Time Series Analysis:

9. Long-memory processes

P. F. Góra

http://th-www.if.uj.edu.pl/zfs/gora/

2021

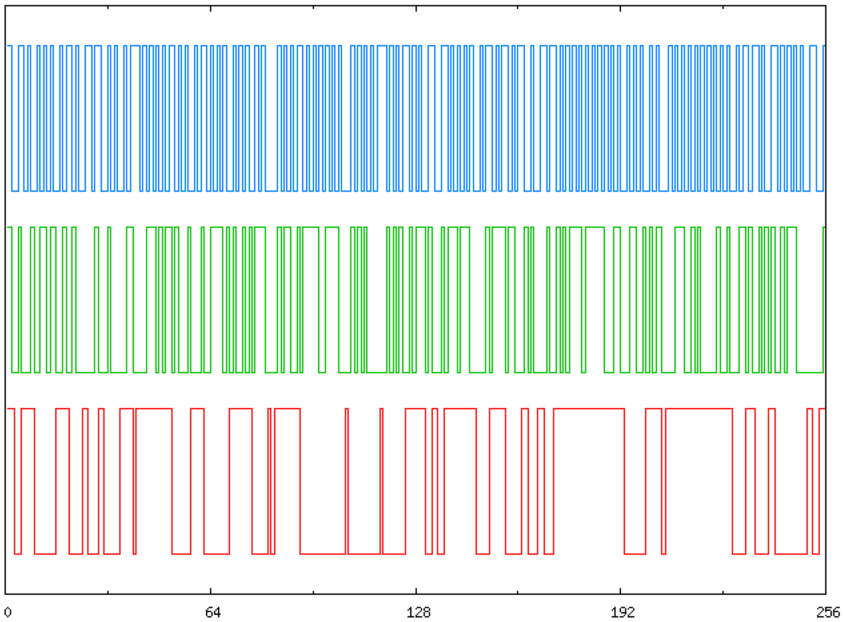
Joseph effect

According to the *Book of Genesis*, patriarch Joseph interpreted dreams of a pharaoh:

> *Pharaoh dreamed that he was standing by the Nile, and behold, there came up out of the Nile seven cows, attractive and plump, and they fed in the reed grass. And behold, seven other cows, ugly and thin, came up out of the Nile after them, and stood by the other cows on the bank of the Nile. And the ugly, thin cows ate up the seven attractive, plump cows.*
>
> Gen. 41

Joseph interpreted this dream that after seven years of abundant harvests there would be seven years of poor harvests and provisions needed to be made to feed the people during the years of poor harvests. As harvests in ancient Egypt depended entirely on Nile flooding, this can be interpreted as seven years of high flooding to be followed by seven years of low flooding. We can see that good/bad years tend to *cluster*. Benoit Mandelbrot has coined the term "Joseph effect" for this phenomenon. The ancient people understood that this clustering was a very important natural phenomenon and included its description, in a literary form, in their holy book.

Consider a *random* time series that can assume only two values, $\{+1,-1\}$.



The lowest curve: similar states tend to *cluster*. The middle curve: an uncorrelated series. The upper curve: an anticorrelated series that tends to flip its state at each step (a blue noise).

