CNR-IMATI Milano

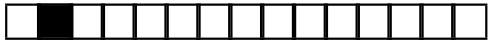
December 14-16, 2004

WAVELETS AND SELF-SIMILARITY: THEORY AND APPLICATIONS

Lecture 5: Estimation of Hurst Exponent

P L A N

1. **Scaling Slope versus H**
2. **Some Methods of estimating H**

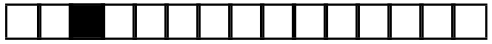


Detection and Assessment of Scaling

■ By observing the data at time domain it is impossible to detect scaling without resorting to such tools such as, for example, Structure Functions, Spectrograms, Logscale Diagrams, Structure Functions, q-th order Logscale Diagrams, etc.

■ If, for example, in Fourier Log-Spectrograms or Logscale Diagrams it is possible to align a straight line with particular slope of $-\alpha$, over duration of several decades (octaves, "binary-decades") then the scaling in the data is present.

■ The key tool is of course the linear regression analysis for which the goodness of fit is assessed by standard regression measure, such as R^2. Plethora of tests are available here including nonparametric goodness of fit tests (χ^2, Kolmogorov-Smirnov), bootstrap, empirical MSE, etc.

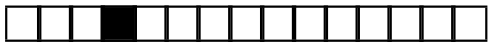


■ It is possible also to automatically select the range of scales $[j_1, j_2]$ for which the goodness of fit measure is minimized. The selection of this range is important since in many situations estimation of the slope is non-robust to change in j_1 and/or j_2. This robustness is influenced by several factors, including quality of data, a slope close to zero, presence of a periodicity, or injection of energy at a particular scale.

■ Especially critical is the selection of low scale j_1. The high variability of spectra at low scales is influenced by several factors – some of which have nothing to do with the nature of data. For example, in the Logscale diagrams points at low scales are obtained by averaging substantially less empirical values of energy (squared wavelet coefficients). The difference in the number of averaged values is huge, for example if the scale $j_2 = 10$ averages 1024 energies, the scale $j_1 = 3$ averages only 8 values.

■ By the assessment of scaling we consider two tasks: identification of the phenomenon (model selection) and evaluation of the scaling slope.

■ As in any model selection task, it is important to have some prior information about the data – because that may be decisive in the model selection and subsequent inference.

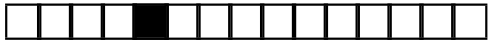


■ If the observed scaling α is in $(0, 1)$ [slope of $-\alpha$], and we believe (or confirm) that data are stationary – then the estimated slope α corresponds to the LRD parameter α and the link with Hurst exponent is $H = \frac{1-\alpha}{2}$.

■ For example, in Gait Data, the stride intervals for a normal person are scaling down with the slope of -0.74. from which we find $H = 0.87$. Stationarity is a reasonable assumption and selected model is LRD. Additional examples are Nile River data and Internet traces data.

■ If, on the other hand, the estimated slopes are greater than one, and the data may have been obtained as cumulative process and/or are not stationary (stock market prices, turbulence measurements, seizure data) then they correspond to an stationary increments or even stationary k-th increments, $k > 1$.

■ In this cases $H = \frac{\alpha-1}{2}$ and in case of a monofractal, H coincides with the regularity (in Hölder sense) of the non-differentiable sample paths. For instance, the turbulence signals give slope of $-5/3$, defining $H = 1/3$.



Estimation of the Scaling Law

■ We discuss some methods for estimating the scaling exponent α or equivalently corresponding Hurst exponent, H.

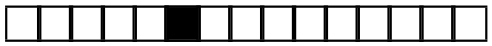
■ The methods fall into two general categories: time domain and scale/frequency domain methods. We already hinted and even discussed some of those, such as R/S and Logscale methods. More comprehensive discussion is provided in this chapter.

■ **R/S Theory**

The rescaled range was analyzed by E. Hurst in 1951. Let $X_1, \ldots, X_n$ be a sequence of random variables with partial sums $Q_t = \sum_{j=1}^{t} X_j,\ t = 1, \ldots, n$, and $Q_0 = 0$. Let $Q_{t,n} = Q_t - \frac{t}{n} Q_n,\ t = 0, 1, \ldots, n$ be the adjusted partial sums. The adjusted range is defined as

$$R_n = \max_{0 \le t \le n} Q_{t,n} - \min_{0 \le t \le n} Q_{t,n}.$$

Let $S_n = 1/n \sum_{j=1}^{n} (X_j - \bar{X})^2$, where $\bar{X} = 1/n \sum_{j=1}^{n} X_j$.



