The Abdus Salam International Centre for Theoretical Physics

United Nations Educational, Scientific and Cultural Organization
IAEA International Atomic Energy Agency

1864-30

Ninth Workshop on Non-linear Dynamics and Earthquake Predictions

1 - 13 October 2007

Long Memory Processes

Filippo Petroni

GRAPES, B5, Sart Tilman,
Liege, Belgium

Long Memory Processes

Filippo Petroni

GRAPES, B5, Sart Tilman,

B-4000 Liège, Belgium

Ninth Workshop on Non-linear Dynamics and Earthquake Predictions

Outline

1. Brief Introduction
2. Definition of Long Memory Processes
3. Relation between Long Memory and Self-Similar Processes
4. Short overview of the methods to detect long memory
5. An example with real data

Introduction

A great challenge in modern times is the construction of predictive theories for non-linear dynamical systems for which the evolution equations are barely known, if known at all.

The more dependence there is between X_{n+k} and the past values $X_n, X_{n+1}, \ldots$ the better X_{n+k} can be predicted.

Stochastic process may be used to model the behavior of an observed time series in a purely statistical way, without a direct physical interpretation of the parameters.

Definition Let X_t be a stochastic process with finite mean $\mu = E[X_t]$ and variance $\sigma^2 = E[(X_t - \mu)^2]$, the autocorrelation between X_i and X_j is defined as

$$\rho(i,j) = \frac{\gamma(i,j)}{\sigma^2} \quad (1)$$

where $\gamma(i,j)$ is the autocovariance between X_i and X_j and is defined as

$$\gamma(i,j) = E[(X_i - \mu)(X_j - \mu)] \quad (2)$$

In the following we will indicate $\rho(i, i+k)$ as $\rho(k)$.

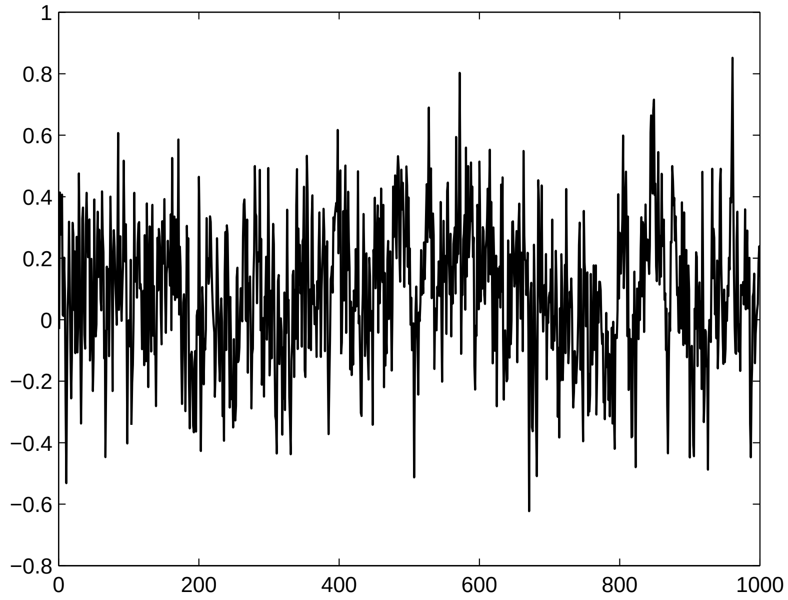
Definition The spectral density for the same stochastic process X_t is defined as

$$f(\lambda) = \frac{\sigma^2}{2\pi} \sum_{k=-\infty}^{\infty} \rho(k) e^{ik\lambda} \tag{3}$$

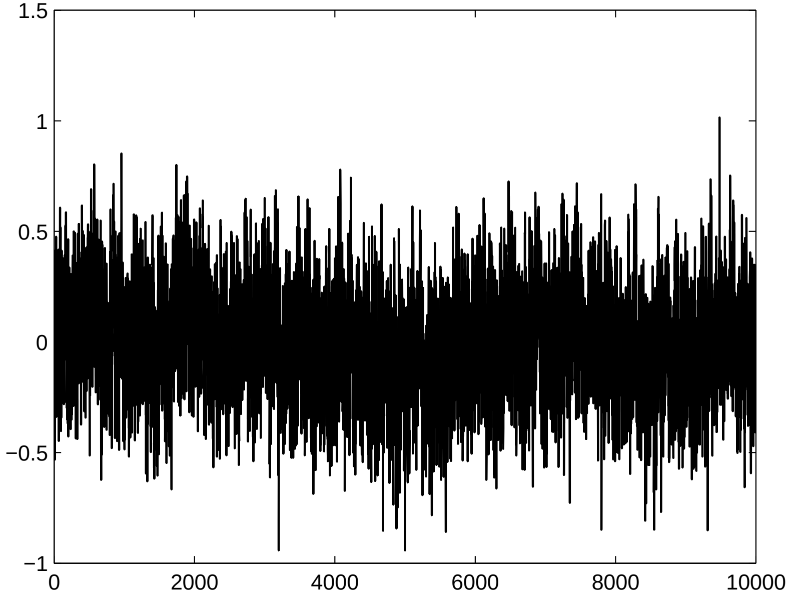
Stationary processes with long memory

A. Qualitative features:

1. There are relatively long periods where the observations tend to stay at high (low) level
2. Looking only at short time periods, there seems to be cycles/trends. However, looking at the whole time series, there is no apparent persisting cycles/trend.
3. Overall the series look stationary



