Wavelet-Based Surrogates for Testing Time Series

Don Percival

Applied Physics Lab, University of Washington

Department of Statistics, University of Washington

Data Analysis Products Division, MathSoft

overheads for talk available via

`http://staff.washington.edu/dbp/talks.html`

joint work with S. Sardy & A. C. Davison, EPFL (Lausanne, Switzerland)

Overview of Talk

- background on surrogate data/bootstrapping
  (current methodology not ideal)
- will look at wavelet-based methodology
  - basics of discrete wavelet transform (DWT)
  - DWT as a time series decorrelator
- works for 'fractal' time series; can fail on others
- generalization: discrete wavelet packet transforms
  - adaptive selection of decorrelating transform
    ('top down' via white noise tests)
- results for sample autocorrelation

Background: I

- Theiler *et al.* (1992): method of surrogate data (useful for identifying nonlinear time series)
  - let $\mathbf{X}$ be time series of interest
  - assume null hypothesis H_0: linear & stationary
  - given $\mathbf{X}$, generate surrogates (simulated series with properties dictated by H_0)
  - compute test statistic for $\mathbf{X}$ and its surrogates
  - reject H_0 if test for $\mathbf{X}$ in tails of histogram
- 'surrogate data' basically same as 'bootstrapping'
  - originally devised for IID data $\mathbf{Z}$ (independent and identically distributed)
  - let $T(\mathbf{Z})$ be statistic with unknown distribution
  - create $\mathbf{Z}_i$ by sampling with replacement from $\mathbf{Z}$
  - use histogram of $T(\mathbf{Z}_0), \ldots, T(\mathbf{Z}_{M-1})$ as stand-in for distribution of $T(\mathbf{Z})$
- time series problematic because of correlation

Background: II

- to generate surrogates, Theiler *et al.* proposed:
  1. compute discrete Fourier transform (DFT)

     $$\widetilde{X}_k \equiv \sum_{t=0}^{N-1} X_t e^{-i2\pi kt/N};$$

     randomize phases; compute inverse DFT
     (critique: fails if T depends on $|\widetilde{X}_k|$)
  2. do 1 with windowed series $h_t X_t$
     (critique: spurious low frequency terms)
  3. rescale $\mathbf{X}$ to be Gaussian; do 1; undo rescaling
     (critique: fails if T depends on histogram)
- Davison & Hinkley, 1998, Chapter 8, discusses bootstrapping in context of time series analysis
  1. fit, e.g., $X_t = \phi X_{t-1} + \epsilon_t$ & bootstrap from model
     (critique: depends critically on choice of model)
  2. block bootstrap
     (critique: fails for 'fractal' time series)
- more work needed on surrogate generation

Overview of DWT

- 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
- $\mathbf{W} = \mathcal{W}\mathbf{X}$ is vector of DWT coefficients
- orthonormality says $\mathbf{X} = \mathcal{W}^T\mathbf{W}$, so $\mathbf{X} \Leftrightarrow \mathbf{W}$
- 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 scale $\tau_j = 2^{j-1}$ (τ_j is jth 'dyadic' scale)
  - 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

DWT in Terms of Filters

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

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

  where $h_{j,l}$ 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 $\mathbf{W}_j$

- Figs. 1 & 2: Haar, D(4), C(6) & LA(8) wavelet filters
- jth wavelet filter is band-pass with pass-band $[\frac{1}{2^{j+1}}, \frac{1}{2^j}]$
- note: jth scale related to interval of frequencies
- similarly, scaling filters yield $\mathbf{V}_{J_0}$
- Figs. 3 & 4: Haar, D(4), C(6) & LA(8) scaling filters
- J_0th scaling filter is low-pass with pass-band $[0, \frac{1}{2^{J_0+1}}]$

Example: DWT of FD Process

- X_t called fractional difference (FD) process if it has a spectral density function (SDF) given by

$$S_X(f) = \frac{\sigma^2}{|2\sin(\pi f)|^{2\delta}},$$

  where $\sigma^2 > 0$ and $-\frac{1}{2} \le \delta < \frac{1}{2}$

- note: for small f, have $S_X(f) \approx C/|f|^{2\delta}$; i.e., '$1/f$ type process' or 'stochastic fractal'
- if $\delta = 0$, FD process is white noise
- if $0 < \delta < \frac{1}{2}$, stationary with 'long memory'
- can extend definition to $\delta \ge \frac{1}{2}$
  - nonstationary $1/f$ type process
  - also called FARIMA(0,δ,0) process
- Fig. 5: LA(8) DWT of FD process with $\delta = 0.4$