The Aswan High Dam

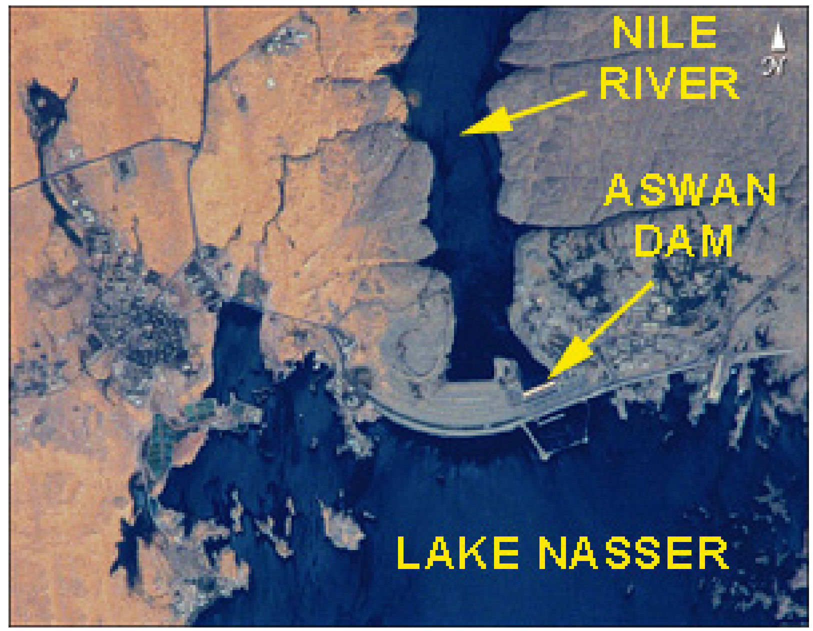
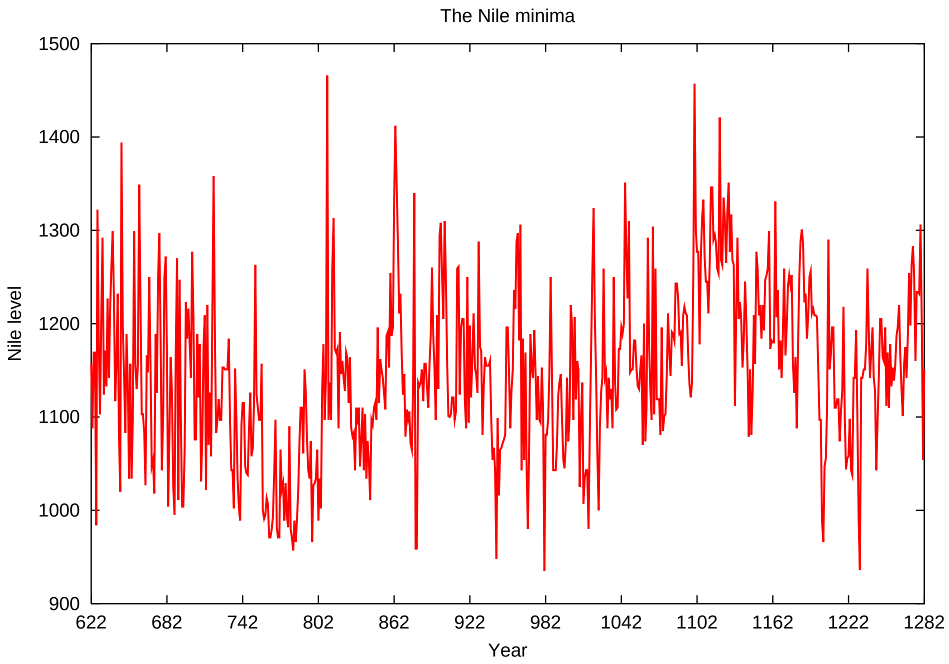


photo: NASA

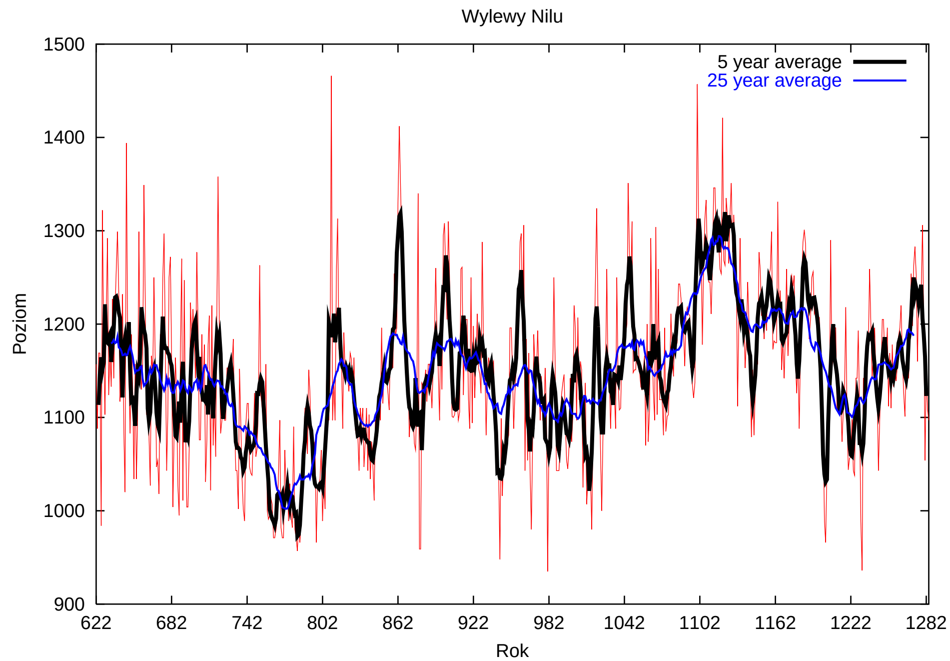
The Aswan Dam history	
around 1000s	The Arabs first attempt to regulate the flooding of the Nile
1889-1902	The British construct the first dam in Aswan
1907-1912	The dam is raised
1929-1933	The dam is raised for the second time
1946	The dam nearly overflows
	Decision is taken to build a new dam
1954	Design of a new dam begins
1960	Construction of The Aswan High Dam begins
1964	The reservoir starts to fill
1970	Construction ends
1976	The reservoir reaches its capacity

