Long Range Dependence

Gennady Samorodnitsky, Cornell University
Long range dependence (or long memory) is a property of certain stationary stochastic processes.

It has been associated historically with

- slow decay of correlations

- certain type of scaling that is connected to self-similar processes.

Both of these can be traced back to the papers of B. Mandelbrot and his co-authors from the 60s and 70s, explaining the Hurst(1951) phenomenon.
The annual minima of the water level in the Nile river from the years 622-1281, measured at the Roda gauge near Cairo.
Certain statistics, for example, the so-called \( R/S \)-statistic, grow at an “unusual rate”

\[ n^\theta \quad \text{for} \quad \theta > 1/2 \]

as a function of the sample size for this data set.

The “usual rate” \( n^{1/2} \) comes from the Central Limit Theorem.

Attempts to explain the Hurst phenomenon via heavy tailed models have failed.

Mandelbrot and his co-workers suggested, as an explanation, “unusual behavior”, in the time dimension, rather than in the space dimension: long memory.
Let $X_n$, $n = 0, 1, 2, \ldots$ be a stationary stochastic process with $0 < \sigma^2 = \text{Var} X_0 < \infty$.

Let $\rho_n = \text{Corr}(X_0, X_n)$, $n = 0, 1, \ldots$ be the correlation function.

Consider the partial sum process
\[ S_n = X_1 + \ldots + X_n, \ n \geq 1, \ S_0 = 0, \]

Then its variance is
\[
\text{Var} S_n = \sum_{i=1}^{n} \sum_{j=1}^{n} \text{Cov}(X_i, X_j)
\]
\[
= \sigma^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{|i-j|} = \sigma^2 (n + 2 \sum_{i=1}^{n-1} (n - i) \rho_i). \]
For most of the “usual” stationary stochastic models (ARMA processes, many Markov and Markov modulated processes, etc.) the correlations are absolutely summable

\[ \sum_{n=0}^{\infty} |\rho_n| < \infty. \]

then

\[ \lim_{n \to \infty} \frac{\text{Var} S_n}{n} = \sigma^2 (1 + 2 \sum_{i=1}^{\infty} \rho_i), \]

and so the partial sums appear to grow at the rate \( S_n \sim n^{1/2} \) of the Central Limit Theorem.
On the other hand, one can assume that the correlations are regularly varying:

$$\rho_n \sim n^{-d}L(n), \quad n \to \infty, \quad 0 < d < 1,$$

with a slowly varying function $L$.

Then by the Karamata theorem,

$$\lim_{n \to \infty} \frac{\text{Var} S_n}{n^2 \rho_n} = \frac{2\sigma^2}{(1 - d)(2 - d)}.$$

That is, the partial sums appear to grow at the rate $S_n \sim n^{1-d/2}L(n)^{1/2}$, which is faster than the rate of the Central Limit Theorem.

This is a possible explanation of the Hurst phenomenon.
Such success has led many to use one of the following as a definition of a process with long range dependence:

- **Lack of summability of correlations**
  \[ \sum_{n=0}^{\infty} |\rho_n| = \infty. \]

- **Correlations are regularly varying at infinity with exponent** \(-1 < d \leq 0\). Of course, this assumption implies lack of summability of correlations.
Sometimes the assumption

- **Correlations are regularly varying at infinity with exponent** $d \leq 0$.

This assumption does not imply non-summability of correlations, and it is designed, rather, for contrast with the case of exponentially decaying correlations.

Sometimes it is referred to as medium range dependence.
Another possible angle of viewing long range dependence that is still closely related to correlations is through the spectral domain.

Let $X_n, n = 0, 1, 2, \ldots$ be a stationary stochastic process with a finite variance $\sigma^2$. If its correlations are summable then the process has a spectral density $f$ satisfying

$$\sigma^2 \rho_n = \int_0^\pi \cos(nx)f(x)dx,$$

$n = 0, 1, 2, \ldots$.