A simple example of fractional noise with H=0.8



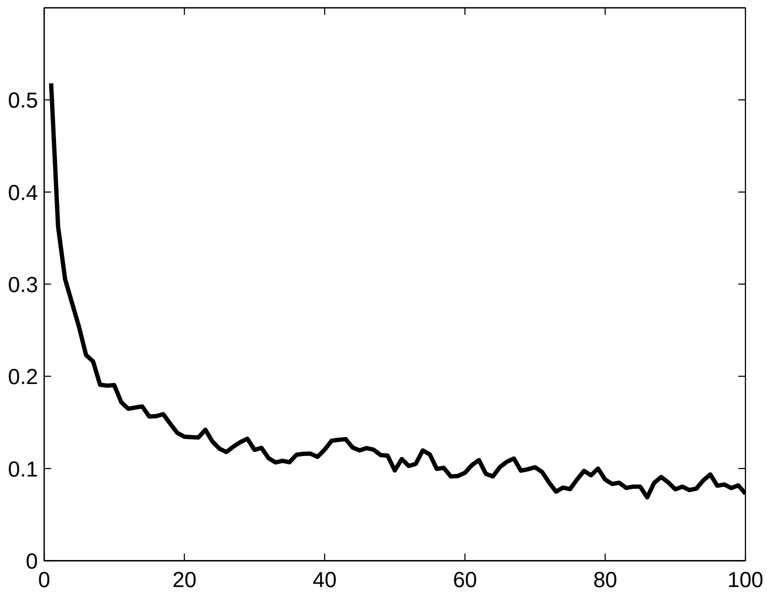
Same as before but for a longer period

B. Quantitative features

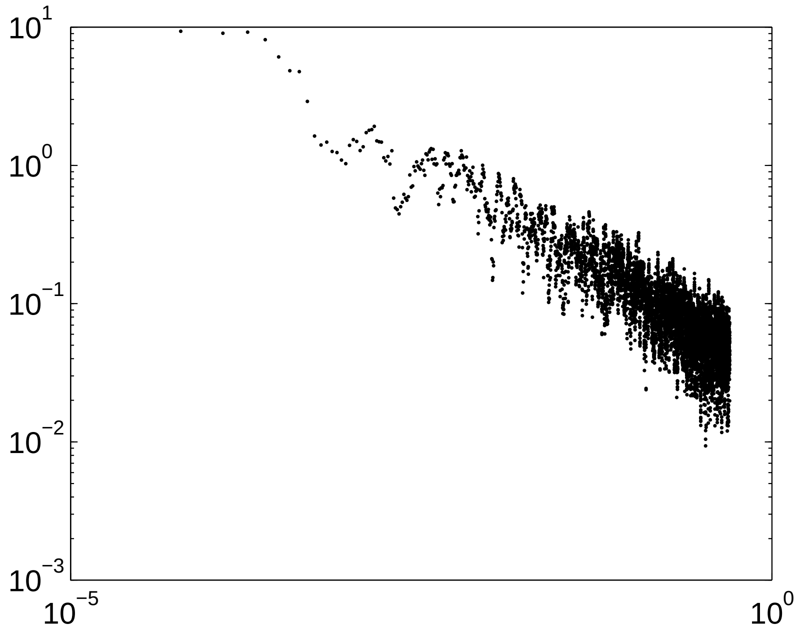
1. The variance of the sample mean seems to decay to zero slower than n^{-1}

2. The sample correlation $\rho(k) = \gamma(k)/\gamma(0)$ decays to zero at a rate that in good approximation is proportional to $k^{-\alpha}$ for some $0 < \alpha < 1$

3. Near the origin the logarithm of the power spectrum $f(\lambda)$ plotted against the logarithm of the frequency appears to be randomly scattered around a straight line with negative slope.



Correlations for the same process (fractional noise with H=0.8)



Power Spectrum for the same process (fractional noise with H=0.8)

Point A.3 implies that at least at first approximation, it is reasonable to assume stationarity. Let us therefore assume that the data are a sample path of a stationary process X_t.

Definition: Let X_t be a stationary process for which the following holds. There exist a real number $\alpha \in (0, 1)$ and a constant c_ρ such that

$$\lim_{k \to \infty} \rho(k)/[c_\rho k^{-\alpha}] = 1 \tag{4}$$

Then X_t is called a stationary process with long memory or long range dependence or strong dependence (or long range correlations).

The Hurst exponent is defined as $H = 1 - \alpha/2$, in terms of Hurst exponent long memory occurs if $\frac{1}{2} < H < 1$.

An equivalent definition of long memory is the following:

Definition: Let X_t be a stationary process for which the following holds: There exists a real number $\beta \in (0,1)$ and a constant $c_f > 0$ such that

$$\lim_{\lambda \to 0} f(\lambda)/[c_f|\lambda|^{-\beta}] = 1 \tag{5}$$

Then X_t is called a stationary process with long memory.

These two definitions are equivalent in the following sense:

Theorem: (i) Suppose 4 holds with $0 < \alpha = 2 - 2H < 1$. Then the spectral density f exists and

$$\lim_{\lambda \to 0} f(\lambda)/[c_f(H)|\lambda|^{1-2H}] = 1 \qquad (6)$$

where

$$c_f = \sigma^2 \pi^{-1} c_\rho \Gamma(2H-1)\sin(\pi - \pi H) \qquad (7)$$

and $\sigma^2 = var(X_t)$

(ii) Suppose 5 holds with $0 < \beta = 2H - 1 < 1$. Then

$$\lim_{k\to\infty} \rho(k)/[c_\rho k^{-2H-2}] = 1 \tag{8}$$

where

$$c_\rho = \frac{c_\gamma}{\sigma^2} \tag{9}$$

and

$$c_\gamma = 2c_f \Gamma(2-2H)\sin(\pi H - \frac{\pi}{2}) \tag{10}$$

It is important to note that the definition of long-range dependence by 4 and 5 is an asymptotic definition.

To summarize...

For a Long Memory Process

$$\rho(k) \approx k^{-\alpha} \tag{11}$$

$$f(\lambda) \approx |\lambda|^{-\beta} \tag{12}$$

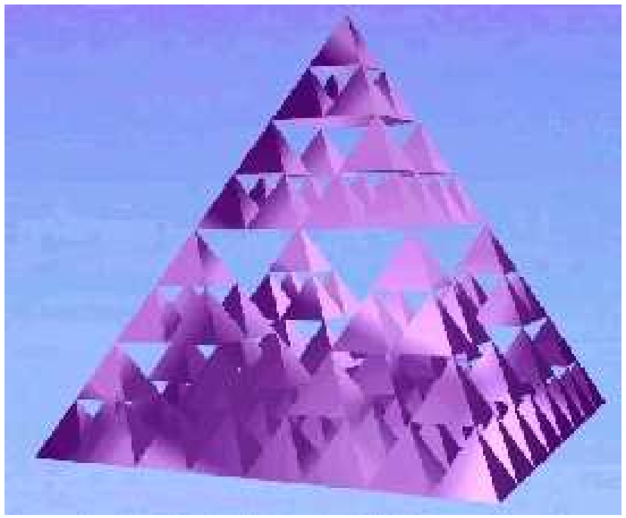
and

$$H = \frac{(2-\alpha)}{2} = \frac{(1+\beta)}{2} \tag{13}$$

with $\frac{1}{2} < H < 1$

Self Similar Processes

A geometric shape is called self-similar in a deterministic way if the same geometric structures are observed, independently of the distance from which one looks at the shape.