■ Theorem: (Beran monograph) (i) If X_t is a process such that $t^{-1/2}\sum_{s=1}^{[tr]} X_s,\ r \in (0,\infty)$, converges weakly to Brownian motion process $B(t)$, as $t \to \infty$ and if X_t^2 is an ergodic process, then

$$\frac{1}{\sqrt{n}} R_n/S_n \xrightarrow{d} \xi,$$

where ξ is a nondegenerate random variable.

(ii) If X_t is a process such that $t^{-H}\sum_{s=1}^{[tr]} X_s,\ r \in (0,\infty)$, converges weakly to a fractional Brownian motion process $B_H(t)$, as $t \to \infty$, and if X_t^2 is an ergodic process, then

$$n^{-H} R_n/S_n \xrightarrow{d} \eta,$$

where η is a nondegenerate random variable.

■ The assumption in (i) holds for most common short memory random processes. That means that the points $(\log n, \log R_n/S_n)$ would scatter around the line with the slope $H = 1/2$. In the second case, the points would scatter around the line with the slope $H > 1/2$.

■ Variance Plots

■ Second order properties of an H-ss process can be explored by variance plots. For example, for a fBm process $\{B_H(t), t \geq 0\}$, i.e., for
$Y(n) = B_H(n+1) - B_H(n)$,
$Var(B_H(t+h) - B_H(t)) = |h^{2H} Var(B_H(u+1) - B_H(u)) = |h|^{2H}\sigma_Y^2$ i.e., for any fixed k, $Var\left[\sum_{k+1}^{k+m} Y_H(n)\right] = m^{2H}\sigma_X^2$.

■ The aggregated process

$$X_H^{(m)}(n) = \frac{1}{m} \sum_{i=(n-1)m+1}^{nm} Y_H(i),$$

would have variance $1/m^2 \cdot m^{2H}\sigma_X^2 = m^{2H-2}\sigma_X^2$.

■ Thus, the Log-Log plot of sample variance $s^2(Y_H^{(m)})$ [For a vector $y = (y_1, \ldots, y_n)$ of length n, $s^2(y) = 1/(n-1)\sum_{i=1}^n (y_i - \bar{y})^2$.] should have a slope of $2H - 2$.

Allan Variance

Allan variance is mainly used in inference involving $1/f$-type processes, especially in statistical signal processing. This measure of variability converges for a wide range of noise processes, it has straightforward relationship to power law spectral density, and it is easy to compute.

For observations $X_1, \ldots, X_n$ the Allan variance is given by

$$\sigma_a^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (X_{i+1} - X_i)^2.$$

The division by 2 calibrates this variance to coincide with the standard variance estimator s^2 when observations are coming from a white noise process.

The advantage of the Allan variance over the standard variance estimator is that it converges for most of the commonly encountered kinds of noise, whereas the classical variance does not always converge to a finite value. For example, for the "flicker" noise the standard variance estimator s^2 does not converge.

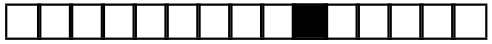
■ Related measure is the Hadamard Variance,

$$\sigma_h^2 = \frac{1}{6(n-2)} \sum_{i=1}^{n-2} (X_{i+2} - 2 \cdot X_{i-1} + X_i)^2,$$

also known as three-point variance. It uses squares of second differences in data.

■ Main use of Allan variance is conventional: to provide a measure of variation in data. In the frequency analysis of time series, measured frequency may be sampled at some rate. The resulting Allan deviation over the sample values represent a general measure of frequency stability at the given sampling rate.

■ The more interesting role of Allan variance is in the discrimination of a residual noise. In frequency measurement work, five different types of noise are encountered: white noise phase modulation, flicker noise phase modulation, white noise frequency modulation, flicker noise frequency modulation, and random walk frequency modulation. A log-log plot of Allan variance versus sampling period produces approximate straight lines with different slopes in four of the five possible cases.



