Foundations and Trends® in
Stochastic Systems
Vol. 1, No. 3 (2006) 163–257

DOI: 10.1561/0900000004

now
the essence of knowledge

Long Range Dependence

Gennady Samorodnitsky

School of Operations Research and Information Engineering and Department of Statistical Science, Cornell University, Ithaca, NY 14853, USA, gennady@orie.cornell.edu

Abstract

The notion of long range dependence is discussed from a variety of points of view, and a new approach is suggested. A number of related topics is also discussed, including connections with non-stationary processes, with ergodic theory, self-similar processes and fractionally differenced processes, heavy tails and light tails, limit theorems and large deviations.

Keywords: Long range dependence, rare events, large deviations.

Mathematical Subject Classification:
Primary 60G10, 62M10; Secondary 60F10

1

Introduction

Long range dependence and *long memory* are synonymous notions, that are arguably very important. This importance can be judged, for example, by the very large number of publications having one of these notions in the title, in areas such as finance [84], econometrics [115], internet modeling [70], hydrology [109], climate studies [142], linguistics [3] or DNA sequencing [71]. These publications address a great variety of issues: detection of long memory in the data, statistical estimation of parameters of long range dependence, limit theorems under long range dependence, simulation of long memory processes, and many others. Surprisingly, very few of these publications address what long range dependence is. When definitions are given, they vary from author to author (the econometric survey [58] mentions 11 different definitions). The notion of long range dependence can also be applied to different aspects of a given stochastic process [63]. More diverse definitions become possible if, instead of looking at the "usual" stationary processes, one studies *stationary point processes*, as in [37], or *random fields*, as in [4].

It is the purpose of this survey to discuss what is meant (often implicitly) by long range dependence, clarify why this notion is

important, mention different point of views on the topic and, hopefully, remove some of the mystery that surrounds it.

The notion of long range dependence has, clearly, something to do with memory in a stochastic process. Memory is, by definition, something that lasts. It is the requirement that the memory has to be "long" that is special. Why is it important that in one model the memory is "a bit longer" than in another model? The first serious argument that this can be important is in a series of papers of B. Mandelbrot and his coauthors, e.g. [89] and [93]. It is also due to the influence of these early papers and subsequent publications of Mandelbrot (especially [90]) that long range dependence has also become associated with scaling and fractal behavior. We survey some of the early history in Section 2.

The "specialness" of long memory indicates that most stationary stochastic processes do not have it. This also makes it intuitive that non-stationary processes can provide an alternative explanation to the empirical phenomena that the notion of long range dependence is designed to address. This connection between long memory and lack of stationarity is very important. It is related to such well known phenomena as unit root problem [111] and regime switching [42]. We discuss the connections with non-stationary processes in Section 3.

A very attractive point of view on long range dependence is based on ergodic-theoretical properties of the dynamical system on which a stationary stochastic process is constructed. Many features that are intuitively associated with long memory are automatically found in such an approach. For several reasons this approach has not become widely accepted. We discuss this in Section 4.

Most of the definitions of long range dependence appearing in literature are based on the second-order properties of a stochastic process. Such properties include asymptotic behavior of covariances, spectral density, and variances of partial sums. The reasons for popularity of the second-order properties in this context are both historical and practical: second-order properties are relatively simple conceptually and easy to estimate from the data. This approach to the notion of long memory is discussed in Section 5.

The term "fractional" appears very frequently in the context of long range dependence. This usually refers to a model constructed using a

generalized operation of a non-integer order, whereas the "usual" order of the operation has to be integer. The examples include differencing or differentiation "non-integral number of times." Certain features often associated with long memory can sometimes be obtained by doing so. Models obtained in this way are discussed in Section 6.