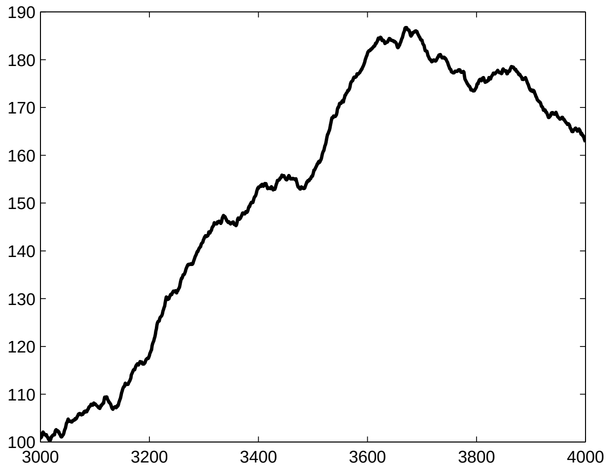
In the context of stochastic processes, self-similarity is defined in terms of the distribution of the process:

Definition: Let Y_t be a stochastic process with continuous time parameter t. Y_t is called self-similar with self-similarity parameter H, if for any positive stretching factor c, the rescaled process with time scale ct, $ct^{-H} Y_{ct}$, is equal in distribution to the original process Y_t.

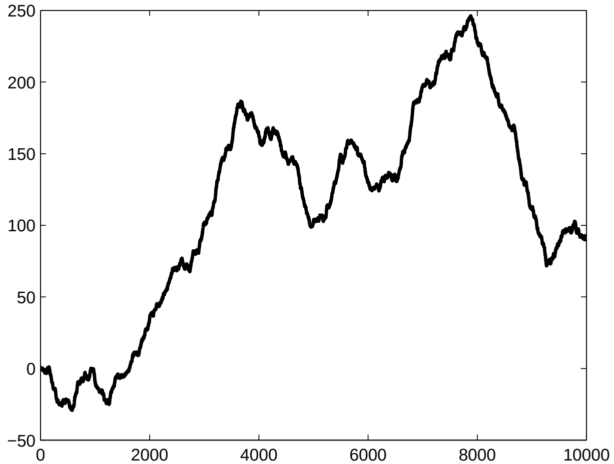
This means that for any sequences of time $t_1, \ldots, t_k$ and any positive constant c, $c^{-H}(Y_{ct_1}, Y_{ct_2}, \ldots, Y_{ct_k})$ has the same distribution of $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_k})$.

Thus, typical sample paths of self-similar process look qualitatively the same, irrespective of the distance from which we look at them.

In contrast to deterministic self-similarity, it does not mean that the same picture repeats itself exactly as we go closer. It is rather the general impression that remains the same.



190
180
170
160
150
140
130
120
110
100
3000
3200
3400
3600
3800
4000



250
200
150
100
50
0
−50
0
2000
4000
6000
8000
10000

Self-similarity arises in a natural way from limit theorems for sums of random variables.

Definition: if for any $k \geq 1$ and any k time point $t_1, \ldots, t_k$, the distribution of $(Y_{t_1+c} - Y_{t_1+c-1}, \ldots, Y_{t_k+c} - Y_{t_k+c-1})$ does not depend on $c \in R$, than we say that Y_t has stationary increments.

Theorem: (i) Suppose that Y_t is a stochastic process such that $Y_1 \neq 0$ with positive probability and Y_t is the limit in distribution of the sequence of normalized partial sums

$$a_n^{-1} S_{nt} = a_n^{-1} \sum_{i=1}^{[nt]} X_i, \quad n = 1, 2, \ldots \tag{14}$$

Here $[nt]$ denotes the integer part of nt, $X_1, X_2, \ldots$ is a stationary sequence of random variables, and $a_1, a_2, \ldots$ is a sequence of positive normalizing constants such that $\log a_s \to \infty$. Then there exists an $H > 0$ such that for any $u > 0$

$$\lim_{n \to \infty} \frac{a_{nu}}{a_n} = u^H \qquad (15)$$

and Y_t is self-similar with self-similarity parameter H, and has stationary increments.

(ii) All self-similar processes with stationary increments and $H > 0$ can be obtained by partial sums as given in (i).

Part (i) says that whenever a process is the limit of normalized partial sums of random variables, it is necessarily self-similar.

Suppose that Y_t is a self-similar process with self-similarity parameter H. The property

$$Y_t =_d t^H Y_1 \quad (t > 0) \tag{16}$$

where $=_d$ is equality in distribution, implies the following limiting behavior of Y_t as t tends to infinity:

1. If $H < 0$, then $Y_t \to_d 0$

2. If $H = 0$, then $Y_t =_d Y_1$

3. if $H > 0$ and $Y_t \neq 0$, then $|Y_t| \to_d \infty$

Analogously, for t converging to zero, we have:

1. If $H < 0$ and $Y_t \neq 0$, then $|Y_t| \to_d \infty$

2. If $H = 0$, then $Y_t =_d Y_1$

3. if $H > 0$ and $Y_t \neq 0$, then $Y_t \to_d 0$

If we exclude the trivial case $Y_t \equiv 0$, then these properties imply that Y_t is not stationary unless $H = 0$.

For the purpose of modelling data that look stationary, we need only to consider self-similar processes with stationary increments. The range of H can then be restricted to $H > 0$.

The form of the covariance function $\gamma_y(t, s) = cov(Y_t, Y_s)$ of a self-similar process Y_t with stationary increments follows from these two properties. Let us assume that $E[Y_t] = 0$.

Let $s < t$ and denote by $\sigma^2 = E[(Y_t - Y_{t-1})^2] = E[Y_1^2]$ the variance of the increments process $X_t = Y_t - Y_{t-1}$. Then

$$E[(Y_t - Y_s)^2] = E[(Y_{t-s} - Y_0)^2] = \sigma^2 (t - s)^{2H} \quad (17)$$

The covariance $\gamma(k)$ of the increments sequence $X_i = Y_i - Y_{i-1}$, between X_i and X_{i+k} for $k > 0$ is

$$\gamma(k) = \frac{1}{2}\{E[(Y_{k+1}-Y_0)^2]-E[(Y_{k-1}-Y_0)^2]-2E[(Y_k-Y_0)^2]\} \tag{18}$$

Using self-similarity, we obtain the formula

$$\gamma(k) = \frac{1}{2}\sigma^2[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}] \tag{19}$$

The correlations are then given by

$$\rho(k) = \frac{1}{2}[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}] \tag{20}$$

The asymptotic behavior of $\rho(k)$ follows from Taylor expansion

$$\frac{\rho(k)}{[H(2H-1)k^{2H-2}]} \to 1 \tag{21}$$

as $k \to \infty$. For $1/2 < H < 1$ this means that the correlations decay to zero so slowly that

$$\sum_{k=-\infty}^{\infty} \rho(k) = \infty \tag{22}$$

The process $X_i \, (i = 1, 2, \dots)$ has long memory.

