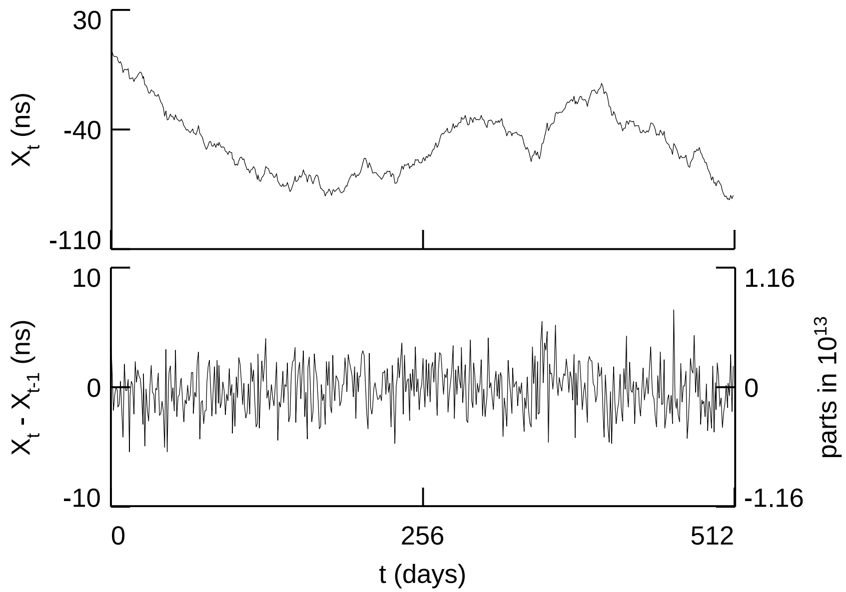
Clock Statistics: A Tutorial

Don Percival

Applied Physics Laboratory
University of Washington, Seattle

Motivating Example: I

- consider following measurements:



  – top: X_t = time (phase) difference between clock 55 and USNO time scale at day t (adjusted for systematic drift)

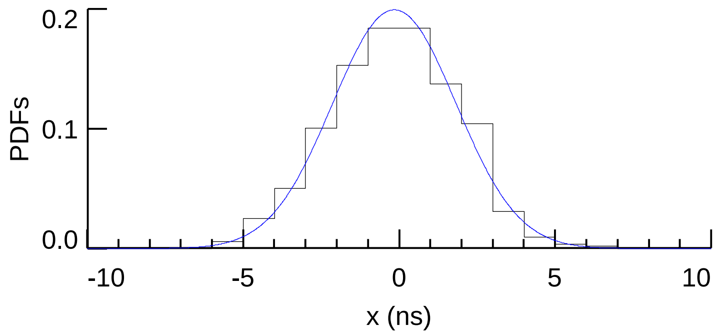
  – bottom: $X_t^{(1)} = X_t - X_{t-1} \propto$ fractional frequency deviate averaged over one day

Motivating Example: II

- clock statistics used to summarize performance
  - if $X_t^{(1)}$ constant, clock 55 agrees with time scale (essentially)
  - $X_t^{(1)}$ has stochastic (noise-like) fluctuations
  - statistics used to quantify fluctuations
- sample statistics
  - mean: $\hat{\mu} = \frac{1}{N}\sum_{t=0}^{N-1} X_t^{(1)}$
    (here $N = 512 = \#$ of measurements)
  - variance: $\hat{\sigma}^2 = \frac{1}{N}\sum_{t=0}^{N-1}(X_t^{(1)} - \hat{\mu})^2$
  - $\hat{\sigma}$ (standard deviation) is measure of spread
- easiest to interpret $\hat{\mu}$ & $\hat{\sigma}$ if data taken to be independent samples from Gaussian (i.e., normal) distribution

Motivating Example: III

- Q: is Gaussian assumption reasonable?
- comparison of histogram to probability density function:



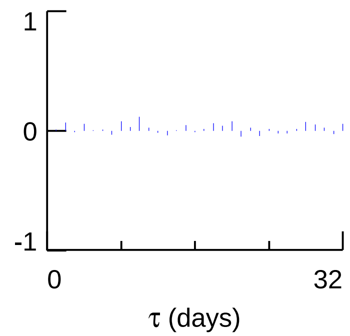
  – Gaussian assumption seems reasonable

Motivating Example: IV

- Q: is independent assumption reasonable?
- under Gaussianity, uncorrelatedness implies independence
- sample autocorrelation sequence measures uncorrelatedness:

$$\hat{\rho}_\tau = \frac{\sum_{t=0}^{N-\tau-1}(X_t^{(1)} - \hat{\mu})(X_{t+\tau}^{(1)} - \hat{\mu})}{\sum_{t=0}^{N-1}(X_t^{(1)} - \hat{\mu})^2}, \quad \tau = 1, 2, \ldots, N-1$$

- can interpret $\hat{\rho}_\tau$ as correlation coefficient:



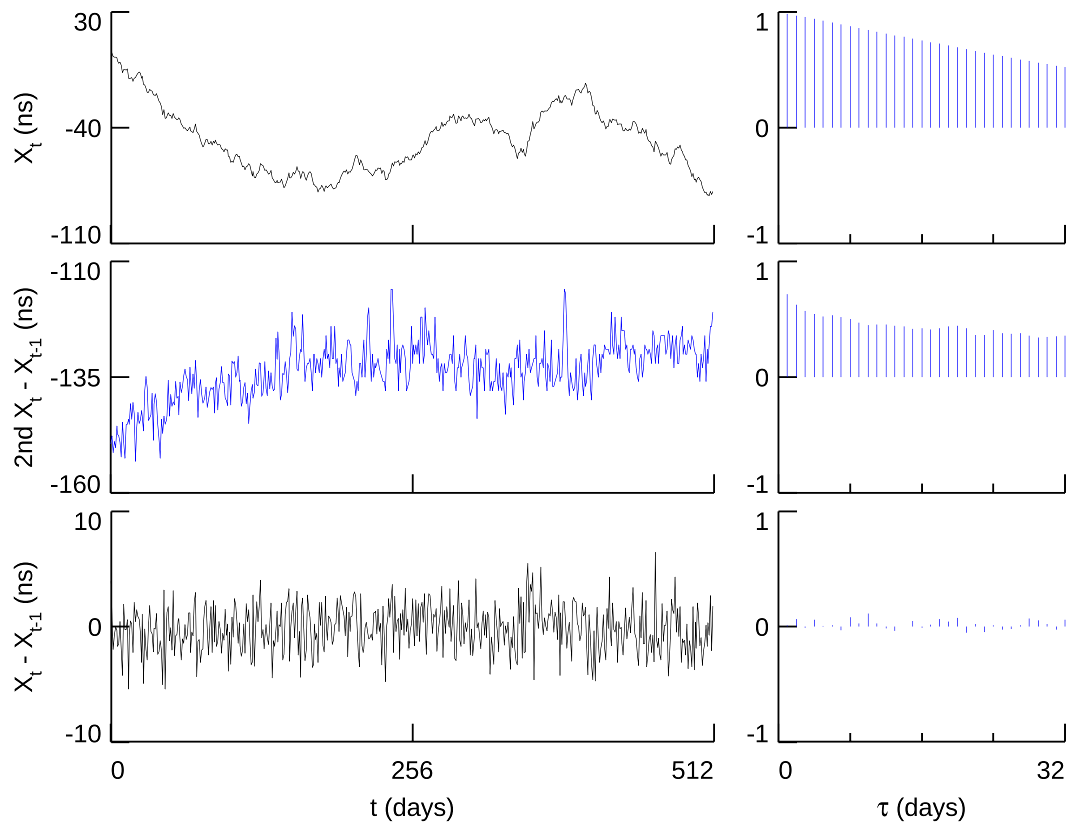
– since $\hat{\rho}_\tau \approx 0$, uncorrelatedness seems reasonable

Conclusions from Motivating Example

- $X_t^{(1)}$ well-modeled as uncorrelated Gaussian deviates (sometimes called Gaussian white noise)
- theory says $\hat{\mu}$ & $\hat{\sigma}^2$ are sufficient statistics for summarizing statistical information about clock 55
- implies 'random walk' model for time difference data X_t
- seems we need little more than what is taught in 'Statistics 101'

Reality Bites!