It is, once again, largely due to the early history that the notion of long range dependence has become closely associated with self-similar processes. Self-similar processes are stochastic models with the property that a scaling in time is equivalent to an appropriate scaling in space. The connection between the two types of scaling is determined by a constant often called *the Hurst exponent*, and it has been argued that the value of this exponent determines whether or not the increments of a self-similar process with stationary increments possess long range dependence. We discuss self-similar processes in Section 7.

The final part of this survey, Section 8, introduces a different approach to understanding long memory, a one that is related to the notion of phase transitions. We argue that this approach makes the notion of long range dependence both intuitive and practical. One should hope for major future research effort in this direction.

2

Some History: The Hurst Phenomenon

The history of long range dependence as a concrete phenomenon believed to be important in its own right should be regarded as beginning in the 1960s with a series of papers of Benoit Mandelbrot and his co-workers, even though even earlier empirical findings had occurred. The cause was a need to explain an empirical phenomenon observed by Hurst [67, 68] who studied the flow of water in the Nile river. A particular data set Hurst looked at appears on Figure 2.1. There are many things that are interesting about this data set (one of which is how far back in time the data go). Harold Hurst, who was interested in dam design, looked at this data through a particular statistic. Given a sequence of n observations $X_1, X_2, \ldots, X_n$, define the partial sum sequence $S_m = X_1 + \cdots + X_m$ for $m = 0, 1, \ldots$ (with $S_0 = 0$). The statistics Hurst looked at is

$$\frac{R}{S}(X_1, \ldots, X_n) = \frac{\max_{0 \le i \le n}(S_i - \frac{i}{n}S_n) - \min_{0 \le i \le n}(S_i - \frac{i}{n}S_n)}{(\frac{1}{n}\sum_{i=1}^{n}(X_i - \frac{1}{n}S_n)^2)^{1/2}}. \quad (2.1)$$

Note that S_n/n is the sample mean of the data. Therefore, $\max_{0 \le i \le n}(S_i - \frac{i}{n}S_n)$, for example, measures how far the partial sums get above the straight line they would follow if all observations were

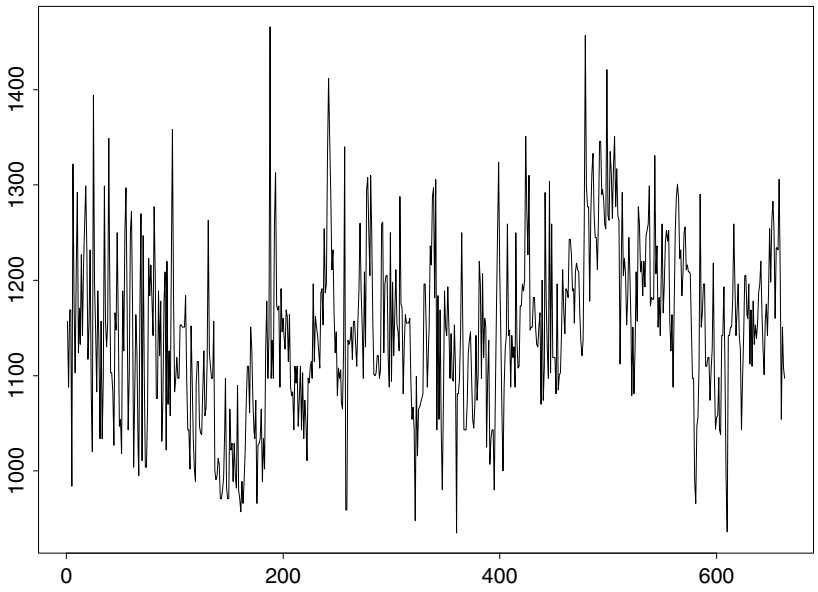

Fig. 2.1 Annual minima of the water level in the Nile river for the years 622–1281, measured at the Roda gauge near Cairo.

equal (to the sample mean), and the difference between the maximum and the minimum of the numerator in (2.1) is the difference between the highest and lowest positions of the partial sums with respect to the straight line of uniform growth. It is referred to as the range of the observations. The denominator of (2.1) is, of course, the sample standard deviation. The entire statistic in (2.1) has, then, been called *the rescaled range* or *the R/S statistic*.