Moreover, in this case the spectral density is continuous on $[0, \pi]$. 
It has also been observed that a particular slow decay of correlations, namely

- the correlations are regularly varying at infinity with exponent $0 \leq d < 1$

often goes together with the spectral density ”blowing up” at the origin. In fact, it has been observed that often

- the spectral density is regularly varying at the origin, with exponent $-(1 - d)$
In fact, the equivalence between the regular variation of the correlations at infinity and the regular variation of the spectral density at the origin have become taken for granted, and often stated as a theorem (without proof).

To the best of our knowledge this equivalence is false in general, without extra regularity assumptions.

Below is a rigorous result. Let $R_n = \sigma^2 \rho_n$, $n = 0, 1, 2, \ldots$ be the covariance function a a (weakly) stationary second order process.
Theorem

(i) Assume that

\[ R_n = n^{-d}L(n), \quad n = 0, 1, 2, \ldots, \]

where \( 0 < d < 1 \) and \( L \) is slowly varying at infinity, satisfying the following assumption:

for every \( \delta > 0 \) both functions

\[ g_1(x) = x^\delta L(x) \quad \text{and} \quad g_2(x) = x^{-\delta} L(x) \quad (1) \]

are eventually monotone.

Then the process has a spectral density, say, \( f \), satisfying

\[ f(x) \sim x^{-(1-d)}L(x^{-1}) \frac{2}{\pi} \Gamma(1 - d) \sin \frac{1}{2} \pi d \]

as \( x \to 0 \).
(ii) Conversely, assume that the process has a spectral density $f$ satisfying

$$f(x) = x^{-d} L(x^{-1}), \ 0 < x < \pi,$$

where $0 < d < 1$, and $L$ is slowly varying at infinity, satisfying assumption (1) above.

Suppose, further, that $f$ is of bounded variation on the interval $(\varepsilon, \pi)$ for any $0 < \varepsilon < \pi$.

Then the covariances of the process satisfy

$$R_n \sim n^{-(1-d)} L(n) \Gamma(1 - d) \sin \frac{1}{2} \pi d$$

as $n \to \infty$. 
This point of view on long range dependence is closely related to the idea of self-similarity and scaling.

A stochastic process \((Y(t), t \geq 0)\) is called self-similar with exponent \(H \in \mathbb{R}\) of self-similarity if for all \(c > 0\)

\[
(Y(ct), t \geq 0) \overset{d}{=} c^H (Y(t), t \geq 0)
\]

in the sense of equality of the finite-dimensional distributions.

If the process \((Y(t), t \geq 0)\) also has stationary increments, then this process is often denoted SSSI (self-similar stationary increments).
A few facts about SSSI processes. Let 
\((Y(t), t \geq 0)\) be an SSSI process with exponent 
\(H\) of self-similarity.

- If \(H < 0\) then \(Y(t) = 0\) a.s. for every 
  \(t \geq 0\)

- If \(H = 0\) and \((Y(t), t \geq 0)\) has a measurable modification (in particular, if it is 
  continuous in probability), then for all 
  \(t \geq 0\), \(P(X(t) = X(0)) = 1\)
  (therefore, one assumes that \(H > 0\) when 
  studying SSSI processes)

- If for some \(0 < \gamma < 1\), \(E|Y(1)|^\gamma < \infty\), 
  then \(0 < H < 1/\gamma\)
• Let $E|Y(1)| < \infty$. Then

- $H \leq 1$
- If $H = 1$ then for every $t \geq 1$ we have $X(t) = tX(1)$ with probability 1
- If $0 < H < 1$ then $EY(1) = 0$

In particular, the above conclusions apply to any SSSI process with a finite variance, a specific example of which are Gaussian SSSI processes.
Let \((Y(t), t \geq 0)\) be an SSSI process with exponent \(0 < H < 1\) of self-similarity, and a finite non-zero variance at time 1. Then the process has to have zero mean.

It is elementary to check that scaling and stationary increments imply that the covariance function of the process must have the form

\[
\text{Cov}(Y(s), Y(t)) = \frac{1}{2}[EY(t)^2 + EY(s)^2 - E(Y(t) - Y(s))^2]
\]

\[
= \frac{EY(1)^2}{2}[t^{2H} + s^{2H} - (t - s)^{2H}],
\]

\(0 \leq s \leq t\).
There is a unique up to a scale zero mean SSSI Gaussian process with \(0 < H < 1\). This process is called *Fractional Brownian motion* (FBM).

