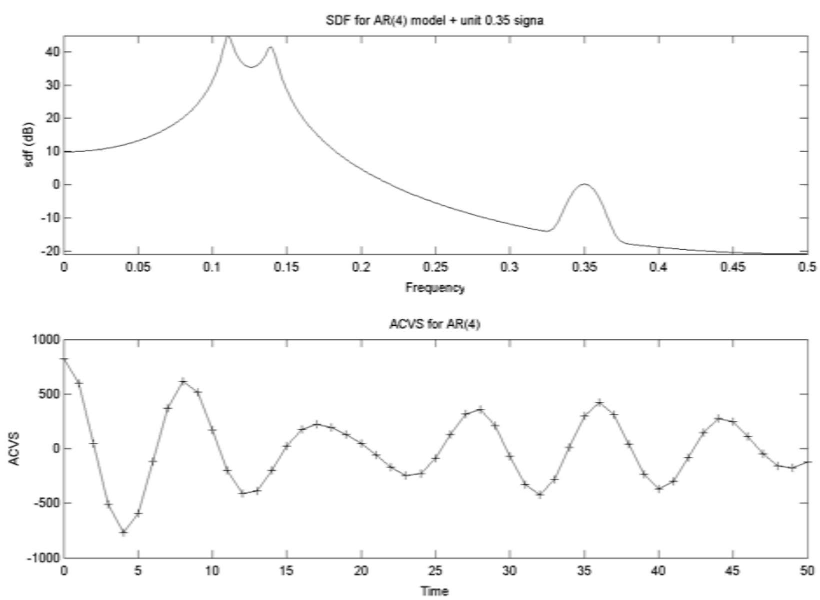
Comparison of SDF and ACVS

- In the example just shown, the sampling interval of 1/2 would seem to under sampled, however, these samples will be close to uncorrelated and this could be advantageous. Some suggest sampling at twice the Nyquist rate to avoid problems near the Nyquist rate (called *oversampling*).
- While the SDF and ACVS contain the same information (Fourier transform of each other), often structure can be seen in the SDF that is not obvious in the ACVS as shown in the next figures.

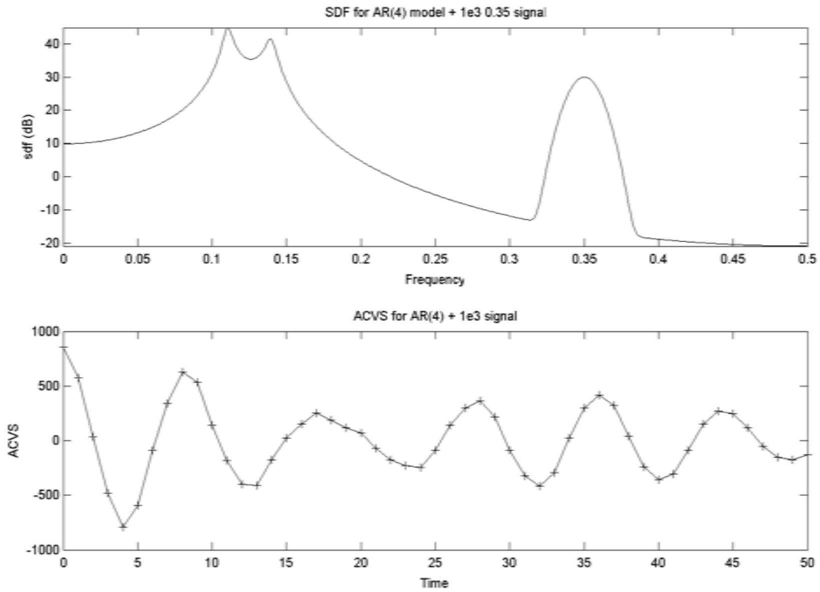
AR(4) SDF and AVCS



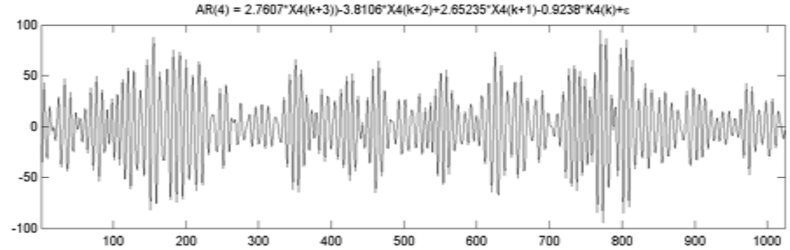
AR(4) SFD and AVCS + 0.35 f



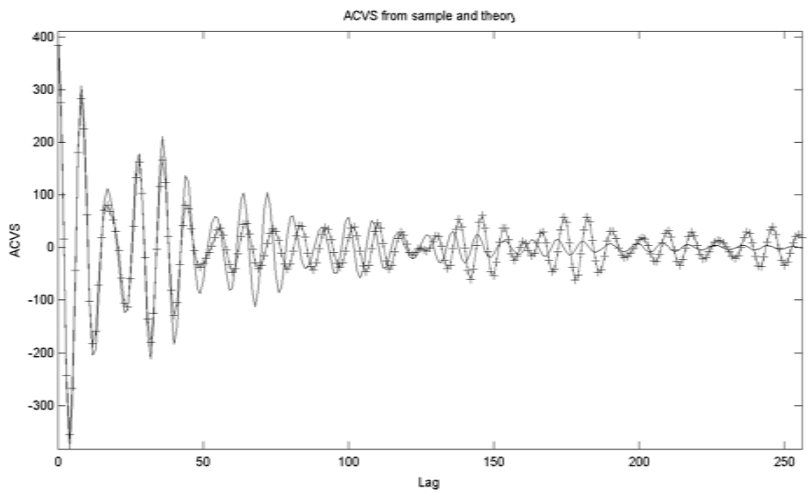
AR(4) SFD and AVCS + bigger f



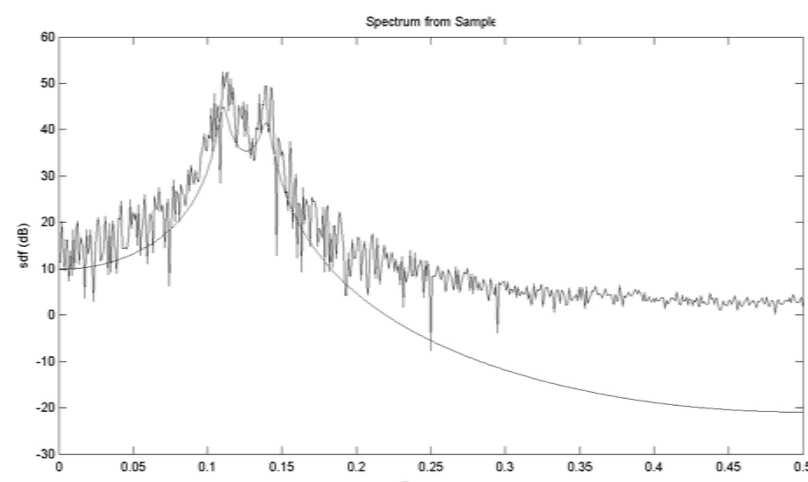
Example of AR(4)



Comparison of sample acvs with theory



Spectrum from sample



Summary

- Stochastic Spectral Estimation
  – Spectral representation of a stationary process
  – Basic properties of the spectrum
  – Classification of spectra
  – Sampling and Aliasing
  – Comparisons of spectral density function and auto covariance sequence: Note that changes to the spectrum can be clear but not so obvious is acvs.