Suppose now that $X_1, X_2, \ldots$ is a sequence of random variables. One can apply the R/S statistic to the first n observations $X_1, X_2, \ldots, X_n$ for increasing values of n. What would one expect the resulting sequence of values of the R/S statistic to be like? Let us do some simple calculations.

Consider the space $D[0,1]$ of right continuous and having left limits functions on $[0,1]$ equipped with the Skorohod J_1 topology [18]. The function $f : D[0,1] \to \mathbb{R}$ defined by

$$f(\mathbf{x}) = \sup_{0 \le t \le 1} (x(t) - tx(1)) - \inf_{0 \le t \le 1} (x(t) - tx(1)),$$

$\mathbf{x} = (x(t), 0 \le t \le 1) \in D[0,1]$, is easily seen to be continuous. We would like to apply this function to the $D[0,1]$-valued version of the partial sum sequence, the so-called partial sum process.

Suppose that $X_1, X_2, \ldots$ is, in fact, a stationary sequence of random variables with a finite variance, and a common mean μ. The partial sum process is defined by

$$S^{(n)}(t) = S_{[nt]} - [nt]\mu, \quad 0 \le t \le 1\,. \tag{2.2}$$

The classical Functional Central Limit Theorem (Donsker's Theorem, invariance principle) says that, if $X_1, X_2, \ldots$ are i.i.d., then

$$\frac{1}{\sqrt{n}} S^{(n)} \Rightarrow \sigma_* B \quad \text{weakly in } D[0,1], \tag{2.3}$$

where σ_*^2 is equal to the common variance σ^2 of the observations, and B is the standard Brownian motion on $[0,1]$ (Theorem 14.1 in [18]). In fact, the Functional Central Limit Theorem is known to hold for stationary processes with a finite variance that are much more general than an i.i.d. sequence (with the limiting standard deviation σ_* not equal, in general, to the standard deviation of the X_is); see a recent survey in [96].

It is straightforward to check that the range of the first n observations (the numerator in the R/S statistic) is equal to $f(S^{(n)})$. Therefore, if the invariance principle (2.3) holds, then by the continuous mapping theorem,

$$\frac{1}{\sqrt{n}}(\text{the range of the first } n \text{ observations}) = f\left(\frac{1}{\sqrt{n}} S^{(n)}\right)$$

$$\Rightarrow f(\sigma_* B) \;=\; \sigma_* \left[\sup_{0 \le t \le 1} (B(t) - tB(1)) - \inf_{0 \le t \le 1} (B(t) - tB(1)) \right]$$

$$:= \sigma_* \left[\sup_{0 \le t \le 1} B_0(t) - \inf_{0 \le t \le 1} B_0(t) \right],$$

where B_0 is a Brownian bridge on $[0,1]$. Furthermore, if the stationary sequence $X_1, X_2, \ldots$ is ergodic, then the sample standard deviation is a consistent estimator of the population standard deviation, and so

$$\left(\frac{1}{n} \sum_{i=1}^{n} \left(X_i - \frac{1}{n} S_n \right)^2 \right)^{1/2} \to \sigma \quad \text{with probability 1.}$$

Under these circumstances we see that

$$\frac{1}{\sqrt{n}}\frac{R}{S}(X_1,\ldots,X_n) \Rightarrow \frac{\sigma_*}{\sigma}\left[\sup_{0\le t\le 1} B_0(t) - \inf_{0\le t\le 1} B_0(t)\right]. \tag{2.4}$$

That is, the R/S statistic grows as the square root of the sample size.