The Nile minima

In 1951, a British engineer, Harold Hurst, studied historic records on Nile minima:



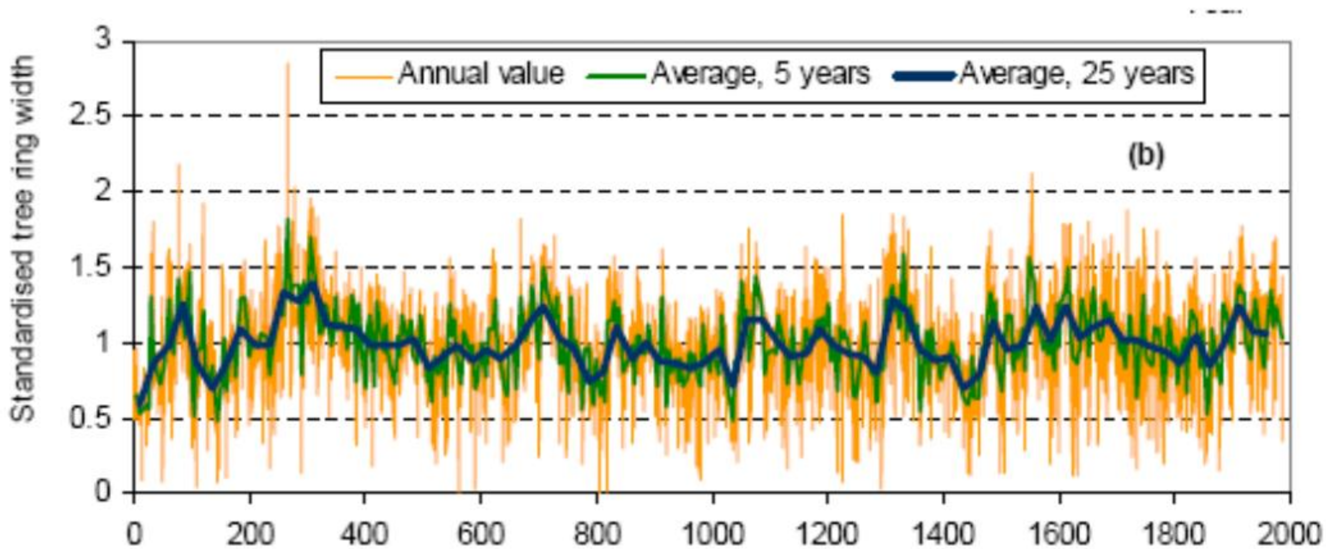
Moving averages *do not* flatten the series!



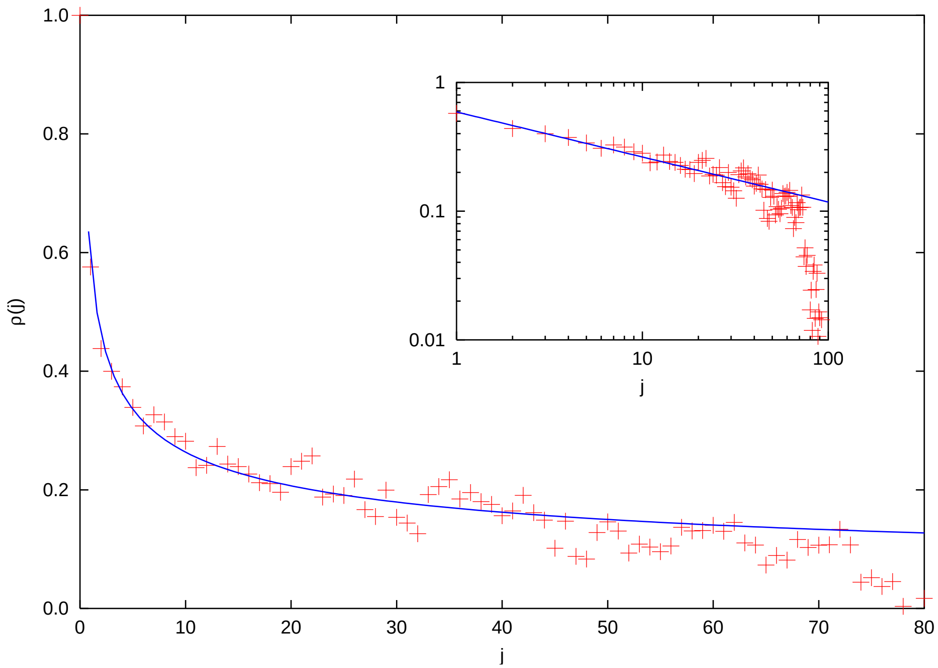
Wet years cluster, dry years cluster — Joseph effect