For $H = 1/2$, all correlations at non-zero lags are zero, i.e., the observations X_i are uncorrelated.

For $0 < H < 1/2$, the correlations are summable. In fact a more specific equation holds, namely,

$$\sum_{k=-\infty}^{\infty} \rho(k) = 0 \qquad (23)$$

In practice, this case is rarely encountered and the sum is equal to some constant c different from zero.

We will assume that the variance is always finite and $\lim_{k \to \infty} \rho(k) = 0$ and then $0 < H < 1$.

Under this assumptions, the spectral density of the increments X_i can be calculated directly and expanded near the origin giving

$$f(\lambda) \approx c_f |\lambda|^{1-2H} \tag{24}$$

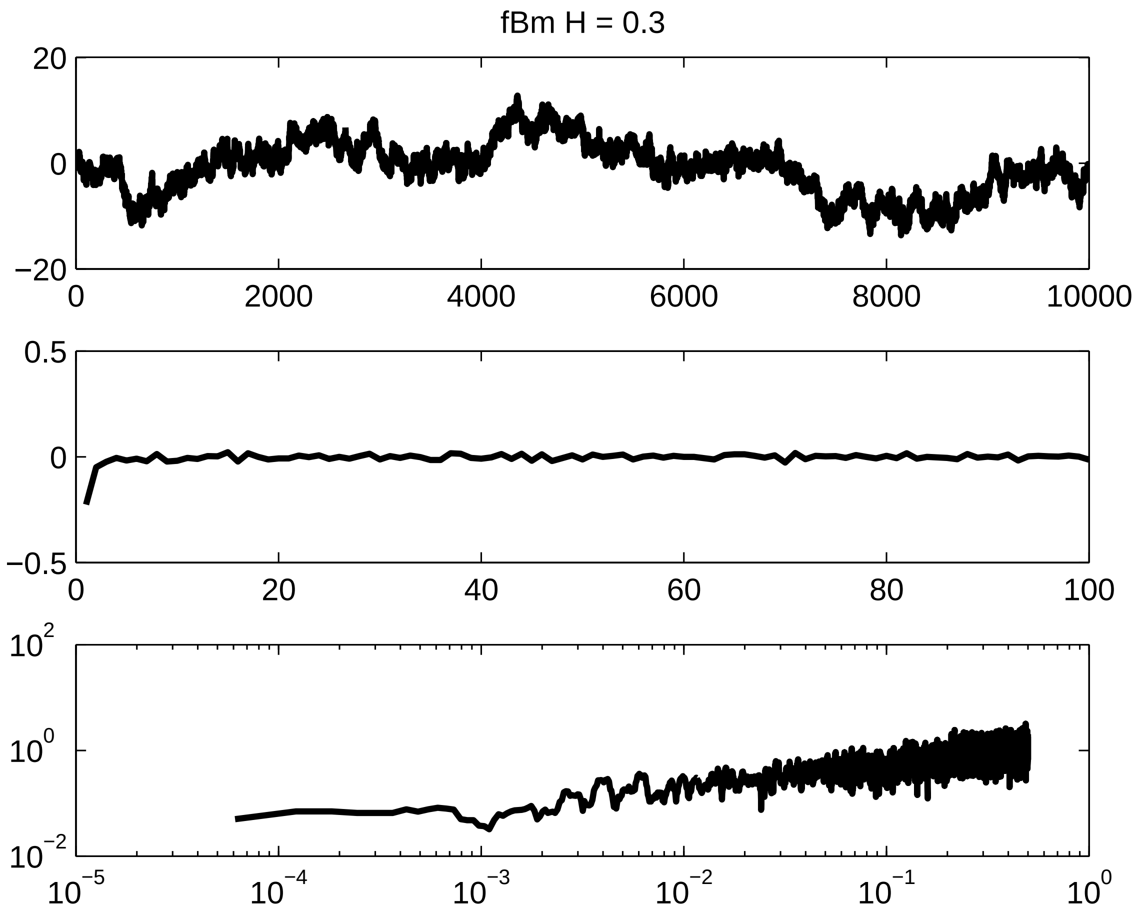
To summarize... again

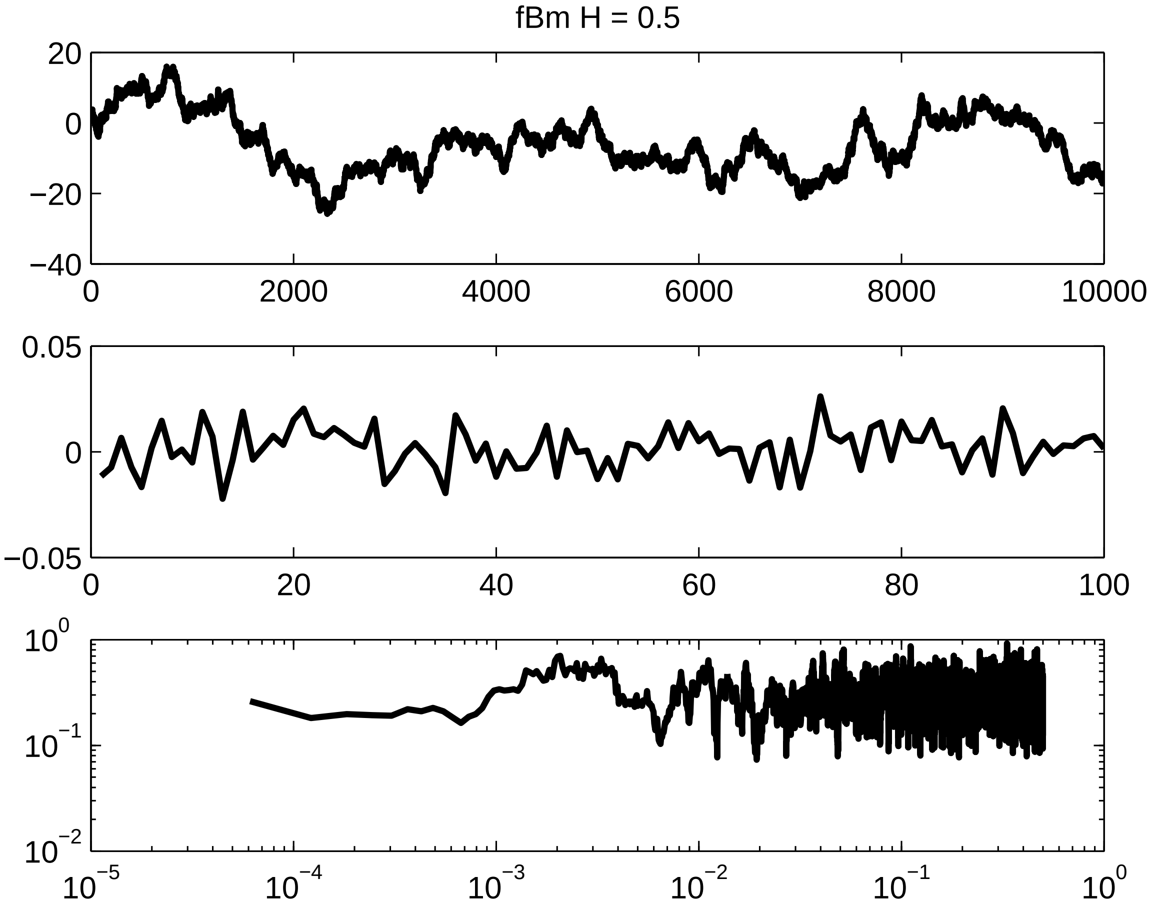
Let X_i be a stationary stochastic process

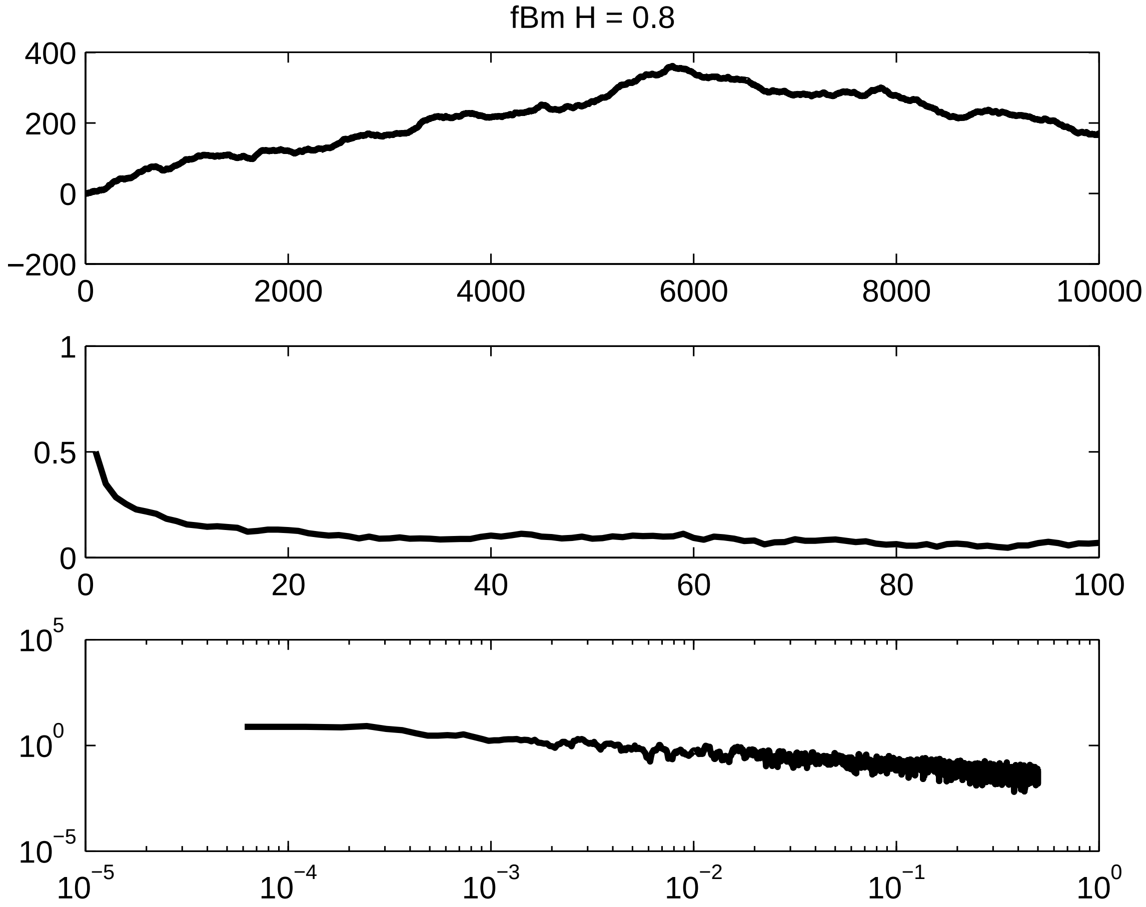
- $\frac{1}{2} < H < 1$ or $0 < \alpha < 1$ or $0 < \beta < 1$ then X_i has Long Memory (persistent)

- $H = \frac{1}{2}$ then correlations are zero

- $0 < H < \frac{1}{2}$ correlations are, in fact, negative and the process is anti-persistent


fBm H = 0.3
20
0
−20
0
2000
4000
6000
8000
10000
0.5
0
−0.5
0
20
40
60
80
100
10^2
10^0
10^−2
10^−5
10^−4
10^−3
10^−2
10^−1
10^0


fBm H = 0.5
20
0
−20
−40
0
2000
4000
6000
8000
10000
0.05
0
−0.05
0
20
40
60
80
100
10^{0}
10^{-1}
10^{-2}
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}


fBm H = 0.8
400
200
0
−200
0
2000
4000
6000
8000
10000
1
0.5
0
0
20
40
60
80
100
10^{5}
10^{0}
10^{-5}
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}

Fractional Brownian Motion and Gaussian Noise

Suppose that Y_t is a self-similar process with stationary increments (N.B. we do not request stationarity of the process itself). That the increments $X_i = Y_i - Y_{i-1}$ have zero mean and that they are Gaussian.