For \(H = 1/2\) the covariance function given by the right hand side of (2) reduces to
\[
\text{Cov}(Y(s), Y(t)) = EY(1)^2t.
\]
That is, a FBM with \(H = 1/2\) is a Brownian motion.

Let \((Y(t), t \geq 0)\) be a FBM. By the stationarity of the increments of FBM the increment process
\[
X_i = Y(i) - Y(i - 1), \quad i = 1, 2, \ldots
\]
is a stationary. It is commonly referred to as the *Fractional Gaussian Noise* (FGN).
Correlations of the Fractional Gaussian Noise satisfy

\[
\rho_n = \frac{(n + 1)^{2H} + (n - 1)^{2H} - 2n^{2H}}{2}
\]

\[
\sim H(2H - 1)n^{-2(1-H)}
\]

as \( n \to \infty \).

The correlations are

- summable if \( 0 < H \leq 1/2 \)
- non-summable if \( 1/2 < H < 1 \)
Fractional Brownian motion has a spectral density given by

\[ f(x) = C (1 - \cos x) \sum_{j=-\infty}^{\infty} |2\pi j + x|^{-(1+2H)} \]

\[ \sim \frac{C}{2} x^{-(2H-1)} \]

as \( x \to 0 \).

The density is

- continuous if \( 0 < H \leq 1/2 \)
- has a pole at the origin if \( 1/2 < H < 1 \)
Therefore:

- Fractional Gaussian noise with $H > 1/2$ is viewed as having long range dependence

- scaling at the rate higher than $t^{1/2}$ is viewed as an indication of long range dependence in the increments

**However:** Correlations provide only very limited information about the process unless the process is "very close" to being Gaussian.

Everyone agrees that Fractional Gaussian noise with $H > 1/2$ is not long range dependent, and in this case, indeed, correlations tell the entire story.
However

- for processes like ARCH or GARCH processes, or fractionally differenced processes of this kind, correlations are zero in spite of a very rich dependence structure in the process

- it is often difficult to relate correlations to functional of the process that are of real interest

- what to do when the variance is infinite?
It is better to concentrate on functionals of interest, for example: *partial sums, maxima, long strange segments, ruin probabilities*, etc.

Suppose that \((\mathcal{P}_\theta, \theta \in \Theta)\) is a family of laws of a stationary stochastic process \((X_0, X_1, X_2, \ldots)\), where \(\Theta\) is some parameter space.

Assume that the marginal laws of the process do not change much as \(\theta\) varies (perhaps, the marginal laws remain constant, or only the global scale changes, if we are considering, say, Gaussian or S\(\alpha\)S processes).

Suppose we are given a measurable functional \(R = R(X_0, X_1, \ldots)\). The behavior of this functional is different, in general, under different laws \(\mathcal{P}_\theta\).
Suppose that there is a partition of the parameter space $\Theta$ into two parts, $\Theta_0$ and $\Theta_1$, such that the behavior of the functional changes dramatically as one crosses the boundary between $\Theta_0$ and $\Theta_1$.

Then it may make sense to talk about that boundary as a boundary between short range dependence and long range dependence.

That is, the change from short memory to long memory occurs as a phase transition.
Such phase transitions have been observed in many applications, many of which related to large deviations.

For stationary symmetric $\alpha$-stable processes with $0 < \alpha < 2$ the variance is infinite, and the change in memory can be traced to ergodic-theoretical properties of the flow generated the process.

Here we will look at a familiar class of stochastic models and a particular functional of interest.
Let \((X_1, X_2, \ldots)\) be a stationary and ergodic stochastic process with a finite mean \(\mu\).

For a \(\theta > \mu\) we define

\[
R_n(\theta) = \sup \left\{ j - i : 0 \leq i < j \leq n, \frac{X_{i+1} + \ldots + X_j}{j - i} > \theta \right\},
\]

(defined to be equal to zero if the supremum is taken over the empty set).