The series seems to be self-similar — it behaves similarly on (nearly) all scales

Another example — tree rings



Correlations in the Nile minima



Long-range correlations, decreasing according to a power law (the exponent ~ -0.35)

“Ordinary” correlations

The autocorrelation function of a stationary ARMA(p,q) process decreases rapidly as it is bounded from above

$$|\rho_k| \leqslant C\,R^k\,, \qquad C > 0\,,\ 0 < R < 1\,. \tag{1}$$

In contrast to that, long-memory processes have autocorrelations that are not necessarily very large but persist for very long times.

Long Range Dependencies

A stochastic process $\{\xi_n\}$ is said to have Long Range Dependencies (or a long range memory) if its autocorrelation function $\rho_k = \left\langle \xi_n \xi_{n+k} \right\rangle$ has the asymptotic form

$$\rho_k \sim k^{2d-1} L(k) \quad \text{as} \quad k \to \infty \tag{2}$$

where $-\frac{1}{2} < d < \frac{1}{2}$ and $L(k)$ is slowly varying at infinity:

$$\forall z > 0 \colon \lim_{t \to \infty} \frac{L(zt)}{L(t)} = 1 \tag{3}$$

Example: Constant functions and logarithms are slowly varying at infinity.

Examples of Long Range Dependencies

- Internet traffic — can impact heavily on queuing
- Telecommunications
- Hydrological phenomena
- Financial data (at least before the crisis ☺)
- Sunspots
- Climate; sequences of dry/wet years
- Movements of foraging animals
- fMRI signals, etc

Remember: GWN and "mild" forms of Gaussian coloured noises implicitly assumed some forms of thermal equilibrium.
Why should we expect GWN in non-equilibrium systems?
No reason. Not the case.

Fractional Gaussian noise (fGn)

$\{\xi_n\}$ — a Gaussian process, usually zero-mean, whose autocorrelation function satisfies

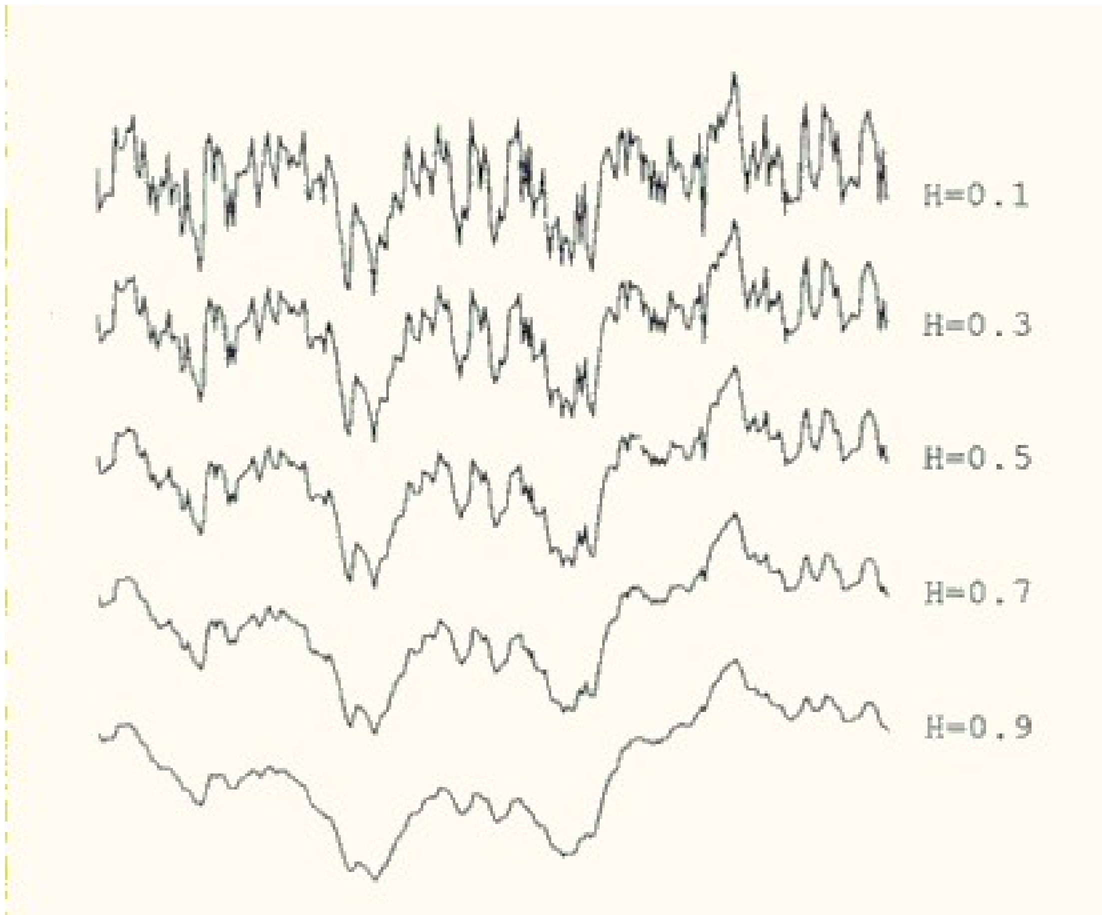
$$\rho_k = \left\langle \xi_n \xi_{n+k} \right\rangle = \frac{\sigma^2}{2}\left(|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}\right) \tag{4a}$$