■ An application of Allan variance in a testing problem. Let $X_1, \ldots, X_n$ be a random sample. It is often of interest to test the hypothesis of constant mean, i.e., $H_0 : EX_i = C,\ i = 1, \ldots n$.

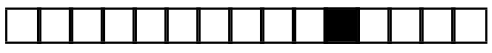
■ The alternative H_1 could be general, but is often specified, for instance, $H_1 : EX_i = EX_{i-1} + h,\ i = 2, \ldots, n$, a systematic shift in the data is present.

When the sample comes from the normal distribution and H_0 is true, the test statistics

$$R = \frac{\sigma_a^2}{s^2}$$

has moments $ER = 1,\ Var(R) = \frac{1}{n+2} + o(n^{-3})$ and $Z = (R-1)\sqrt{\frac{n^2-1}{n-2}}$ has approximately standard normal distribution. When $n > 60$, an α-level approximate rejection region for H_0 is $\{R < R_\alpha^*\}$, where $R_\alpha^* = 1 + z_\alpha/\sqrt{n + \frac{1+z_\alpha^2}{2}}$.

■ Both Allan and Hadamard Variances are special cases of so called generalized variations which use differencing filters. When viewed from the prospective of wavelets, Allan variance is Haar wavelet variance.



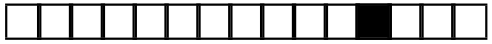
Zero-Crossing Method

The zero-crossing method is based on counting the number of zero crossings Z_N, producing an estimate of the Hurst exponent as:

$$\hat{H} = \frac{1}{2}\{1 + \log_2(1 \pm |\cos(\pi S_N)|)\}$$

where $S_N = Z_N/(N-1)$ is an average number of zero-crossings for the differenced time series of length N, and sign + (alternatively –) in $\pm$ is taken if the true exponent H is above (below) 1/2.

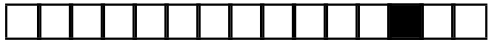
Usually, it is not hard to judge whether the true exponent H is above 1/2 by the observing the time series plot if this value is not sufficiently close to zero. It was also demonstrated that the $\hat{H}$ is asymptotically Gaussian for the fractional Brownian motion (fBm) models if the length of the time series is large enough and the true Hurst exponent does not exceed 3/4.



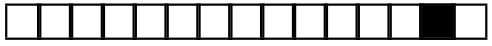
■ The estimation of $\hat{H}$ via zero-crossings is valid only for time series with stationary increments. For signals lacking stationary increments, such as the case for several turbulence time series measurements, we call $\hat{H}$ the *quasi-Hurst* exponent.

■ One of the attributes differentiating turbulence signals from fBm is the distinction in $\hat{H}$. Theoretically, the quasi-Hurst and Hurst exponents coincide for fBm, since fBm has stationary increments.

■ We empirically demonstrate this convergence using 1000 fBm paths. The resulting $\hat{H}$ is 0.3331 with standard deviation of 0.06. On the other hand, the quasi-Hurst exponent for turbulence signals is quite variable and significantly exceeds 1/3, which may be utilized to diagnose how atmospheric stability alters the global scaling parameter. That is, by analyzing deviations of $\hat{H}$ from 1/3 for turbulence measured under different stability conditions, a logical basis for tracking how atmospheric stability conditions impact global scaling of inertial subrange turbulence can be developed.



```
 %
function [Hhat, sn] = ZeroCross(seq, sign)
 % ========================================================
 % function [Hhat, sn] = ZeroCross(seq)
 %
 %  seq -- sequence to be estimated for selfsimilarity
 %  sign --  (-1) if expected H < 0.5; (1) if expected H > 0.5
 %  Hhat -- estimator of Hurst exponent.
 %  sn -- propostion of zerocrossings among neighboring pairs
 % ========================================================
 finest = diff(seq,1);
 nfin = length(finest);
 sn = 0;
 for i = 1: nfin-1
     sn = sn + (finest(i) * finest(i+1) < 0);
 end
 sn = sn/(nfin-1);
 Hhat = 1/2 * (1 + log2( 1 + sign *  abs( cos( pi * sn )) ) );
 %
```



OTHER METHODS TO ESTIMATE H

- Quadratic Variations
- Log-Periodogram
- Lobato's Spectral CDF Method
- Wavelet Based Methods `log2spectra.m`
- Whittle Method
- Convex Rearrangements