What happens to \(R_n(\theta)\) as \(n\) increases?
According to the strong law of large numbers for long time intervals \( i + 1, i + 2, \ldots, j \)

\[
\frac{X_{i+1} + \ldots + X_j}{j - i}
\]

should be about the mean \( \mu \).

One would expect that, nonetheless, over some of the many different intervals of this type between 1 and \( n \) for large \( n \) this average may exceed a given \( \theta > \mu \).

The statistic \( R_n(\theta) \) gives the length of the longest interval over which this happens. The law of large numbers appears to break down on such intervals.

Hence: \textit{long strange intervals}.
*Moving average processes*

These are stochastic processes in discrete time defined by

\[ X_n = \mu + \sum_{j=-\infty}^{\infty} \varphi_{n-j} \varepsilon_j, \quad n = 0, 1, \ldots \]

Here \((\varepsilon_n, n = \ldots, -1, 0, 1, 2, \ldots)\) (the noise variables) are iid random variables, and \(\mu\) is a constant.

If the random variables \((\varepsilon_n)\) have a finite mean, we will assume that the mean is equal to zero (it is simply incorporated in the constant \(\mu\)).
We assume that the random variables \((\varepsilon_n)\) have regular varying and balanced tails: for \(\varepsilon = \varepsilon_0\)

\[
P(|\varepsilon| > \lambda) \text{ is regularly varying}
\]

with exponent \(-\alpha < -1\),

and

\[
\lim_{\lambda \to \infty} \frac{P(\varepsilon > \lambda)}{P(|\varepsilon| > \lambda)} = p, \quad \lim_{\lambda \to \infty} \frac{P(\varepsilon < -\lambda)}{P(|\varepsilon| > \lambda)} = q.
\]

Here \(p, q \geq 0\) and \(p + q = 1\).
The following assumptions are sufficient for the series to converge:

If $\alpha > 2$,

$$\sum_{j=-\infty}^{\infty} \varphi_j^2 < \infty,$$

if $1 < \alpha \leq 2$,

$$\sum_{j=-\infty}^{\infty} |\varphi_j|^\alpha \epsilon < \infty$$

for some $\epsilon > 0$. 

Here the boundary between short and long memory appears to be that

$$\sum_{j=-\infty}^{\infty} |\varphi_j| < \infty \quad \text{and} \quad = \infty.$$  

Indeed, if $\sum_{j=-\infty}^{\infty} |\varphi_j| < \infty$, then

$$\frac{R_n(\theta)}{a_n} \Rightarrow M(\theta),$$

a non-degenerate limit. Here

$$a_n = \left(\frac{1}{F}\right)^{(n)},$$

where $F$ is the distribution function of $|\varepsilon|$.

That is, long strange segments have length of about $n^{1/\alpha}$. 
On the other hand, assume that the coefficients \((\varphi_j)\) are regularly varying and balanced. That is, there is a function \(\varphi : [0, \infty) \to [0, \infty)\) such that

\[
\varphi(t) = L_2(t) \, t^{-h}
\]
as \(t \to \infty\) and such that

\[
\lim_{j \to \infty} \frac{\varphi_j}{\varphi(j)} = \xi_+ , \quad \lim_{j \to \infty} \frac{\varphi_{-j}}{\varphi(j)} = \xi_-
\]
for some \(\xi_+, \xi_- \geq 0\), at least one of which is positive.

Here

\[
h > \max\left\{\frac{1}{\alpha}, \frac{1}{2}\right\}
\]
and \(L_2\) is a slowly varying function.
Then

\[ \frac{R_n(\theta)}{b_n} \Rightarrow M_h(\theta), \]

a non-degenerate limit. Here

\[ b_n = \left( \frac{1}{\varphi} \right)^\leftarrow (a_n). \]

That is, long strange segments have length of about \( n^{1/\alpha h} \).

**Note:** in the long memory case the rate of growth of a functional of interest depends on the parameters.