and $H \in [0,1]$. For large k and $H \neq \frac{1}{2}$:

$$\rho_k \simeq \sigma^2 H(2H-1)k^{-(2-2H)} \tag{4b}$$

H is also called the Hurst exponent. $H = d + \frac{1}{2}$, cf. (2).

Examples of fGn's



Properties of fGn

- fGn is stationary
- fGn is self-similar
- For $H \in \left[\frac{1}{2}, 1\right]$, fGn displays Long Range Dependence
- For $H = \frac{1}{2}$, fGn reduces to GWN
- For $H \in \left[0, \frac{1}{2}\right]$, fGn has negative correlations

Power spectrum of fGn

Since fGn is stationary, Wiener-Khinchin theorem holds and we can calculate the power spectrum by Fourier transforming (4a). We get

$$P(f) = C\sigma^2 \left| e^{2\pi i f} - 1 \right|^2 \sum_{k=-\infty}^{\infty} \frac{1}{|f+k|^{2H+1}} \tag{5a}$$

for $|f| \leqslant \frac{1}{2}$. If $H \neq \frac{1}{2}$,

$$P(f) \sim C\sigma^2 |f|^{1-2H} \quad \text{for} \quad f \to 0 \tag{5b}$$

fGn displays power-law behaviour for low frequencies (or long waves).

Estimating the Hurst exponent

We can see from (5b) that if a straight line with a slope α ($\alpha < 0$) (in a log-log scale) is fitted to the power spectrum, then

$$H = \frac{1-\alpha}{2} \tag{6}$$

Fractional Brownian Motion (fBm)

Let $\{\xi_k\}$ be a fGn. Then

$$X_n = \sum_{k=1}^{n} \xi_k \tag{7}$$

is called a Fractional Brownian Motion. Observe that

$$X_n = X_{n-1} + \xi_n \tag{8}$$

The autocorrelation function

$$\langle X_n X_m \rangle = \frac{\sigma^2}{2} \left(|n|^{2H} + |m|^{2H} - |n-m|^{2H} \right) \tag{9}$$

The variance then satisfies

$$\left\langle X_n^2 \right\rangle = \sigma^2 |n|^{2H} \tag{10}$$

Properties of fBm

- fBm is non-stationary
- For $H \in \left[\frac{1}{2}, 1\right]$, fBm is a superdiffusion
- For $H = \frac{1}{2}$, fBm is a normal diffusion
- For $H \in \left[0, \frac{1}{2}\right]$, fBm is a subdiffusion
- Super- and subdiffusion go under a joint term, anomalous diffusion
  - ◇ But not all anomalous diffusions are fBm's!

Detrended Fluctuation Analysis (DFA)