- alas, other clocks do not have such simple statistical properties



- $\hat{\mu}$ & $\hat{\sigma}^2$ not sufficient summaries for clock in middle plot

Overview of Remainder of Tutorial

- discussion of models for interpreting clock statistics
  - models specified via spectrum (spectral density function)
  - while white noise & random walk models depend on μ & σ^2, more comprehensive models depend on μ and spectrum
  - in simplest case, spectrum itself depends on 2 parameters
    * σ_ϵ^2, a parameter setting overall level of spectrum
    * α, a so-called 'power law' parameter
- look at clock statistics based upon 2 variance decompositions
  - spectral analysis
  - wavelet analysis

The Spectrum

- let X_t be a stochastic process, i.e., collection of random variables (RVs) indexed by t
- suppose further that X_t is stationary
- implies certain theoretical properties do not change with time
- in particular, its variance $\sigma^2 = \mathrm{var}\,\{X_t\}$ is the same for all t
- spectrum $S_X(\cdot)$ decomposes σ^2 across frequencies f:

$$\mathrm{var}\,\{X_t\} = \int_{-1/2}^{1/2} S_X(f)\,df$$

here f is a Fourier frequency with units of cycles per unit time (e.g., cycles per day for process sampled once per day)

Physical Interpretation of Spectrum via Filtering

- let a_u be a filter, and form $Y_t = \sum_{u=-\infty}^{\infty} a_u X_{t-u}$
- Y_t has spectrum $S_Y(f) = \mathcal{A}(f) S_X(f)$, where

$$\mathcal{A}(f) = \Big| \sum_{u=-\infty}^{\infty} a_u e^{-i2\pi f u} \Big|^2 \text{ is squared gain function}$$

- if a_u narrow-band of bandwidth Δf about f, i.e.,

$$\mathcal{A}(f') = \begin{cases} \frac{1}{2\Delta f}, & f - \frac{\Delta f}{2} \le |f'| \le f + \frac{\Delta f}{2} \\ 0, & \text{otherwise,} \end{cases}$$

  then have following interpretation for $S_X(f)$:

$$\operatorname{var}\{Y_t\} = \int_{-1/2}^{1/2} S_Y(f')\, df' = \int_{-1/2}^{1/2} \mathcal{A}(f') S_X(f')\, df' \approx S_X(f)$$

Spectrum for White Noise Process

- simplest stationary process is white noise
- ϵ_t is white noise process if
  - $E\{\epsilon_t\} = \mu_\epsilon$ for all t (usually take $\mu_\epsilon = 0$), where $E\{\epsilon_t\}$ denotes expected value of RV ϵ_t
  - $\text{var}\,\{\epsilon_t\} = \sigma_\epsilon^2$ for all t
  - ϵ_t and $\epsilon_{t'}$ are uncorrelated for all $t \neq t'$
- spectrum for white noise is just $S_\epsilon(f) = \sigma_\epsilon^2$
- note that

$$\int_{-1/2}^{1/2} S_\epsilon(f)\,df = \int_{-1/2}^{1/2} \sigma_\epsilon^2\,df = \sigma_\epsilon^2 = \text{var}\,\{\epsilon_t\},$$

as required

First Order Backward Difference of White Noise

- consider first order backward difference of white noise:

$$X_t = \epsilon_t - \epsilon_{t-1} = \sum_{u=-\infty}^{\infty} a_u \epsilon_{t-u} \;\&\; a_u = \begin{cases} 1, & u = 0; \\ -1, & u = 1; \\ 0, & \text{otherwise.} \end{cases}$$

- squared gain function is

$$\mathcal{A}(f) = \left| \sum_{u=-\infty}^{\infty} a_u e^{-i2\pi f u} \right|^2 = \left|1 - e^{-i2\pi f}\right|^2 = |2\sin(\pi f)|^2$$

- have $S_X(f) = \mathcal{A}(f) S_\epsilon(f) = \sigma_\epsilon^2 \, |2\sin(\pi f)|^2$
- note that $S_X(f) \approx \sigma_\epsilon^2 \, |2\pi f|^2$ at low frequencies (using $\sin(x) \approx x$ for small x)

Higher Order Backward Differences of White Noise

- let B be backward shift operator: $B\epsilon_t = \epsilon_{t-1}$, $B^2\epsilon_t = \epsilon_{t-2}$, $(1-B)\epsilon_t = \epsilon_t - \epsilon_{t-1}$, etc.
- consider dth order backward difference of white noise:

$$X_t = (1-B)^d \epsilon_t = \sum_{k=0}^{d} \frac{d!}{k!(d-k)!}(-1)^k \epsilon_{t-k}$$
$$= \sum_{k=0}^{\infty} \frac{\Gamma(1+\frac{\alpha}{2})}{\Gamma(k+1)\Gamma(1+\frac{\alpha}{2}-k)}(-1)^k \epsilon_{t-k}$$

with $\alpha = 2d$, i.e., $\alpha = 2, 4, \ldots$

- spectrum given by

$$S_X(f) = \mathcal{A}(f)S_\epsilon(f) = \sigma_\epsilon^2 \left|2\sin(\pi f)\right|^\alpha \approx \sigma_\epsilon^2 \left|2\pi f\right|^\alpha$$

Fractional Differences of White Noise

- for α not necessary an integer,

$$X_t = \sum_{k=0}^{\infty} \frac{\Gamma(1+\frac{\alpha}{2})}{\Gamma(k+1)\Gamma(1+\frac{\alpha}{2}-k)}(-1)^k \epsilon_{t-k} = \sum_{k=0}^{\infty} a_k(\alpha)\epsilon_{t-k}$$

  makes sense as long as $\alpha > -1$

- X_t is stationary fractionally differenced (FD) process
- note: FD processes introduced in 1980 paper co-authored by C.W.J. Granger, co-winner of 2003 Nobel Prize for economics!
- spectrum is as before:

$$S_X(f) = \sigma_\epsilon^2 \, |2\sin(\pi f)|^\alpha \approx \sigma_\epsilon^2 \, |2\pi f|^\alpha$$

- obeys power law at low frequencies with exponent α
- note: FD process reduces to white noise when $\alpha = 0$

Nonstationary FD Processes: I

- let $X_t^{(1)}$ be FD process with parameter $-1 < \alpha^{(1)} \leq 1$
- define X_t as cumulative sum of $X_t^{(1)}$: $X_t = \sum_{l=0}^{t} X_l^{(1)}$
- since

$$X_t^{(1)} = X_t - X_{t-1} \;\&\; S_{X^{(1)}}(f) = \sigma_\epsilon^2 \, |2\sin(\pi f)|^{\alpha^{(1)}},$$

filtering theory suggests using relationship

$$S_{X^{(1)}}(f) = |2\sin(\pi f)|^2 S_X(f)$$

to *define* spectrum for X_t, i.e.,

$$S_X(f) = \frac{S_{X^{(1)}}(f)}{|2\sin(\pi f)|^2} = \sigma_\epsilon^2 \, |2\sin(\pi f)|^{\alpha}$$

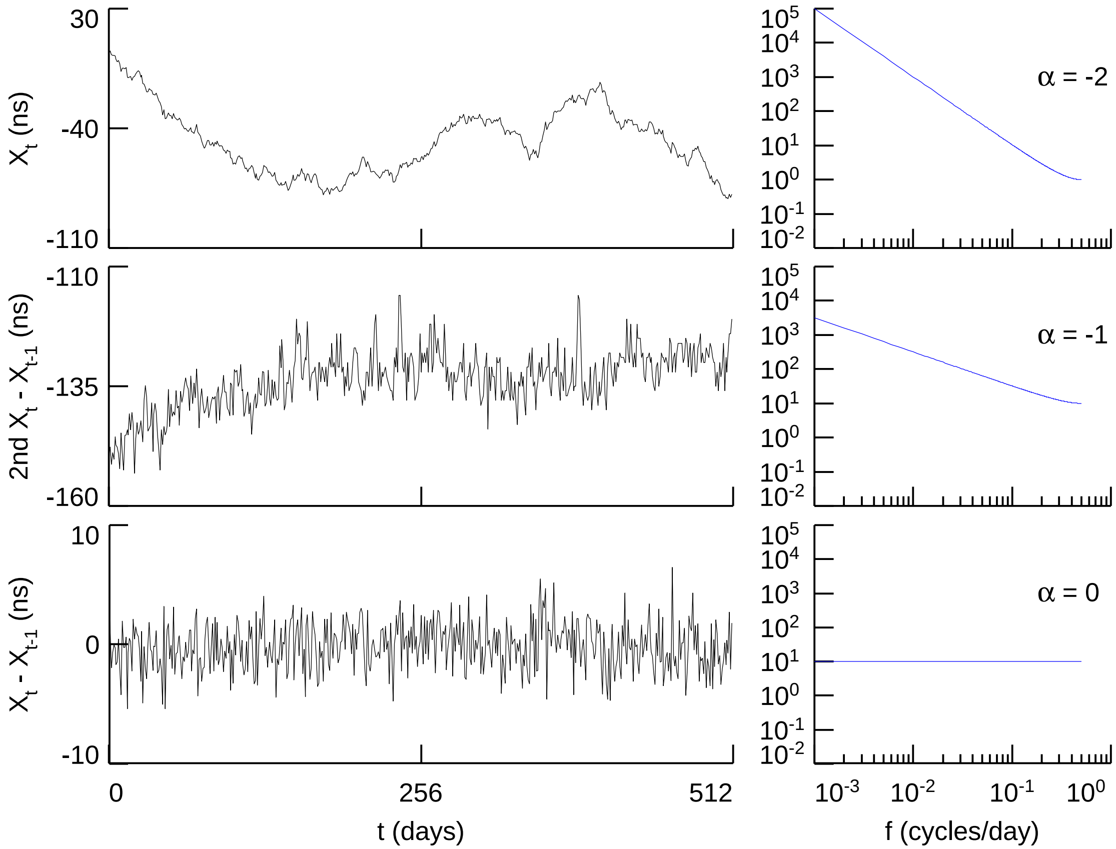
with $\alpha = \alpha^{(1)} - 2$

Nonstationary FD Processes: II

- X_t said to have stationary 1st order backward differences
- special case: if $\alpha^{(1)} = 0$ so that $X_t^{(1)}$ is white noise, then X_t is a random walk process and has spectrum
$$S_X(f) = \sigma_\epsilon^2 \left|2\sin(\pi f)\right|^{-2} \approx \sigma_\epsilon^2 \left|2\pi f\right|^{-2};$$
i.e., random walk is FD process with $\alpha = -2$
- one cumulative sum defines FD processes for $-3 < \alpha \leq -1$
- two cumulative sums define FD processes for $-5 < \alpha \leq -3$
- special case: if $X_t^{(2)}$ is white noise and if
$$X_t^{(1)} = \sum_{l=0}^{t} X_l^{(2)} \;\; \& \;\; X_t = \sum_{l=0}^{t} X_l^{(1)},$$
X_t is a random run, and $S_X(f) \approx \sigma_\epsilon^2 \left|2\pi f\right|^{-4}$ so $\alpha = -4$