For each value of $H \in (0, 1)$ there is exactly one Gaussian process X_i that is the stationary increment of a self similar process Y_t. This process is called **fractional Gaussian noise**.

The corresponding self-similar process Y_t is called **fractional Brownian motion** denoted by $B_H(t)$

In the simple case of $H = 1/2$, X_i are independent normal variable $\rightarrow B_{\frac{1}{2}}(t)$ is an ordinary Brownian motion.

Definition: Let $B(t)$ be a stochastic process with continuous sample paths and such that

- $B(t)$ is Gaussian
- $B(0) = 0$ almost surely
- $B(t)$ has independent increments

- $E[B(t) - B(s)] = 0$

- $var[B(t) - B(s)] = \sigma^2|t - s|$

Then $B(t)$ is called Brownian motion.

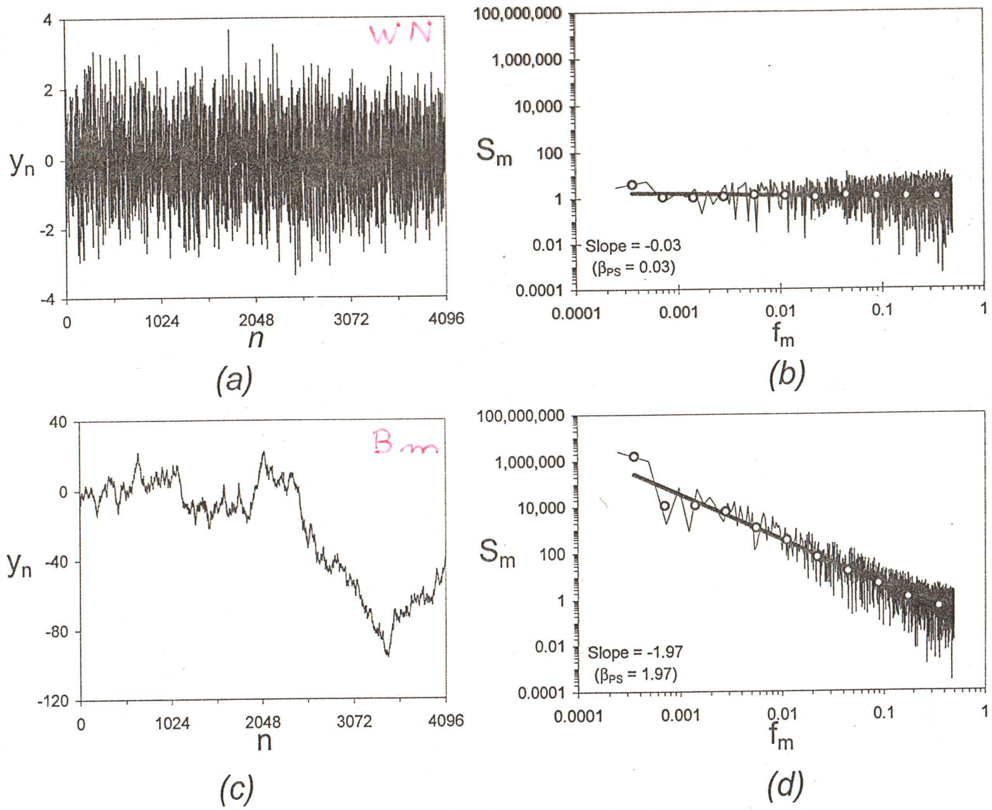
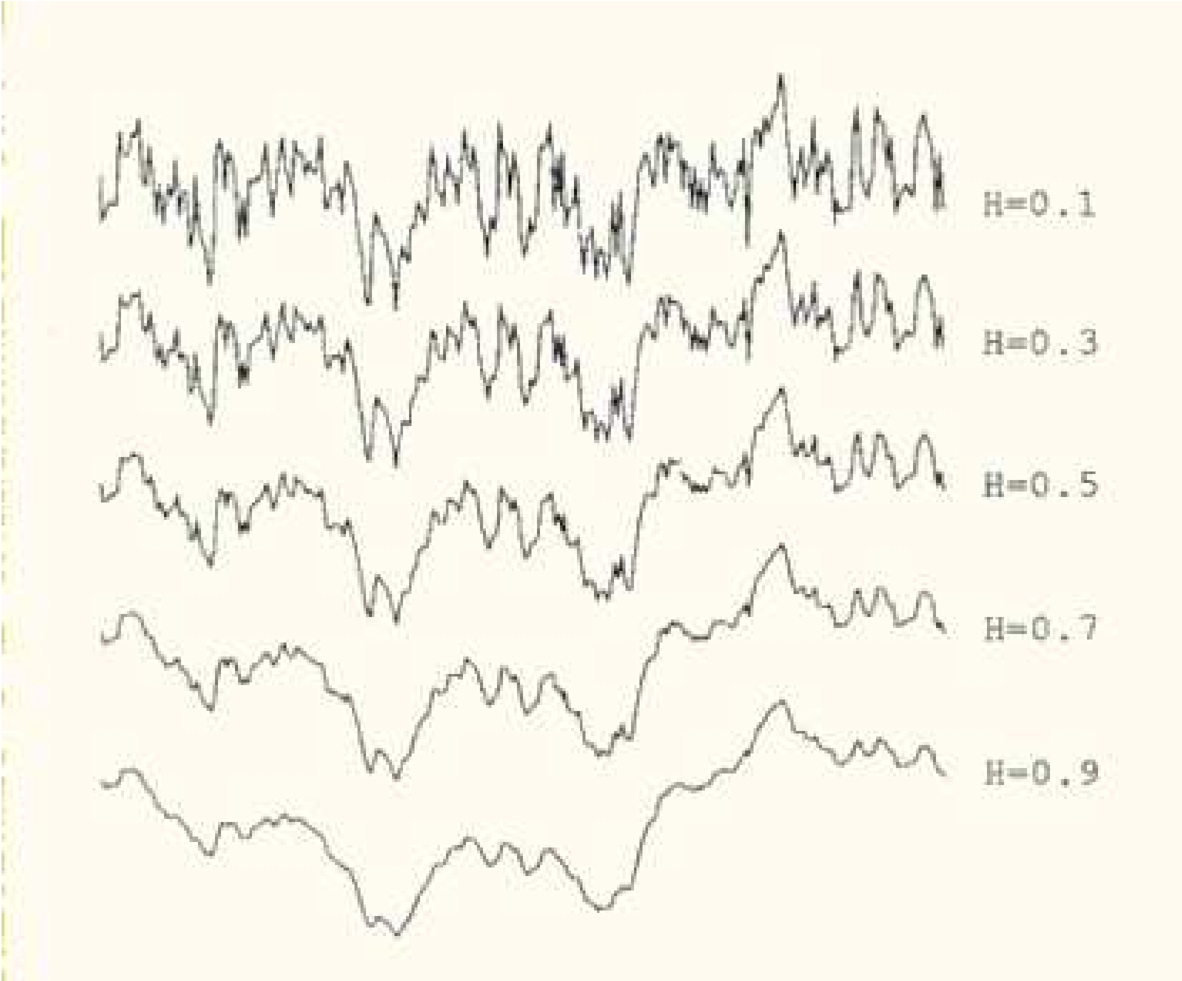