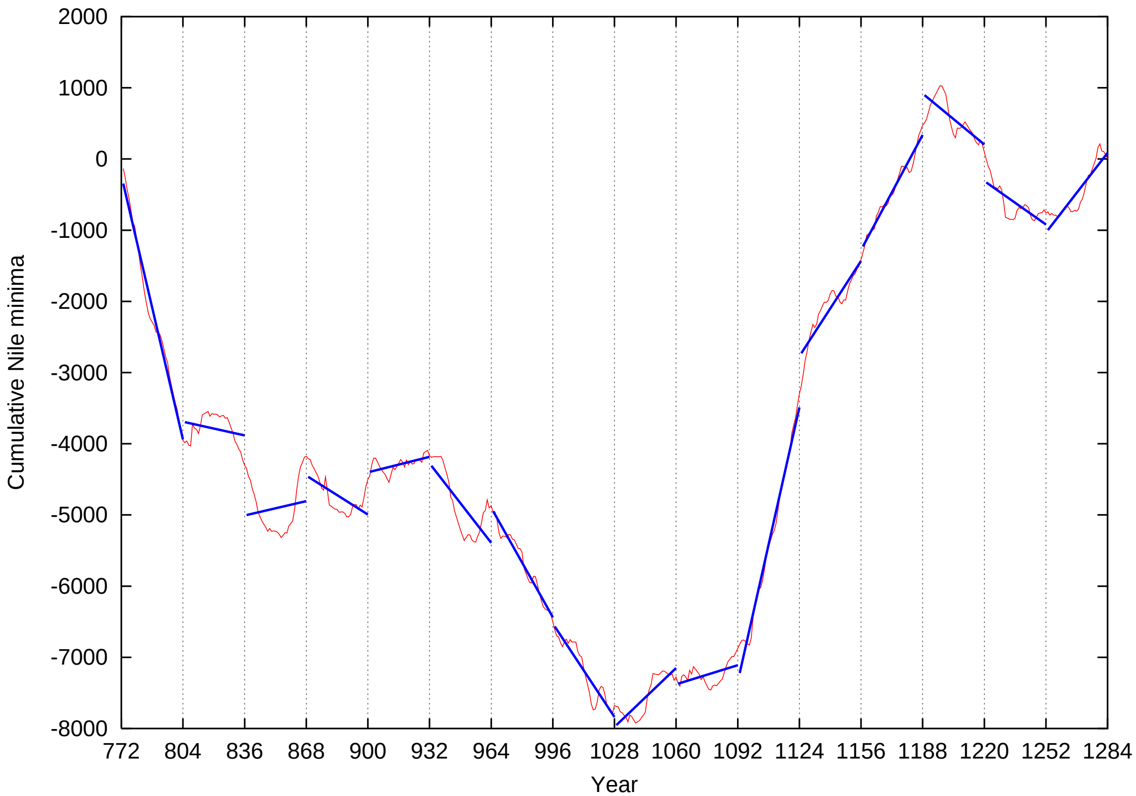
Given a (possibly) fractional series $\{x_n\}$, do the following:

1. Calculate the average $\langle x_n \rangle$

2. Convert the series into a "random walk", i.e., calculate the cumulative sums:

$$X_n = \sum_{k=1}^{n} (x_k - \langle x_n \rangle) \tag{11}$$

3. Divide the series into segments of length L and fit a straight line within each segment. Let the line be $X_n = a \cdot n + b$.



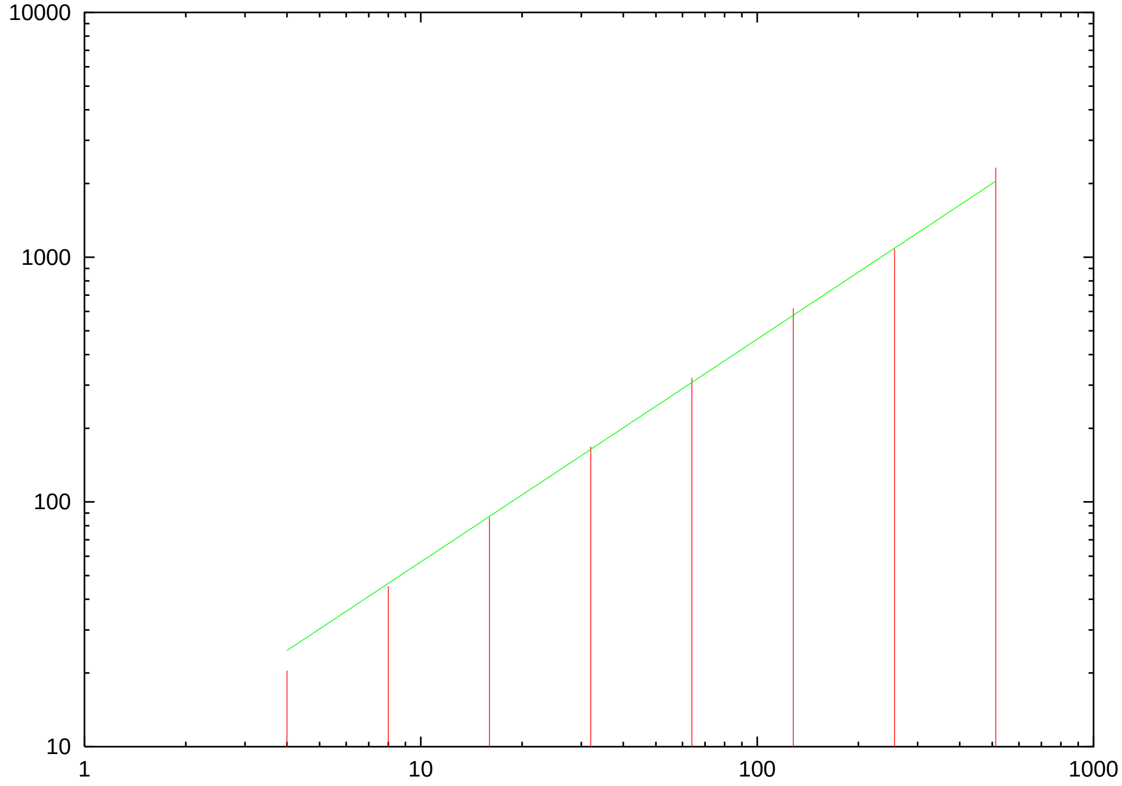
Cumulative Nile minima series divided into segments of length 32.
A straight line, representing a local trend, is fitted within each segment.