Examples of Spectra for FD Processes

- three examples of clock noise well-modelled by FD processes



- on log/log plot, power law spectra appear linear with slope α

Summary of FD Processes

- X_t said to be FD process if its spectrum is given by
$$S_X(f) = \sigma_\epsilon^2 \, |2\sin(\pi f)|^\alpha$$
- well-defined for any real-valued exponent α
- at low frequencies, have $S_X(f) \approx \sigma_\epsilon^2 \, |2\pi f|^\alpha$; i.e., FD spectrum is approximately a power law with exponent α
- if $\alpha > -1$, FD process stationary
- if $\alpha \leq -1$, FD process nonstationary but its dth order backward difference is stationary FD process with parameter $\alpha^{(d)}$, where
$$d = 1 + \left\lfloor \frac{-\alpha - 1}{2} \right\rfloor \text{ and } \alpha^{(d)} = \alpha + 2d$$
(here $\lfloor x \rfloor$ is largest integer $\leq x$)

Generalization: Composite FD Process

- FD process not always an adequate model, so of interest to consider generalizations
- suppose $X_t(\alpha_m)$ is FD process with power law α_m and $\sigma_\epsilon^2 = 1$
- suppose $X_t(\alpha_m)$ & $X_t(\alpha_{m'})$ are independent when $m \neq m'$
- form composite FD process $X_t = \sum_{m=0}^{M-1} a_m X_t(\alpha_m)$
- has spectrum given by

$$S_X(f) = \sum_{m=0}^{M-1} a_m^2 \, |2\sin(\pi f)|^{\alpha_m}$$

Generalization: ARFIMA Process

- autoregressive, fractionally integrated, moving average
- idea is to replace ϵ_t in

$$X_t = \sum_{k=0}^{\infty} a_k(\alpha)\epsilon_{t-k}$$

with ARMA process U_t (models high-frequency part of noise):

$$U_t = \sum_{k=1}^{p} \phi_k U_{t-k} + \epsilon_t - \sum_{k=1}^{q} \theta_k \epsilon_{t-k}$$

- yields process with spectrum

$$S_X(f) = \sigma_\epsilon^2 \left|2\sin(\pi f)\right|^\alpha \frac{\left|1 - \sum_{k=1}^{q} \theta_k e^{-i2\pi f k}\right|^2}{\left|1 - \sum_{k=1}^{p} \phi_k e^{-i2\pi f k}\right|^2}$$

Generalization: Time-Varying FD Process

- can define time-varying FD (TVFD) process via

$$X_t = \sum_{k=0}^{\infty} a_k(\alpha_t)\epsilon_{t-k}$$

as long as $\alpha_t > -1$ for all t

- can use representation

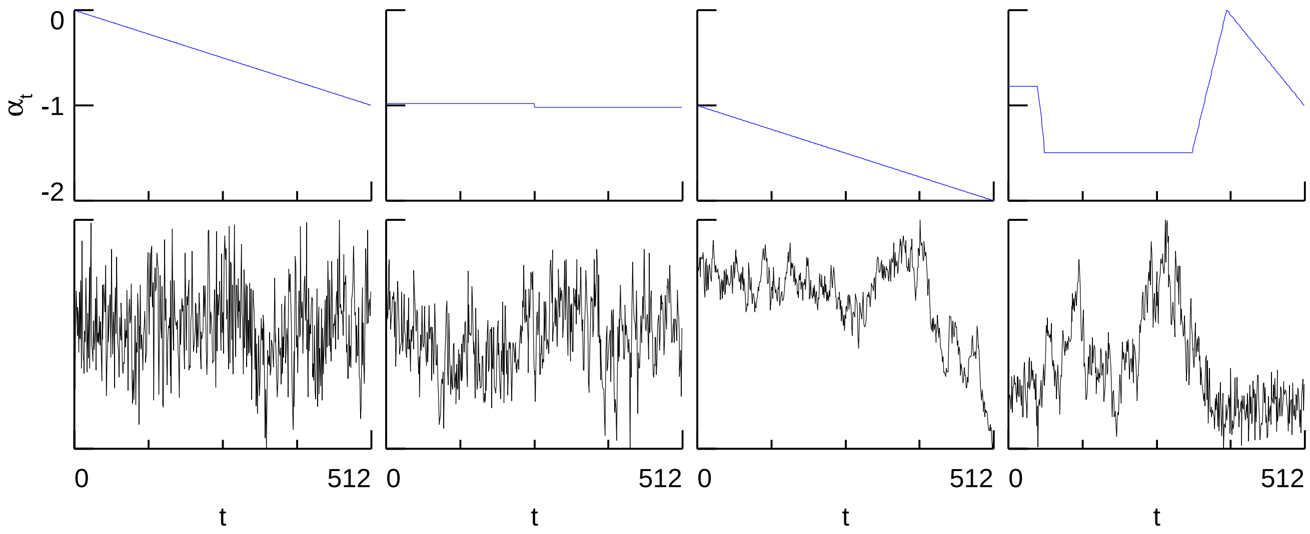
$$X_t = \sum_{k=0}^{2N-1} c_{t,k}(\alpha_t)\varepsilon_k, \quad t = 0, 1, \ldots, N-1,$$

to extend definition to handle arbitrary α_t

- can also make σ_ϵ^2 time-varying

Examples of Time-Varying FD Processes

- realizations from four TVFD processes



FD Process Parameter Estimation

- Q: given sample $X_0, \ldots, X_{N-1}$ that is assumed to be realization of FD process, how can we estimate α & σ_ϵ^2?
- *many* different estimators have been proposed!
  (area of active research)
- will concentrate on estimators based on
  - spectral analysis (frequency-based)
  - wavelet analysis (scale-based)

Why Spectral and Wavelet Analysis?

- both physically interpretable
- both are analysis of variance techniques
  - useful for more than just estimating α & σ_ϵ^2
  - provide useful characterizations of clock performance
- can assess need for models more complex than FD process (e.g., composite FD process)
- provide preliminary estimates for more complicated schemes (maximum likelihood estimation)

Estimation via Spectral Analysis

- recall that spectrum for FD process given by

$$S_X(f) = \sigma_\epsilon^2 \, |2\sin(\pi f)|^\alpha$$

and thus

$$\log\left(S_X(f)\right) = \log\left(\sigma_\epsilon^2\right) + \alpha \log\left(|2\sin(\pi f)|\right);$$

i.e., plot of $\log\left(S_X(f)\right)$ vs. $\log\left(|2\sin(\pi f)|\right)$ linear with slope α

- for $0 < f < 1/8$, have $\sin(\pi f) \approx \pi f$, so

$$\log\left(S_X(f)\right) \approx \log\left(\sigma_\epsilon^2\right) + \alpha \log\left(2\pi f\right);$$

i.e., plot of $\log\left(S_X(f)\right)$ vs. $\log\left(2\pi f\right)$ approximately linear at low frequencies with slope α

Basic Spectral Estimation Scheme

- estimate $S_X(f)$ via $\hat{S}_X(f)$
- fit linear model to $\log(\hat{S}_X(f))$ vs. $\log(2\pi f)$ over low f's
- use estimated slope $\hat{\alpha}$ to estimate α
- manipulate estimated intercept to estimate σ_ϵ^2
- lots of possible estimators $\hat{S}_X(f)$ in the literature
- will consider periodogram & multitaper spectral estimator

The Periodogram: I

- basic estimator of $S_X(f)$ is periodogram:

$$\hat{S}_X^{(p)}(f) = \frac{1}{N}\Big|\sum_{t=0}^{N-1}(X_t - \hat{\mu})e^{-i2\pi ft}\Big|^2$$

- gives decomposition of sample variance:

$$\int_{-1/2}^{1/2} \hat{S}_X^{(p)}(f)\,df = \hat{\sigma}^2 = \frac{1}{N}\sum_{t=0}^{N-1}(X_t - \hat{\mu})^2$$

The Periodogram: II

- for stationary processes & large N, theory says

$$\hat{S}_X^{(p)}(f) \stackrel{\mathrm{d}}{=} S_X(f)\frac{\chi_2^2}{2}, \quad 0 < f < 1/2,$$

approximately, implying that

  - $E\{\hat{S}_X^{(p)}(f)\} \approx E\{S_X(f)\chi_2^2/2\} = S_X(f)$
  - $\mathrm{var}\,\{\hat{S}_X^{(p)}(f)\} \approx \mathrm{var}\,\{S_X(f)\chi_2^2/2\} = S_X^2(f)$
    * '$\stackrel{\mathrm{d}}{=}$' means 'equal in distribution'
    * χ_2^2 is chi-square RV with 2 degrees of freedom