Figure 1, Malamud and Turcotte


H=0.1
H=0.3
H=0.5
H=0.7
H=0.9

Detection of Long Memory

1. Detrended Fluctuation Analysis

2. Rescaled Range Analysis

3. Spectral Analysis

Detrended Fluctuation Analysis

The method has been used to identify whether long range correlations exist in many research fields such as e.g. finance, cardiac dynamics, meteorology etc. etc.

- The signal time series $X(i)$, $i = 1, 2, \ldots, N$ is $first$ $integrated$

$$Y(k) = \sum_{i=1}^{k} [X(i) - < X >] \tag{25}$$

where $< X >$ is the mean;

- The time axis (from 1 to N) is next divided into non-overlapping boxes of equal size n;

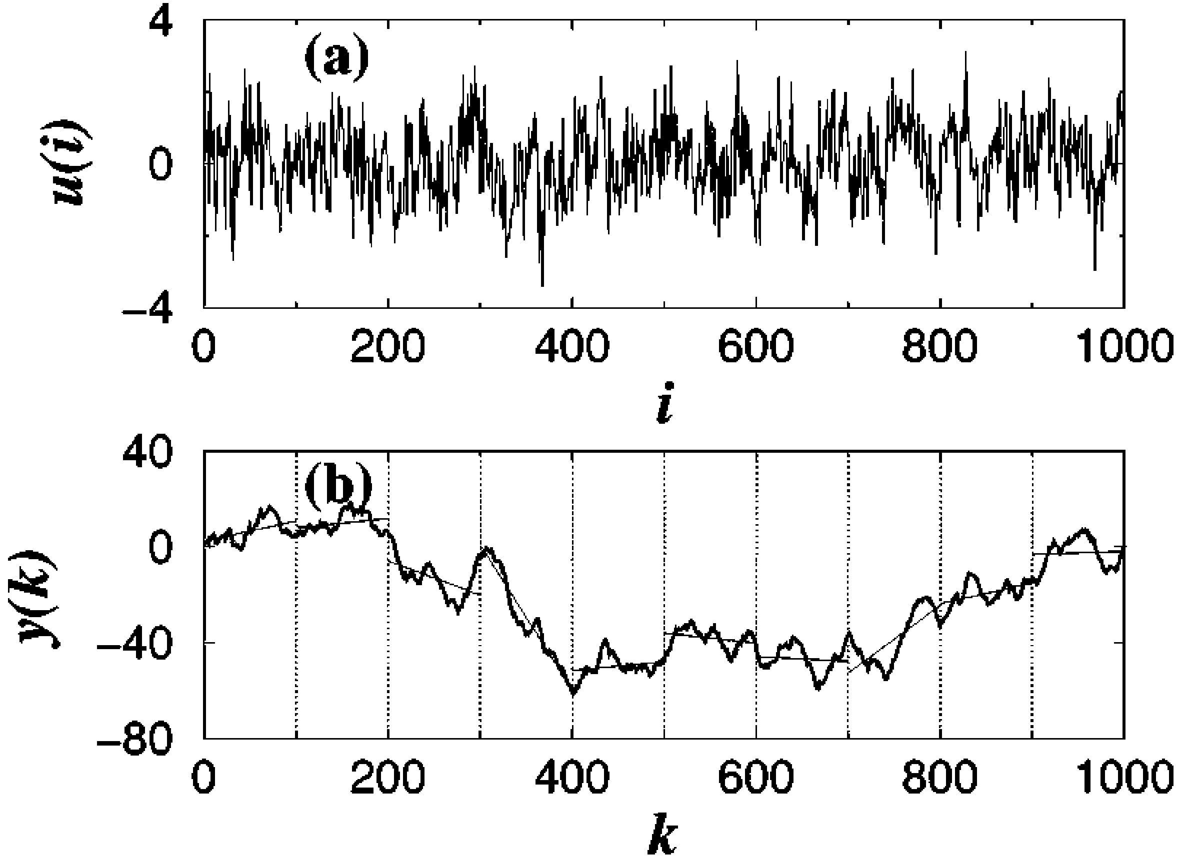
- In each box of length n one looks thereafter for the best (polynomial, of degree m) trend, $z_{n,m}$;

- The integrated signal in each box is detrended by subtracting the local trend;

- For each box size n, the root mean square deviation of the (integrated) signal is calculated

$$F_m(n) \equiv \sqrt{\frac{1}{N}\sum_{k=1}^{N} [Y(k) - z_{n,m}(i)]^2} \ ; \qquad (26)$$

- the above computation is repeated for a broad range of scales (box size n) to provide a relationship between the fluctuation function $F_m(n)$ and the box size n;

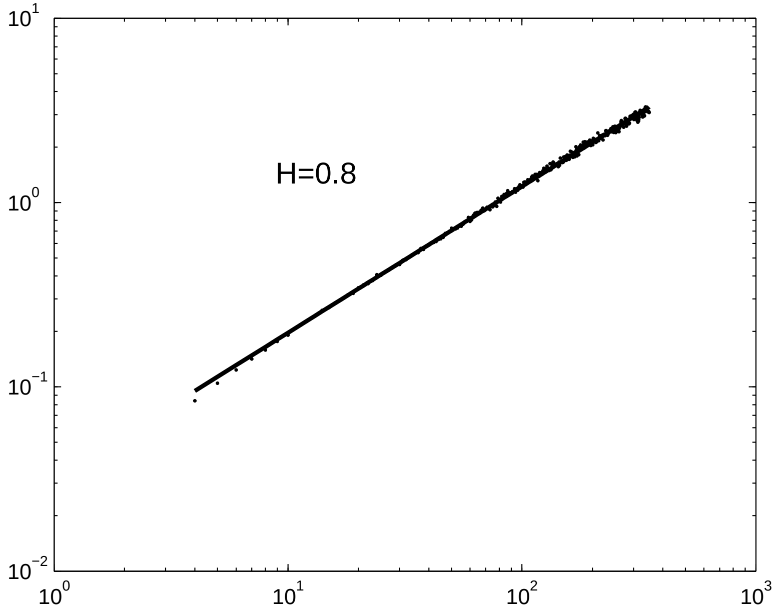


(a)
u(i)
4
0
−4
0
200
400
600
800
1000
i
(b)
y(k)
40
0
−40
−80
0
200
400
600
800
1000
k