4. Within each segment, calculate

$$F(L) = \sqrt{\frac{1}{L} \sum_{i=i_0}^{i_0+L-1} (X_i - a \cdot i - b)^2} \tag{12}$$

5. Calculate the average $\bar{F}(L)$ over all segments of the same length.

6. If $\bar{F}(L) \sim L^{\alpha}$, the series is fractional. For fGn, α is the Hurst exponent, H.



Results of DFA applied to the Nile minima. The fit gives $H \simeq 0.93$.

Another approach - ARFIMA

Remember the time shift operator (cf. Lecture 5, Eq. 2): $Bz_n = z_{n-1}$. With this operator, an ARMA(p,q) process can be written as

$$\phi(B)x_n = \sigma\,\theta(B)\eta_n\,, \tag{13}$$

where $\{\eta_n\}$ is GWN, ϕ, θ are polynomials or orders $[, q$, respectively, whose roots lie outside the unit circle. Similarly, an ARIMA(p,d,q) can be written as

$$\phi(B)(1-B)^d x_n = \sigma\,\theta(B)\eta_n\,, \tag{14}$$

where, additionally, $d \in \mathbb{N}$.

A presence of long-range correlations in a time series suggests a need for differentiating to achieve stationarity, but taking a first difference may be too extreme. This motivates the notion of fractional differencing.

Fractionally integrated processes

Therefore, we consider processes of the form (14), but with $-\frac{1}{2} < d < \frac{1}{2}$:

$$\phi(B)(1-B)^d x_n = \sigma\,\theta(B)\eta_n\,, \quad -\frac{1}{2} < d < \frac{1}{2} \tag{15}$$

Such processes are called ARFIMA(p,d,q).

For $d > -1$, the operator $(1-B)^d$ can be defined by the binomial expansion:

$$(1-B)^d = \sum_{j=0}^{\infty} \pi_j B^j \tag{16a}$$

where

$$\pi_0 = 1\,,\ \pi_j = \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} = \prod_{0<k\leqslant j} \frac{k-1-d}{k} \tag{16b}$$

Fractionally integrated white noise

An ARFIMA(0,d,0) is a fractionally integrated white noise

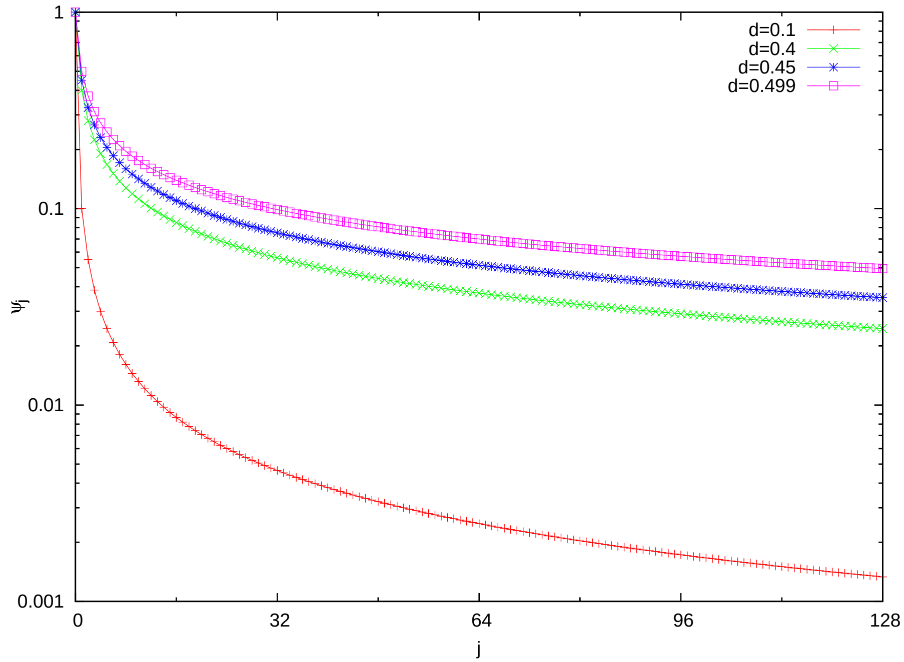
$$(1-B)^d w_n = \sigma\,\eta_n \tag{17a}$$