When Harold Hurst calculated the R/S statistic on the Nile river data on Figure 2.1 he found, however, the empirical rate of growth closer to $n^{.74}$ (with n being the number of observations). This phenomenon became known as *the Hurst phenomenon*, and finding a stochastic model that would explain it proved to be tricky. The assumptions made above to guarantee the convergence in (2.4) are reasonably mild, and one would expect that even if exact convergence in (2.4) was difficult to establish, the square root of the sample size was still the order of magnitude of the R/S statistic. A drastic departure from the assumptions was needed.

One such departure was suggested by Moran in [103], and it consisted of dropping the assumption of a finite variance of the observations $X_1, X_2, \ldots$ and assuming, instead, that the observations are in the domain of attraction of an infinite variance α-stable distribution with $0 < \alpha < 2$ (Moran, actually, assumed the observations to have a symmetric α-stable law). It was pointed out in [91], however, that the self-normalizing feature of the R/S statistic prevents infinite variance alone from explaining the Hurst phenomenon. Let us sketch why.

We will assume, for simplicity, that the observations (in addition to being i.i.d.) have *balanced power tails*:

$$\begin{gathered} P(|X_1| > x) \sim cx^{-\alpha} \quad \text{as } x \to \infty, \text{ and} \\ \lim_{x\to\infty} \frac{P(X_1 > x)}{P(|X_1| > x)} = p, \quad \lim_{x\to\infty} \frac{P(X_1 < -x)}{P(|X_1| > x)} = q \end{gathered} \tag{2.5}$$

for some $0 < \alpha < 2$, $c > 0$ and $0 \le p, q \le 1$, $p + q = 1$. The general domain of attraction assumption allows a slowly varying function in the tail of $|X_1|$ [53]. The greater generality may introduce a slowly varying (in the sample size) function in the order of magnitude of the R/S statistic, but it cannot change the crucial exponent of n in that order of magnitude.

Assuming (2.5), one can use a very important point process convergence result (from which a large number of other *heavy tailed* limit theorems follow). For $n \geq 1$ we define

$$N_n = \sum_{j=1}^{n} \delta_{(j/n, X_j/n^{1/\alpha})}$$

and view N_n as a point process on $[0,1] \times ([-\infty,\infty]\backslash\{0\})$. Here $\delta_{(t,x)}$ is the point mass at a point (t,x). Then

$$N_n \Rightarrow N := \sum_{j=1}^{\infty} \delta_{(U_j, J_j)} \tag{2.6}$$

weakly in the space of Radon discrete random measures on the space $[0,1] \times ([-\infty,\infty] \setminus \{0\})$, where the Radon property means that a measure assigns finite values to sets bounded away from the origin. In the right-hand side of (2.6), (J_j) are the points of a Poisson process on $\mathbb{R}$ with mean measure μ given by

$$m\Big((x,\infty)\Big) = cpx^{-\alpha}, \quad m\Big((-\infty,-x)\Big) = cqx^{-\alpha}$$

for $x > 0$, while (U_j) are i.i.d. standard uniform random variables independent of the Poisson process. The space of Radon discrete random measures is endowed with the topology of vague convergence. The result is (4.70) in [113], which can also be consulted for technical details.

It is possible to apply (2.6) to understand the "size" of the R/S statistic, starting with a typical "truncation" step, needed because various sums of points are not continuous functionals of point processes in the topology of vague convergence (but see [38] for recent progress toward topologies that may make certain sums of the points continuous functionals). Using Theorem 3.2 in [18] and verifying certain technical conditions, one obtains

$$\frac{1}{\sqrt{n}} \frac{R}{S}(X_1, \ldots, X_n) \Rightarrow g(N), \tag{2.7}$$

where

$$g(N) = \frac{\sup_{0\leq t\leq 1} \sum_{j=1}^{\infty} \big(\mathbf{1}(U_j \leq t) - t\big) J_j - \inf_{0\leq t\leq 1} \sum_{j=1}^{\infty} \big(\mathbf{1}(U_j \leq t) - t\big) J_j}{(\sum_{j=1}^{\infty} J_j^2)^{1/2}};$$

we omit the details. Note that (2.7) means that even in the heavy tailed case the R/S statistic grows as the square root of the sample size.