Wavelets as Whitening Filters

- since FD process is stationary, $\mathbf{W}_j$ is also (ignoring terms influenced by circularity)
- Fig. 6: SDFs $S_j(\cdot)$ for each $\mathbf{W}_j$
- DWT acts as whitening filter for FD series because SDFs for $\mathbf{W}_j$ are $\approx$ flat over pass-bands $[\frac{1}{2^{j+1}}, \frac{1}{2^j}]$
- Figs. 7 & 8: auto- and cross-correlations
- can regard $\mathbf{W}_1, \ldots, \mathbf{W}_{J_0}$ as $\approx$ uncorrelated (but $N/2^{J_0}$ scaling coefficients are NOT)
- since $\mathbf{X} \Leftrightarrow \mathbf{W}$, can write $T(\mathbf{X}) = T(\mathbf{W})$
- if Gaussian, close to independently distributed
  - $\approx$ IID within given $\mathbf{W}_j$, but not between ($\mathbf{W}_j$ & $\mathbf{W}_{j'}$ can have different variances)
  - bootstrap OK with this departure from IID

DWT-Based Bootstrapping: I

- simple example: lag 1 autocorrelation estimate

$$T(\mathbf{X}) = \hat{r}_1 \equiv \frac{\Sigma_{t=0}^{N-2}(X_t - \overline{X})(X_{t+1} - \overline{X})}{\Sigma_{t=0}^{N-1}(X_t - \overline{X})^2}$$

  for FD process

- very simple – but challenging – statistic:
  - $\hat{r}_1$ is function of $|\widetilde{X}_k|^2$ (hence DFT + phase randomization fails)
  - block bootstrap etc tend to fail for FD processes
- to get standard error of $\hat{r}_1$,
  - compute DWT $\mathbf{W}$ of $\mathbf{X}$
  - sample with replacement from $\mathbf{W}_j$ to form $\mathbf{W}_j^{(b)}$ (do same with $\mathbf{V}_{J_0}$)
  - synthesize $\mathbf{X}^{(b)}$ using $\mathbf{W}_j^{(b)}$'s & $\mathbf{V}_{J_0}^{(b)}$ (see Fig. 9)
  - compute $\hat{r}_1^{(b)}$ for $\mathbf{X}^{(b)}$
  - repeat until computer gets tired
  - use standard error of $\hat{r}_1^{(b)}$'s for $\mathbf{X}^{(b)}$'s to assess standard error of $\hat{r}_1$ for $\mathbf{X}$

DWT-Based Bootstrapping: II

- Monte Carlo study
  - LA(8) DWT
  - FD process with $\delta = 0.45$
  - 1024 simulated FD series
  - 64 DWT-based bootstraps from each series
  - FDP with $d = 0.45$
- comparison of standard errors

N	block	DWT	true
128	0.077	0.099	0.107
1024	0.034	0.047	0.053

WaveStrapping: I

- key to scheme is decorrelation property
  - DWT well suited for FD processes,
    but can fail for other processes
  - Fig. 10 considers $X_t = \phi X_{t-1} + \epsilon_t$:
    DWT OK for $\phi = 0.9$, but not $\phi = -0.9$
- need different partitioning of $[0, \frac{1}{2}]$
- Fig. 11: discrete wavelet packet transform (DWPT)
- jth DWPT consists of $\mathbf{W}_{j,n}$, $n = 0, \ldots, 2^j - 1$
- $\{\mathbf{W}_{j,n}\}$, $j = 0, \ldots, J_0$, form DWP table
- can extract many transforms from DWP table
  - Fig. 12: DWT is special case
  - Fig. 13: another special case
- idea: adaptively select transform from DWP table

WaveStrapping: II

- given $\mathbf{W}_{j,n}$, test for white noise using, e.g.,
  - portmanteau test on sample autocorrelations
  - cumulative periodogram test
- if test fails to reject, keep $\mathbf{W}_{j,n}$
- if test rejects, split $\mathbf{W}_{n,j}$ into $\mathbf{W}_{j+1,2n}$ & $\mathbf{W}_{j+1,2n+1}$, and test both for white noise
- yields 'top down' selection from DWT table
- once transform selected, bootstrap as in DWT case
- adaptive procedure called 'WaveStrapping'

WaveStrapping: III

- Monte Carlo study: as before, but now also with
  - white noise (WN) process ϵ_t
  - autoregressive (AR) process: $X_t = 0.9X_{t-1} + \epsilon_t$
  - moving average (MA) process: $X_t = \epsilon_t + \epsilon_{t-1}$
- comparison of standard errors

				WaveStrap		
process	N	block	DWT	port	pgrm	true
WN	128	0.081	0.083	0.086	0.087	0.087
	1024	0.030	0.032	0.032	0.031	0.031
AR(1)	128	0.054	0.055	0.051	0.054	0.048
	1024	0.015	0.016	0.015	0.015	0.014
MA(1)	128	0.065	0.070	0.068	0.066	0.063
	1024	0.022	0.026	0.024	0.024	0.022
FD	128	0.077	0.099	0.088	0.096	0.107
	1024	0.034	0.047	0.045	0.047	0.053

Future Research

- ‘bottom up’ selection with ‘best basis’ algorithm (algorithm tends to pick decorrelating transforms)
- combine with parametric bootstrap: model within scale correlation as AR(1)
- combine with tests for nonlinearity
- test on physiological time series