If the original time series $X(i)$ is long range correlated (has long memory) the fluctuation function $F_m(n)$ increases following a power law:

$$F_m(n) \approx n^{\gamma} \tag{27}$$

and the exponent γ is related to the correlation exponent α. It turns out that $\gamma = H$.



DFA for the same process (fractional noise with H=0.8)

Rescaled Range Analysis

- The signal time series $X(i)$, $i = 1, 2, \ldots, N$ is divided into l boxes of equal size n;

- In the k^{th} box, $(k = 1, \ldots, l)$, there are n elements;

- the local fluctuation at point j in the k^{th} box is

$$F^{(k)}(n) = X^{(k)}(j) - < X >_n^{(k)} \tag{28}$$

  where $<>_n$ is the mean in the k^{th} box;

- Let

$$S^{(k)}(n) = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(F(n)^{(k)})^2} \tag{29}$$

- The cumulative departure $Y_m^{(k)}(n)$ up to the m^{th} point in the k^{th} box (of size n) is next calculated

$$Y_m^{(k)}(n) = \sum_{j=1}^{m} F^{(k)}(n) \tag{30}$$

  for $m = 1, \ldots, n$ and in all k boxes

- look for $\max_{1\le m\le n} Y_m^{(k)}(n)$ and $\min_{1\le m\le n} Y_m^{(k)}(n)$

- The rescaled range function is then defined by

$$\frac{R^{(k)}(n)}{S^{(k)}(n)} = \frac{\max_{1\leq m\leq n} Y_m^{(k)}(n) - \min_{1\leq m\leq n} Y_m^{(k)}(n)}{\sqrt{\frac{1}{n}\sum_{j=1}^{n} (F(n)^{(k)})^2}}$$

for $k = 1, \ldots, l$

- The average of the rescaled range function for boxes of size n is obtained and denoted by $< R/S >_n$

- Everything is repeated for different n

- for Long Memory processes one expects

$$< R/S >_n \approx n^H \tag{31}$$

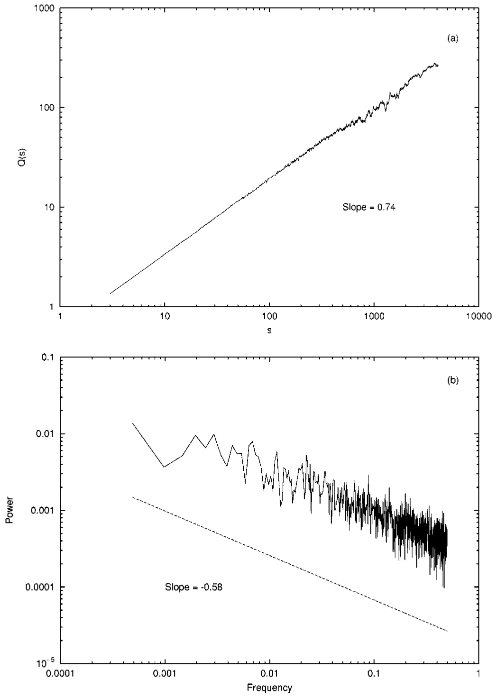


FIG. 1. (a) Log-log plot of the rescaled range statistic $Q(s)$ against the window size s for a true long range correlated process with $\alpha = 0.6$, variance 0.25, and a long data set of 8192 points. (b) Spectral density of the same data.

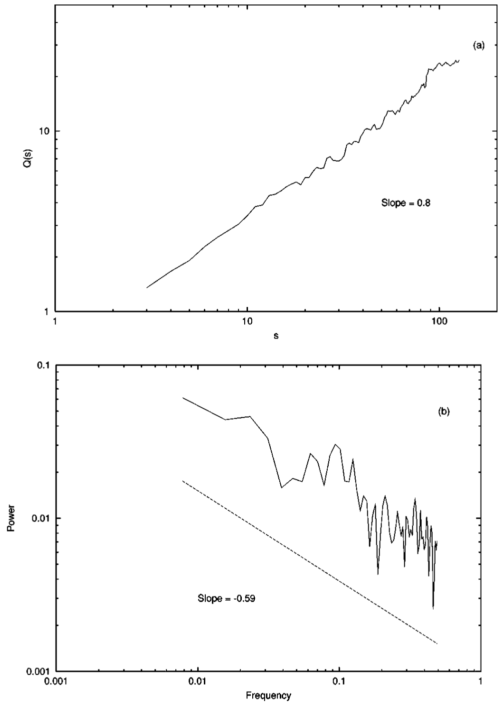


FIG. 2. (a) Log-log plot of the rescaled range statistic $Q(s)$ against the window size s for a true long range correlated process with $\alpha = 0.6$, variance 0.25, and a short data set of 256 points. (b) Spectral density of the same data.

Two examples taken from [2]

Spectral Analysis

The periodogram (the sample analogon of the spectral density) is defined as

$$I(\lambda_j) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} (X_t - < X >_n) e^{it\lambda_j} \right|^2 = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \widehat{\gamma}(k) e^{it\lambda_j}$$

where

$$\widehat{\gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} (X_t - < X >_n)(X_{t+k} - < X >_n)$$

are the sample covariances

It is expected for a long memory process

$$I(\lambda_j) \approx |\lambda|^{1-2H}$$

References

1. J. Beran, Statistics for Long-Memory Processes, Chapman & Hall, New York (1994).

2. G. Rangarajan, M. Ding, Phys. Rev. E 61 (2000) 4991-5001.

3. B.D. Malamud and D. L. Turcotte, J. Stat. Plann. Infer. 80 (1999) 173-196.