or

$$w_n = \sigma\,(1-B)^{-d}\eta_n = \sigma\,\sum_{j=0}^{\infty}\psi_j\eta_{n-j} \tag{17b}$$

where

$$\psi_0 = 1\,,\;\psi_j = \prod_{0<k\leqslant j}\frac{k-1+d}{k} \sim \frac{1}{\Gamma(d)}j^{d-1} \text{ as } j\to\infty\,. \tag{17c}$$



Coefficients ψ_j, cf. Eq. (17c), for various values of d. There is no “natural” cutoff, one needs to be taken arbitrarily.

A cutoff

Formally, the summations in (16) and (17) extend to infinity. *In practice*, we need to introduce a cutoff to make the sums finite so that can be handled in a finite time. There are no *definite* results on how large this cutoff must be, but some research suggest that a cutoff of the order ~ 100 in most cases is large enough. Therefore, instead of (16) we take

$$(1 - B)^d = \sum_{j=0}^{100} \pi_j B^j \tag{18}$$

and similarly for (17).

Variance and autocorrelation of the fractionally integrated noise

The variance of this process is

$$\left\langle w_n^2 \right\rangle = \sigma^2 \frac{\Gamma(1-2d)}{[\Gamma(1-d)]^2}\,. \tag{19}$$

and the autocorrelation function

$$\rho_l(w) = \frac{\Gamma(l+d)\Gamma(1-d)}{\Gamma(l-d+1)\Gamma(d)} = \prod_{0<k\leqslant l} \frac{k-1+d}{k-d}\,. \tag{20}$$

A popular interpretation

ARFIMA(p,d,q) can be interpreted as as *ARMA(p,q) driven by a fractionally integrated white noise*:

$$\phi(B)(1-B)^d x_n = \sigma\,\theta(B)\eta_n \tag{21a}$$

therefore

$$\phi(B)x_n = \sigma\,\theta(B)w_n \tag{21b}$$

where

$$(1-B)^d w_n = \eta_n \tag{21c}$$

The power spectrum

From the above representation we can see that the power spectrum of a general ARFIMA(p,d,q) process is

$$P(f) = 2\sigma^2 \left|1 - e^{-2\pi i f}\right|^{-2d} \frac{\left|\theta(e^{-2\pi i f})\right|^2}{\left|\phi(e^{-2\pi i f})\right|^2} \tag{22}$$

where $0 \leqslant f \leqslant \frac{1}{2}$. Note that $P(f) \sim f^{-2d}$ for $f \to 0$ which is a characteristic feature of the spectrum of long memory processes.

ARFIMA(1,d,0)

In theory, ARFIMA(p,d,q) processes of arbitrary orders can be used. In practice, however, processes with small p,q are deemed to be most useful.

Consider ARFIMA(1,d,0):

$$(1-\beta B)(1-B)^d x_n = \sigma\,\eta_n \tag{23a}$$

$$x_n = (1-\beta B)^{-1} w_n \tag{23b}$$

where w_n is (17) and $-1 < \beta < 1$. This process has the autocorrelation function

$$\rho_l = \frac{\rho_l(w)}{1-\beta}\,\frac{F(d+l,1;1-d+l;\beta)+F(d-l,1;1-d-l;\beta)-1}{F(1+d,1;1-d;\beta)} \tag{24}$$

where $F(a,b;c;x)$ is the hypergeometric function.

ARFIMA(0,d,1)

$$(1 - B)^d x_n = \sigma\,(1 - \alpha B)\eta_n\,, \quad -1 < \alpha < 1\,. \tag{25}$$

The autocorrelation function is

$$\rho_l = \rho_l(w)\frac{al^2 - (1-d)^2}{l^2 - (1-d)^2}\,, \tag{26a}$$

$$a = (1-\alpha)^2\left[1 + \alpha^2 - \frac{2\alpha d}{1-d}\right]^{-1}. \tag{26b}$$