- $\hat{S}_X^{(p)}(f_j)$ and $\hat{S}_X^{(p)}(f_k)$ approximately uncorrelated for $f_j = \frac{j}{N}$, $f_k = \frac{k}{N}$ and $0 < f_j < f_k < 1/2$

The Periodogram: III

- taking log transform yields

$$\log\left(\hat{S}_X^{(p)}(f)\right) \stackrel{\mathrm{d}}{=} \log\left(S_X(f)\frac{\chi_2^2}{2}\right) = \log\left(S_X(f)\right) + \log\left(\frac{\chi_2^2}{2}\right)$$

- Bartlett & Kendall (1946):

$$E\left\{\log\left(\frac{\chi_\eta^2}{\eta}\right)\right\} = \psi\left(\frac{\eta}{2}\right) - \log\left(\frac{\eta}{2}\right) \;\;\&\;\; \mathrm{var}\left\{\log\left(\frac{\chi_\eta^2}{\eta}\right)\right\} = \psi'\left(\frac{\eta}{2}\right)$$

where $\psi(\cdot)$ & $\psi'(\cdot)$ are di– & trigamma functions

- letting $\gamma \doteq 0.57721$ be Euler's constant, yields

$$\begin{aligned} E\{\log\left(\hat{S}_X^{(p)}(f)\right)\} &= \log\left(S_X(f)\right) + \psi(1) - \log\left(1\right) \\ &= \log\left(S_X(f)\right) - \gamma \\ \mathrm{var}\{\log\left(\hat{S}_X^{(p)}(f)\right)\} &= \psi'(1) = \frac{\pi^2}{6} \end{aligned}$$

The Periodogram: IV

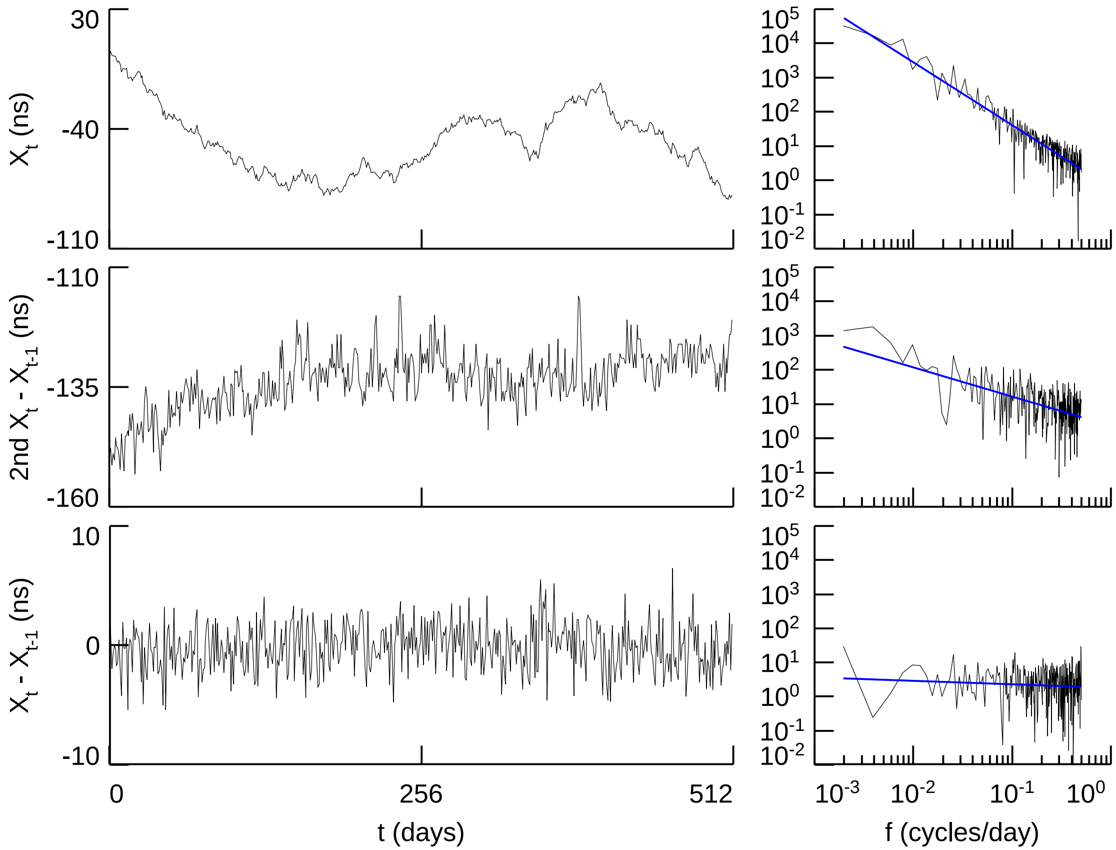
- define $Y^{(p)}(f_j) = \log(\hat{S}_X^{(p)}(f_j)) + \gamma$
- model $Y^{(p)}(f_j)$ over low frequencies indexed by $0 < j < J$ as

$$\begin{aligned} Y^{(p)}(f_j) &\approx \log(S_X(f_j)) + \epsilon(f_j) \\ &\approx \log(\sigma_\epsilon^2) + \alpha \log(2\pi f_j) + \epsilon(f_j) \end{aligned}$$

- error $\epsilon(f_j)$ in linear regression model such that
  - $E\{\epsilon(f_j)\} = 0$ & $\mathrm{var}\{\epsilon(f_j)\} = \frac{\pi^2}{6}$ (known!)
  - can argue that $\epsilon(f_j)$'s approximately pairwise uncorrelated
  - $\epsilon(f_j) \stackrel{\mathrm{d}}{=} \log(\chi_2^2) + \gamma - \log(2)$ markedly non-Gaussian
- least squares procedure yields estimates $\hat{\alpha}$ and $\hat{\sigma}_\epsilon^2$ for α and σ_ϵ^2, along with estimates of variability in $\hat{\alpha}$ and $\hat{\sigma}_\epsilon^2$

Examples of Periodogram-Based Spectral Analysis

- examples of clock noise, periodograms & fitted regression lines



$\hat{\alpha} \doteq -1.83$
95% CI $\doteq [-1.94, -1.72]$

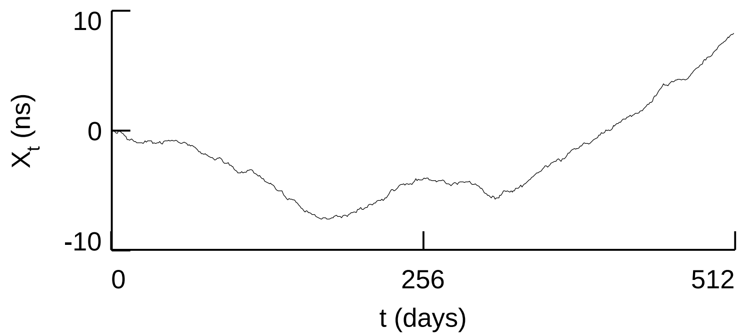
$\hat{\alpha} \doteq -0.85$
95% CI $\doteq [-1.01, -0.69]$

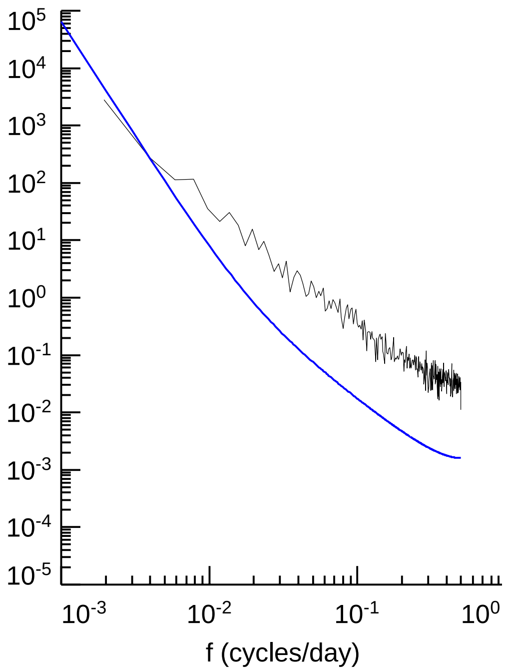
$\hat{\alpha} \doteq -0.11$
95% CI $\doteq [-0.28, 0.06]$

— note: ‘CI’ stands for ‘confidence interval’

Bias in Periodogram due to Leakage

- periodogram can be badly biased for certain processes
- example: periodogram for X_t generated from composite FD process ($\alpha_0 = -4$ and $\alpha_1 = -2$)



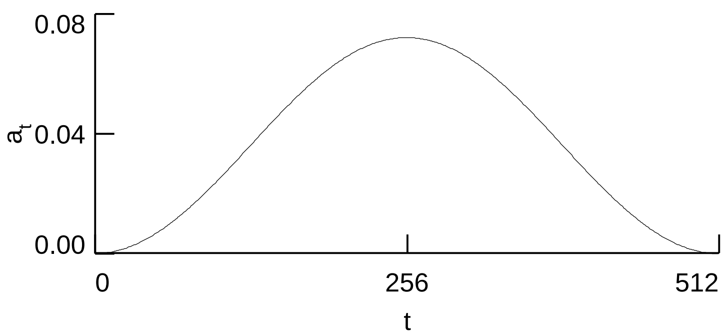


Alleviation of Leakage via Tapering

- tapering is technique for alleviating leakage:

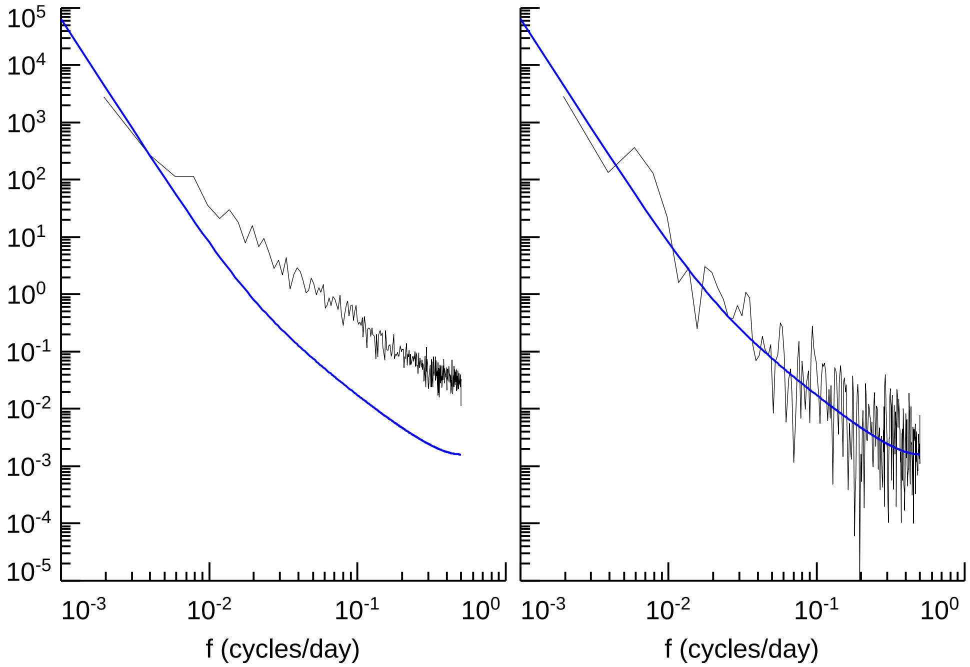
$$\hat{S}_X^{(d)}(f) = \left| \sum_{t=0}^{N-1} a_t (X_t - \hat{\mu}) e^{-i2\pi f t} \right|^2$$

- $\hat{S}_X^{(d)}(\cdot)$ called direct spectral estimator
- a_t called data taper (typically bell-shaped curve)
- example: Hanning data taper



Example of Alleviation of Leakage

- periodogram & direct spectral estimate for composite FD series



— note: used Hanning data taper in forming $\hat{S}_X^{(d)}(\cdot)$

Multitaper Spectral Estimation: I

- critique: tapering loses 'information' at end of series (sample size N effectively shortened)
- Thomson (1982): multitapering recovers 'lost info'
- use set of K orthonormal data tapers $a_{k,t}$:

$$\sum_{t=0}^{N-1} a_{k,t} a_{l,t} = \begin{cases} 1, & \text{if } k = l; \\ 0, & \text{if } k \neq l, \end{cases} \quad 0 \leq n, l \leq K-1$$

- use $a_{k,t}$ to form kth direct spectral estimator:

$$\hat{S}^{(mt)}_{X,k}(f) = \left| \sum_{t=0}^{N-1} a_{k,t}(X_t - \hat{\mu})e^{-i2\pi ft} \right|^2, \quad k = 0, \ldots, K-1$$

Multitaper Spectral Estimation: II

- simplest form of multitaper spectrum estimator:

$$\hat{S}_X^{(mt)}(f) = \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_{X,k}^{(mt)}(f)$$

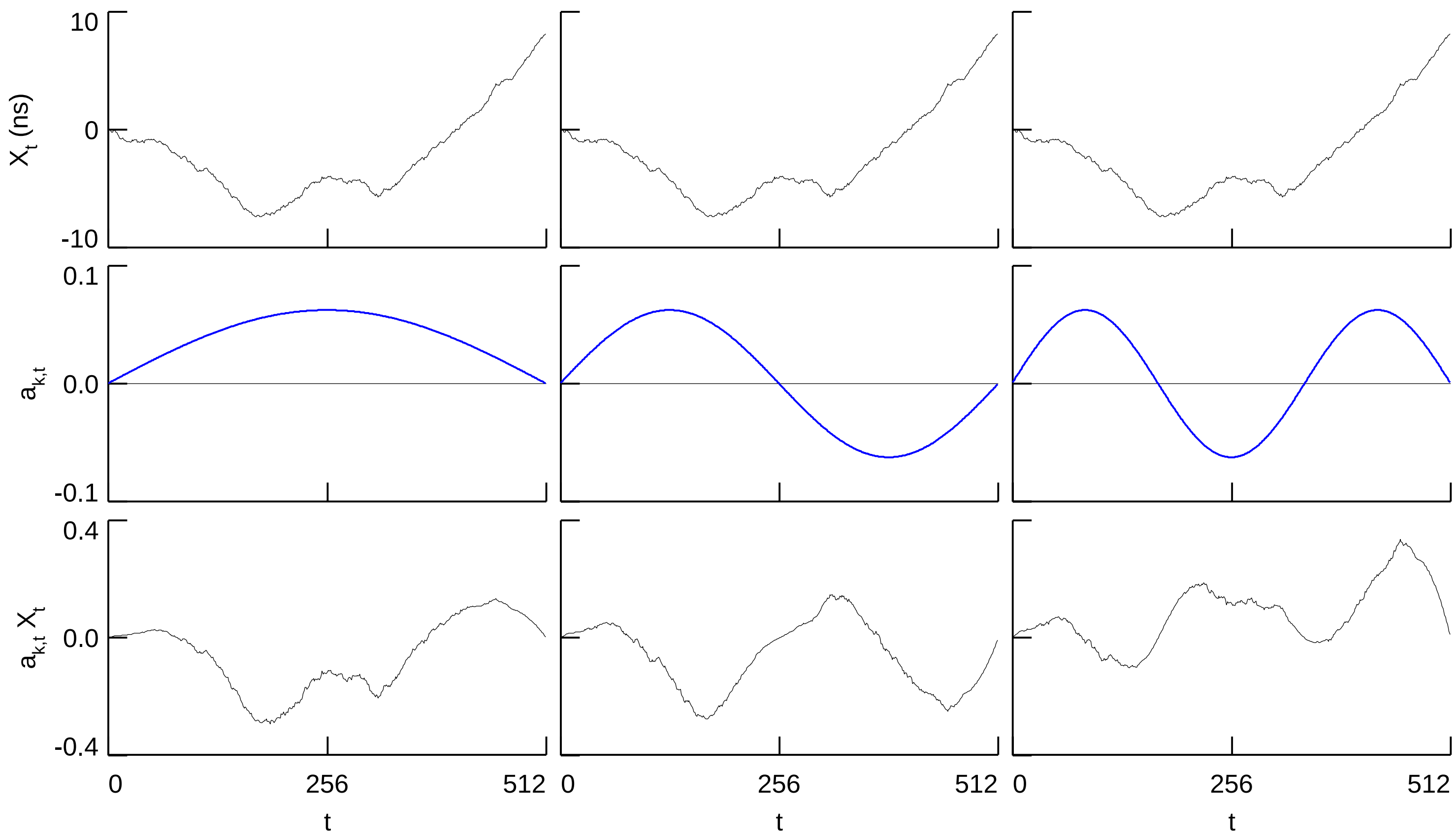
- sinusoidal tapers are one family of multitapers:

$$a_{k,t} = \left(\frac{2}{N+1} \right)^{1/2} \sin \left(\frac{(k+1)\pi(t+1)}{N+1} \right)$$

(Riedel & Sidorenko, 1995)