We conclude, therefore, as did Mandelbort and Taqqu [91], that infinite variance alone cannot explain the Hurst phenomenon. A different drastic departure from the assumptions leading to the square root of the sample size rate of growth of the R/S statistic was suggested in [89] (see also [93]), and it had nothing to do with heavy tails. The idea was, instead, to take as a model a stationary process with a finite variance, but with correlations decaying so slowly as to invalidate the Functional Central Limit Theorem (2.3). The simplest model of that sort is the *Fractional Gaussian Noise*.

Let us start with a zero mean Gaussian process $(B_H(t), t \geq 0)$ satisfying $B_H(0) = 0$ and $E(B_H(t) - B_H(s))^2 = \sigma^2|t-s|^{2H}$ for some $\sigma > 0$ and $0 < H \leq 1$. We will see below that such a process does exist, and it has stationary increments (that is, the law of $(B_H(t+h) - B_H(h), t \geq 0)$ does not depend on $h \geq 0$). It is called a *Fractional Brownian motion*, or *FBM*, and it becomes the usual Brownian motion when $H = 1/2$. Clearly, this process has the self-similarity property $(B_H(ct), t \geq 0) \stackrel{\mathrm{d}}{=} (c^H B_H(t), t \geq 0)$ for any $c > 0$.

The power-like behavior of the incremental variance immediately allows one to check the metric entropy condition [47] or the Kolmogorov criterion [48] to conclude that a Fractional Brownian motion has a continuous version, and we always assume that we are working with such a version. Furthermore, an easy computation of the covariance function shows that for $H = 1$, $B_1(t) = tB_1(1)$ with probability 1 for each $t \geq 0$, and so to avoid trivialities we always take $0 < H < 1$.

A Fractional Gaussian Noise, or *FGN*, is a discrete step increment process of a Fractional Brownian motion defined by $X_j = B_H(j) - B_H(j-1)$ for $j = 1, 2, \ldots$. The stationarity of the increments of the FBM implies that this is a stationary Gaussian process. Using the fact $ab = (a^2 + b^2 - (a-b)^2)/2$ and the incremental variance of the FBM, we easily see that

$$\mathrm{Cov}(X_{j+n}, X_j) = \frac{\sigma^2}{2}\left[(n+1)^{2H} + |n-1|^{2H} - 2n^{2H}\right] \tag{2.8}$$

for $j \geq 1, n \geq 0$. That is,

$$\rho_n := \mathrm{Corr}(X_{j+n}, X_j) \sim H(2H-1)n^{-2(1-H)} \quad \text{as } n \to \infty. \tag{2.9}$$

In particular, $\rho_n \to 0$ as $n \to \infty$. This implies that the FGN is a mixing, hence ergodic process; see [34]. Furthermore, by the self-similarity of the FBM, for every n

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}B_H(n) = \sigma^2 n^{2H}. \tag{2.10}$$

Suppose now that a set of observations $X_1, X_2, \ldots$ forms a Fractional Gaussian Noise as defined above, and let us consider the behavior of the R/S statistic on these observations. The ergodicity of the FGN implies that the denominator of the statistic converges a.s. to the standard deviation of the observations, σ. For the numerator of the R/S statistic we recall that $S_i = B_H(i)$ for every i, and the self-similarity of the FBM gives us

$$\begin{aligned}
&\max_{0\leq i\leq n}\left(S_i - \frac{i}{n}S_n\right) - \min_{0\leq i\leq n}\left(S_i - \frac{i}{n}S_n\right) \\
&\quad = \max_{0\leq i\leq n}\left(B_H(i) - \frac{i}{n}B_H(n)\right) - \min_{0\leq i\leq n}\left(B_H(i) - \frac{i}{n}B_H(n)\right) \\
&\quad \stackrel{\mathrm{d}}{=} n^H\left[\max_{0\leq i\leq n}\left(B_H\left(\frac{i}{n}\right) - \frac{i}{n}B_H(1)\right)\right. \\
&\qquad \left. - \min_{0\leq i\leq n}\left(B_H\left(\frac{i}{n}\right) - \frac{i}{n}B_H(1)\right)\right].
\end{aligned}$$