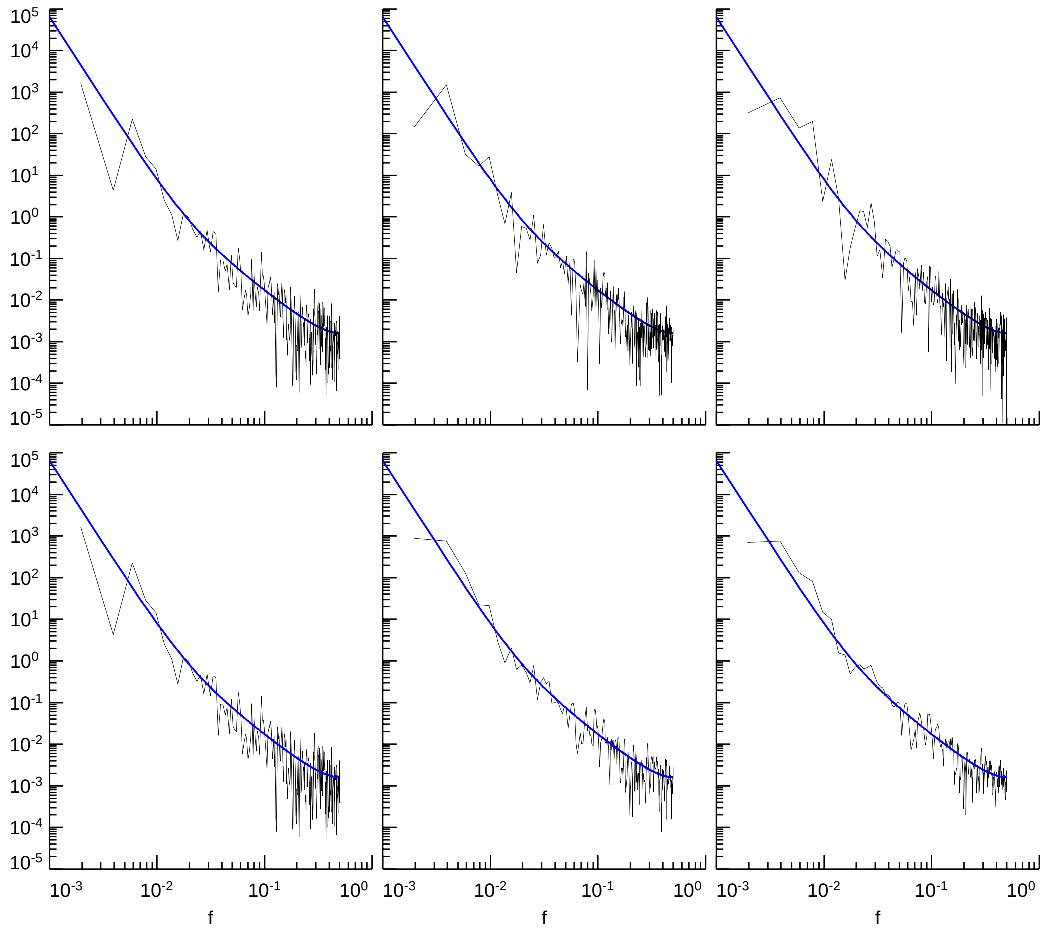
Example of Sinusoidal Tapers & Tapered Series

- X_t (top); $a_{k,t}, k = 0, 1, 2$ (middle); $a_{k,t}X_t$ (bottom)



Example of Multitaper Spectral Estimates

- $\hat{S}^{(mt)}_{X,k}(\cdot), k = 0, 1, 2$ (top); $\hat{S}^{(mt)}_{X}(\cdot)$, $K = 1, 2, 3$ (bottom)



Multitaper Spectral Estimation: III

- if $S_X(\cdot)$ slowly varying around $S_X(f)$ & if N large,
$$\hat{S}_X^{(mt)}(f) \stackrel{\mathrm{d}}{=} \frac{S_X(f)\chi^2_{2K}}{2K}$$
approximately for $0 < f < 1/2$, implying
$$\operatorname{var}\{\hat{S}_X^{(mt)}(f)\} \approx \frac{S^2(f)}{4K^2}\operatorname{var}\{\chi^2_{2K}\} = \frac{S^2(f)}{K}$$
- define $Y^{(mt)}(f_j) = \log(\hat{S}_X^{(mt)}(f_j)) - \psi(K) + \log(K)$
- can model $Y^{(mt)}(f_j)$ as
$$\begin{aligned} Y^{(mt)}(f_j) &\approx \log(S_X(f_j)) + \zeta(f_j) \\ &\approx \log(\sigma^2_\epsilon) + \alpha\log(2\pi f_j) + \zeta(f_j) \end{aligned}$$
over low frequencies indexed by $0 < j < J$

Multitaper Spectral Estimation: IV

- error $\zeta(f_j)$ in linear regression model such that
  - $E\{\zeta(f_j)\} = 0$
  - $\text{var}\,\{\zeta(f_j)\} = \psi'(K)$, a known constant!
  - approximately Gaussian if $K \geq 5$
  - correlated, but with known simple structure
- generalized least squares procedure yields estimates $\hat{\alpha}$ and $\hat{\sigma}_\epsilon^2$ for α and σ_ϵ^2, along with estimates of variability in $\hat{\alpha}$ and $\hat{\sigma}_\epsilon^2$
- multitaper approach superior to periodogram approach

Discrete Wavelet Transform (DWT): I

- let $\mathbf{X} = [X_0, X_1, \ldots, X_{N-1}]^T$ be observed time series (for convenience, assume N integer multiple of 2^{J_0})
- let $\mathcal{W}$ be $N \times N$ orthonormal DWT matrix; i.e., inverse of $\mathcal{W}$ is just its transpose $\mathcal{W}^T$
- $\mathbf{W} = \mathcal{W}\mathbf{X}$ is vector of DWT coefficients
- orthonormality says $\mathbf{X} = \mathcal{W}^T\mathbf{W}$
- implies $\mathbf{X}$ & $\mathbf{W}$ are equivalent (no loss of information in $\mathbf{W}$)

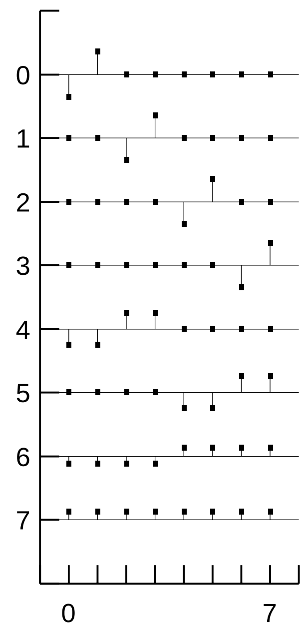
Discrete Wavelet Transform (DWT): II

- can partition $\mathbf{W}$ as follows:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_{J_0} \\ \mathbf{V}_{J_0} \end{bmatrix}$$

- $\mathbf{W}_j$ contains $N_j = N/2^j$ wavelet coefficients
  - related to changes of averages at 'dyadic' scale $\tau_j = 2^{j-1}$
  - related to times spaced 2^j units apart
- $\mathbf{V}_{J_0}$ contains $N_{J_0} = N/2^{J_0}$ scaling coefficients
  - related to averages at scale $\lambda_{J_0} = 2^{J_0}$
  - related to times spaced 2^{J_0} units apart

Example: $\mathcal{W}$ for Haar DWT with $N = 8$



- rows 0, 1, 2 & 3 yield $\mathbf{W}_1 \propto$ *changes* on scale 1
- next 2 rows yield $\mathbf{W}_2 \propto$ *changes* on scale 2
- row 6 yields $\mathbf{W}_3 \propto$ *change* on scale 4
- row 7 yields $\mathbf{V}_3 \propto$ *average* on scale 8

DWT in Terms of Filters

- filter $X_0, X_1, \ldots, X_{N-1}$ to obtain

$$2^{j/2}\widetilde{W}_{j,t} = \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l \bmod N}, \quad t = 0, 1, \ldots, N-1;$$

  $h_{j,l}$ is jth level wavelet filter (note: circular filtering)

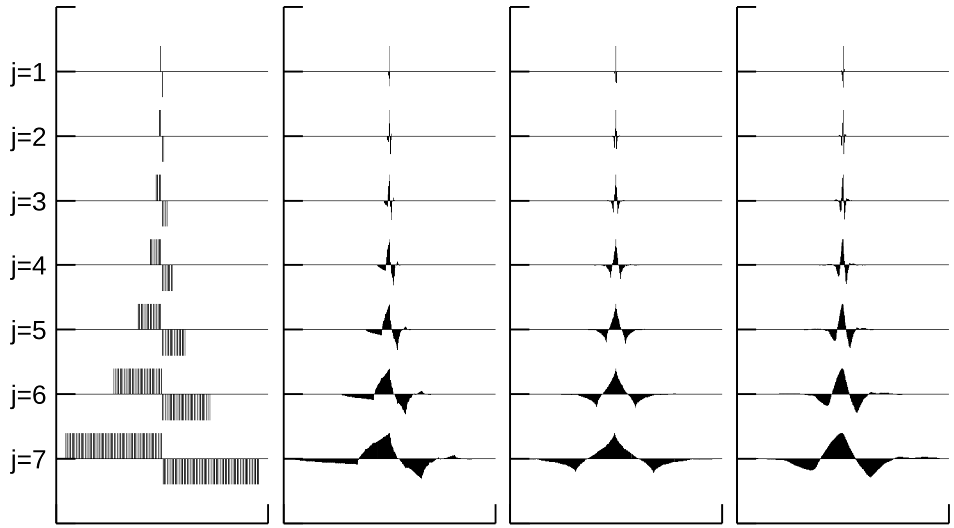
- subsample to obtain wavelet coefficients:

$$W_{j,t} = 2^{j/2}\widetilde{W}_{j,2^j(t+1)-1}, \quad t = 0, 1, \ldots, N_j - 1,$$

  where $W_{j,t}$ is tth element of $\mathbf{W}_j$

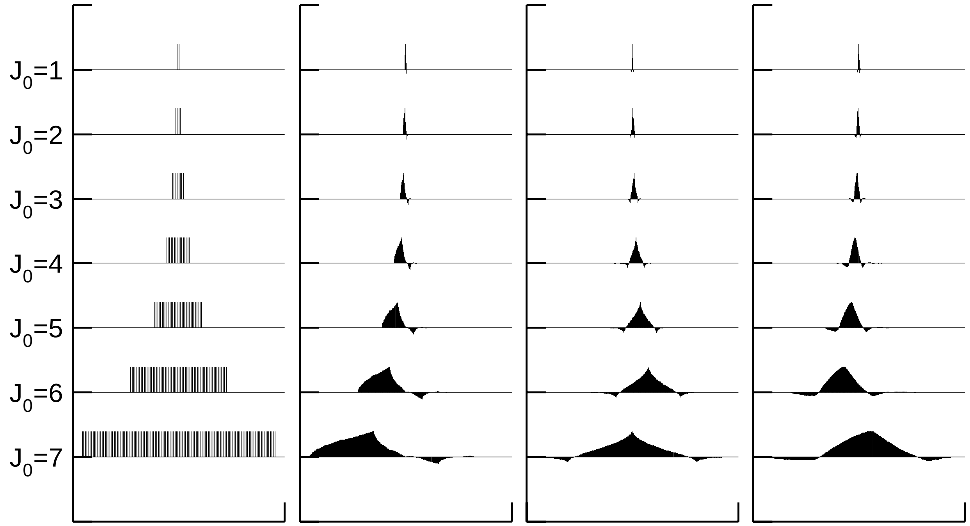
- jth wavelet filter is band-pass with pass-band $[\frac{1}{2^{j+1}}, \frac{1}{2^j}]$ (i.e., scale related to *interval* of frequencies)

Four Examples of Wavelet Filters



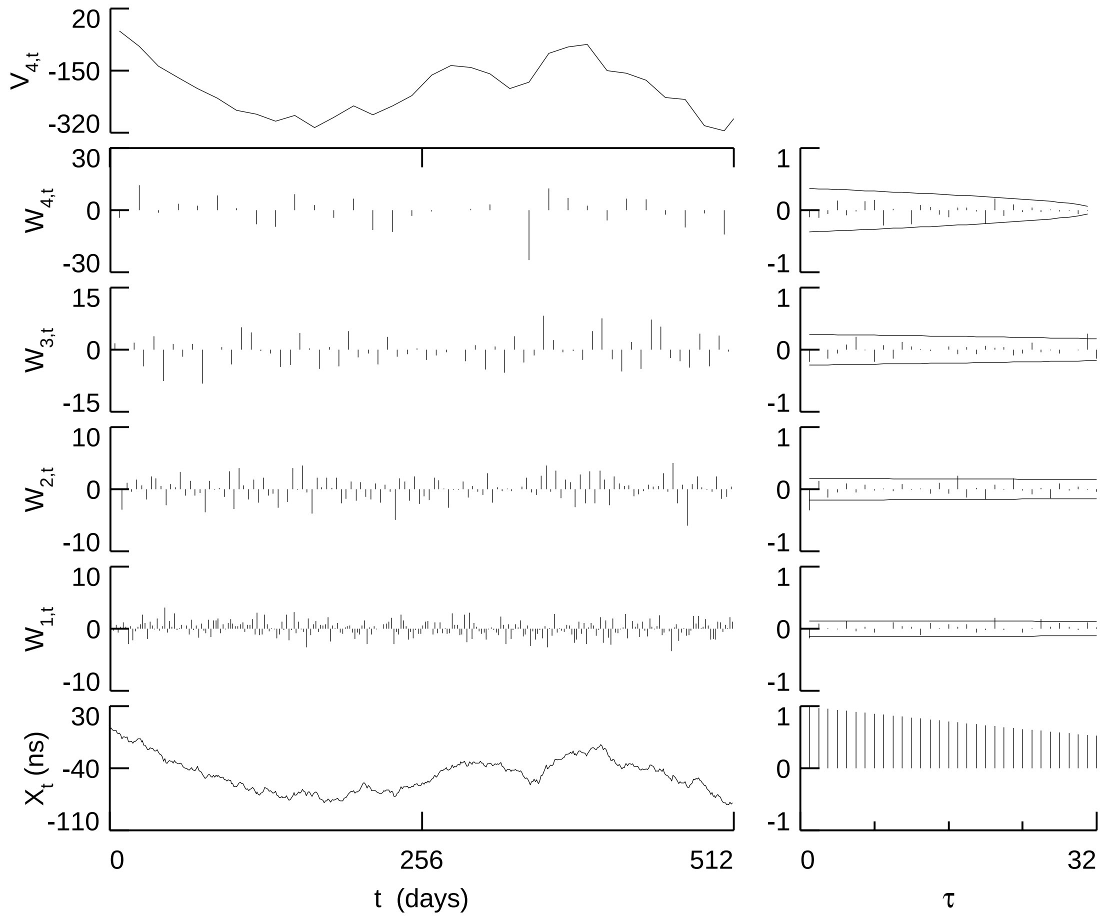
- from left to right, these plots show
  - Haar wavelet filters, for which $L_1 = 2$
  - D(4) filters, i.e., Daubechies' 'extremal phase' with $L_1 = 4$
  - C(6) filters, i.e., Daubechies' 'coiflet' with $L_1 = 6$
  - LA(8) filters, i.e., Daubechies' 'least asymmetic' with $L_1 = 8$

Four Examples of Scaling Filters



- above are scaling filters corresponding to wavelet filters
- scaling filters yield $\mathbf{V}_{J_0}$
- J_0th scaling filter is low-pass with pass-band $[0, \frac{1}{2^{J_0+1}}]$
- as width L_1 of 1st level filter increases, band-pass & low-pass approximations improve

Example: D(4) DWT Coefficients for Clock 55 X_t



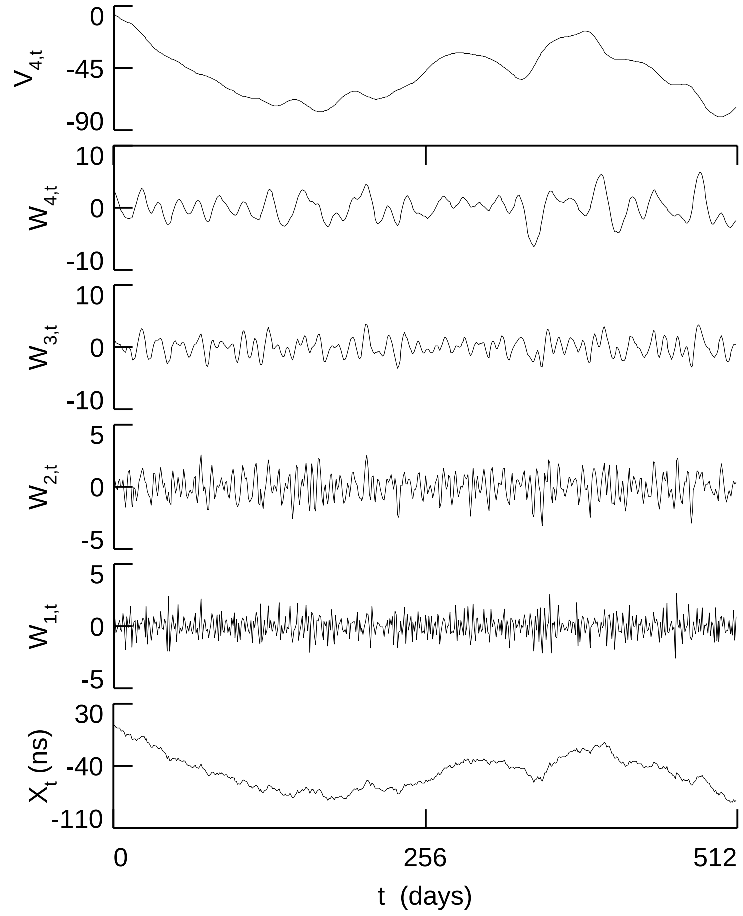
Variation: Maximal Overlap DWT (MODWT)

- can eliminate downsampling and use

$$\widetilde{W}_{j,t} = \frac{1}{2^{j/2}} \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l \bmod N}, \quad t = 0, 1, \ldots, N-1,$$

  to define MODWT coefficients $\widetilde{\mathbf{W}}_j$ (& also $\widetilde{\mathbf{V}}_{J_0}$)
- unlike DWT, MODWT is not orthonormal (in fact MODWT is highly redundant)
- unlike DWT, MODWT works for all samples sizes N (i.e., power of 2 assumption is not required)

Example: D(4) MODWT Coefficients for Clock 55



- can use to track, e.g., time-varying FD process

Wavelet-Based Analysis of Variance: I

- consider 'energy' in time series: $\|\mathbf{X}\|^2 = \mathbf{X}^T\mathbf{X} = \sum_{t=0}^{N-1} X_t^2$
- energy preserved in DWT coefficients:

$$\|\mathbf{W}\|^2 = \|\mathcal{W}\mathbf{X}\|^2 = \mathbf{X}^T\mathcal{W}^T\mathcal{W}\mathbf{X} = \mathbf{X}^T\mathbf{X} = \|\mathbf{X}\|^2$$

- since $\mathbf{W}_1, \ldots, \mathbf{W}_{J_0}, \mathbf{V}_{J_0}$ partitions $\mathbf{W}$, have

$$\|\mathbf{W}\|^2 = \sum_{j=1}^{J_0} \|\mathbf{W}_j\|^2 + \|\mathbf{V}_{J_0}\|^2,$$

  leading to analysis of sample variance:

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{t=0}^{N-1}(X_t - \hat{\mu})^2 = \frac{1}{N}\Big(\sum_{j=1}^{J_0}\|\mathbf{W}_j\|^2 + \|\mathbf{V}_{J_0}\|^2\Big) - \hat{\mu}^2$$

Wavelet-Based Analysis of Variance: II

- energy also preserved in MODWT coefficients:

$$\|\mathbf{X}\|^2 = \sum_{j=1}^{J_0} \|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_{J_0}\|^2,$$

  leading to an analogous analysis of sample variance:

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{t=0}^{N-1} (X_t - \hat{\mu})^2 = \frac{1}{N} \Big(\sum_{j=1}^{J_0} \|\widetilde{\mathbf{W}}_j\|^2 + \|\widetilde{\mathbf{V}}_{J_0}\|^2 \Big) - \hat{\mu}^2$$

- scale-based decomposition (spectrum is frequency-based)

Wavelet Variance Analysis: I

- for FD and related processes X_t, can define wavelet variance
  - run X_t through jth level wavelet filter:

$$\overline{W}_{j,t} = \frac{1}{2^{j/2}} \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l}, \quad t = \ldots, -1, 0, 1, \ldots$$

  - wavelet variance is variance of filter output:

$$\nu_X^2(\tau_j) = \mathrm{var}\,\{\overline{W}_{j,t}\}$$

  - does not depend on t and has same units as X_t^2
  - also called wavelet spectrum
- wavelet variance decomposes σ^2 across scales τ_j:

$$\mathrm{var}\,\{X_t\} = \sum_{j=1}^{\infty} \nu_X^2(\tau_j)$$

Wavelet Variance Analysis: II

- because $h_{j,l} \approx$ bandpass over $[1/2^{j+1}, 1/2^j]$,

$$\nu_X^2(\tau_j) \approx 2 \int_{1/2^{j+1}}^{1/2^j} S_X(f)\, df \qquad (*)$$

- if $S_X(\cdot)$ 'featureless', info in $\nu_X^2(\tau_j)$ equivalent to info in $S_X(\cdot)$
- $\nu_X^2(\tau_j)$ more succinct: only one value per octave band
- recall spectrum for FD process:

$$S_X(f) = \sigma_\epsilon^2 \, |2 \sin(\pi f)|^\alpha \approx \sigma_\epsilon^2 \, |2\pi f|^\alpha$$

- $(*)$ implies $\nu_X^2(\tau_j) \propto \tau_j^{-\alpha-1}$ approximately
- can deduce α from slope of $\log(\nu_X^2(\tau_j))$ vs. $\log(\tau_j)$

Estimation of Wavelet Variance: I

- can base estimator on MODWT of $X_0, X_1, \ldots, X_{N-1}$
- if we compare

$$\widetilde{W}_{j,t} = \frac{1}{2^{j/2}} \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l \bmod N}, \quad t = 0, \ldots, N-1,$$

with

$$\overline{W}_{j,t} = \frac{1}{2^{j/2}} \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l}, \quad t = \ldots, -1, 0, 1, \ldots$$

find $\widetilde{W}_{j,t} = \overline{W}_{j,t}$ if 'mod' not needed: $L_j - 1 \leq t < N$

Estimation of Wavelet Variance: II

- if $N - L_j \geq 0$, unbiased estimator of $\nu_X^2(\tau_j)$ is

$$\hat{\nu}_X^2(\tau_j) = \frac{1}{N - L_j + 1} \sum_{t=L_j-1}^{N-1} \widetilde{W}_{j,t}^2 = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \overline{W}_{j,t}^2,$$

where $M_j = N - L_j + 1$

- can also construct biased estimator of $\nu_X^2(\tau_j)$:

$$\tilde{\nu}_X^2(\tau_j) = \frac{1}{N} \sum_{t=0}^{N-1} \widetilde{W}_{j,t}^2 = \frac{1}{N} \left(\sum_{t=0}^{L_j-2} \widetilde{W}_{j,t}^2 + \sum_{t=L_j-1}^{N-1} \overline{W}_{j,t}^2 \right)$$

first sum in parentheses influenced by circularity

Estimation of Wavelet Variance: III

- biased estimator unbiased if X_t white noise
- biased estimator offers exact analysis of $\hat{\sigma}^2$; unbiased estimator need not
- biased estimator can have better mean square error (Greenhall *et al.*, 1999; need to 'reflect' X_t)

Statistical Properties of $\hat{\nu}^2_X(\tau_j)$

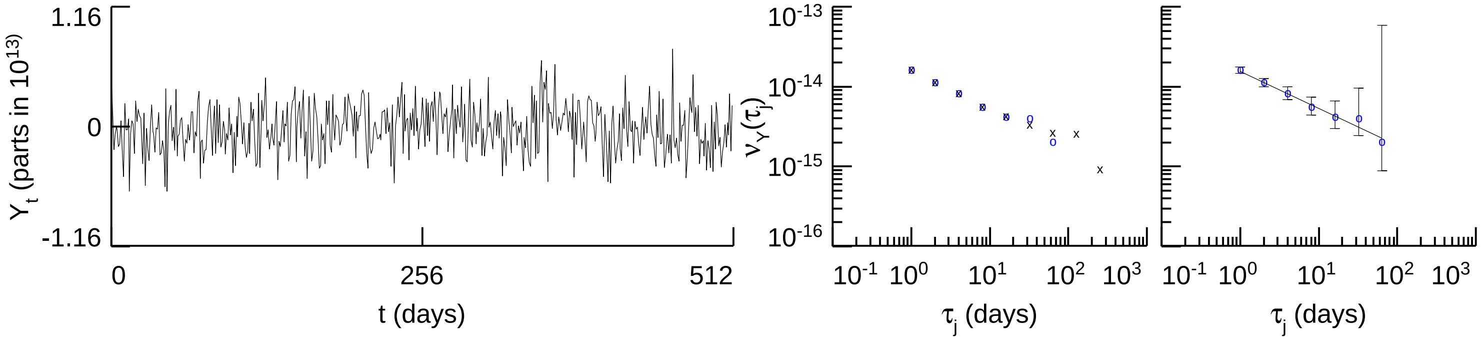
- suppose $\{\overline{W}_{j,t}\}$ Gaussian with mean 0 & spectrum $S_j(f)$
- suppose square integrability condition holds:

$$A_j = \int_{-1/2}^{1/2} S_j^2(f)\, df < \infty \;\; \& \;\; S_j(f) > 0$$

  (holds for FD process if wavelet filter width L_1 large enough)
- can show $\hat{\nu}^2_X(\tau_j)$ asymptotically normal with mean $\nu^2_X(\tau_j)$ & large sample variance $2A_j/M_j$
- can estimate A_j and use with $\hat{\nu}^2_X(\tau_j)$ to construct confidence interval for $\nu^2_X(\tau_j)$

Example: Wavelet Variance Analysis of Clock 55

- use one day average fractional frequency deviates $\overline{Y}_t \propto X_t^{(1)}$



- x's: $\hat{\nu}_{\overline{Y}}(\tau_j)$ using Haar wavelet; related to square root of Allan variance $\sigma^2_{\overline{Y}}(2, \tau_j)$ since

$$\nu^2_{\overline{Y}}(\tau_j) = \tfrac{1}{2}\sigma^2_{\overline{Y}}(2, \tau_j)$$

- o's: $\hat{\nu}_{\overline{Y}}(\tau_j)$ using D(4) wavelet, along with 95% CIs & weighted linear least squares fit of $\log_{10}(\hat{\nu}_{\overline{Y}}(\tau_j))$ versus $\log_{10}(\tau_j)$
- yields $\hat{\alpha} \doteq -0.06$ (very close to white noise $\alpha = 0$)

Summary

- fractionally differenced processes
  - provide statistically tractable models for clock noise
  - extensible to composite, ARFIMA & time-varying processes
- spectral and wavelet analysis can provide
  - estimates of parameters of FD processes
  - decomposition of sample variance across
    - frequencies (in case of spectral analysis)
    - scales (in case of wavelet analysis)
- wavelet variance has some advantages for clock noise
  - has same units as X_t^2 & estimates α & σ_ϵ^2 somewhat better
  - useful with time-varying noise process & polynomial trends

References

1. Bartlett, M. S., Kendall, D. G., *Supplement to the Journal of the Royal Statistical Society*, 1946, **8**, 128—138.

2. Granger, C. W. J., Joyeux, R., *Journal of Time Series Analysis*, 1980, **1**, 15–29.

3. Greenhall, C. A., Howe, D. A., Percival, D. B., *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 1999, **46**, 1183–1191.

4. Percival, D. B., *Metrologia*, 2003, **40**, S289–S304.

5. Percival, D. B., Walden, A. T., *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge, UK, Cambridge University Press, 1993, 583 p.

6. Percival, D. B., Walden, A. T., *Wavelet Methods for Time Series Analysis*, Cambridge, UK, Cambridge University Press, 2000, 594 p.

7. Riedel, K. S., Sidorenko, A., *IEEE Transactions on Signal Processing*, 1995, **43**, 188–195.

8. Thomson, D. J., *Proceedings of the IEEE*, 1982, **70**, 1055–1096.