By the continuity of the sample paths of the FBM we have

$$\begin{aligned}
&\max_{0\leq i\leq n}\left(B_H\left(\frac{i}{n}\right) - \frac{i}{n}B_H(1)\right) - \min_{0\leq i\leq n}\left(B_H\left(\frac{i}{n}\right) - \frac{i}{n}B_H(1)\right) \\
&\quad \to \sup_{0\leq t\leq 1}(B_H(t) - tB_H(1)) - \inf_{0\leq t\leq 1}(B_H(t) - tB_H(1))
\end{aligned}$$

with probability 1. That is, for the FGN,

$$\begin{aligned}
n^{-H}\frac{R}{S}(X_1, \ldots, X_n) \Rightarrow \frac{1}{\sigma}\Bigg[&\sup_{0\leq t\leq 1}(B_H(t) - tB_H(1)) \\
&- \inf_{0\leq t\leq 1}(B_H(t) - tB_H(1))\Bigg],
\end{aligned}$$

and so the R/S statistic grows at the rate n^H as a function of the sample size. Therefore, selecting an appropriate H in the model will, finally, explain the Hurst phenomenon. In particular, the parameter H of Fractional Brownian motion is often referred to as *Hurst parameter*.

This success of the Fractional Gaussian Noise model was, and still is, striking. Of course, self-similarity of the FBM was used in the above computation, but it was quickly realized that the really important fact was the unusually slow decay of correlations in (2.9), especially for high values of H (i.e., close to 1). For these values of H the variance of the partial sums in (2.10) also increases unusually fast. Unlike the previous unsuccessful attempt to explain the Hurst phenomenon by introducing in the model unusually heavy tails (infinite variance in this case), the FGN model succeeds here by introducing unusually long memory. Particularly vivid terminology was introduced in [93], in the context of weather and precipitation: unusually heavy tails were designated as *Noah effect*, referring to the biblical story of Noah and extreme incidents of precipitation, while unusually long memory was designated as *Joseph effect*, referring to the biblical story of Joseph and long stretches (seven years) of time higher than average and lower than average precipitation. One can visually see the domination of extreme observations in the left plot of Figure 2.2, where the observations are Pareto random variables with parameter 1 (and so even fail to have a finite

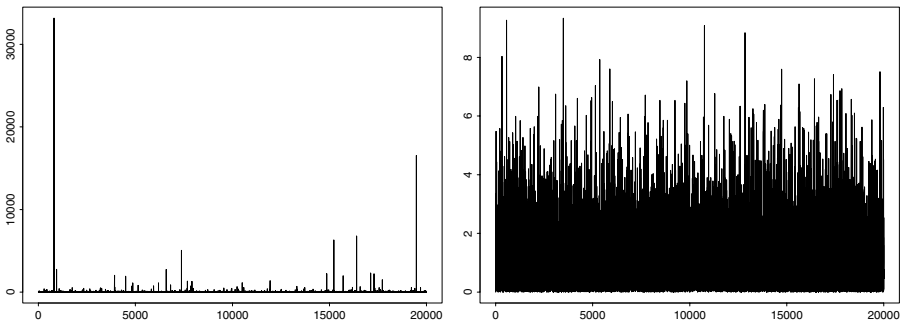

Fig. 2.2 i.i.d. Pareto random variables with parameter 1 (left plot) and i.i.d. exponential random variables with parameter 1 (right plot).

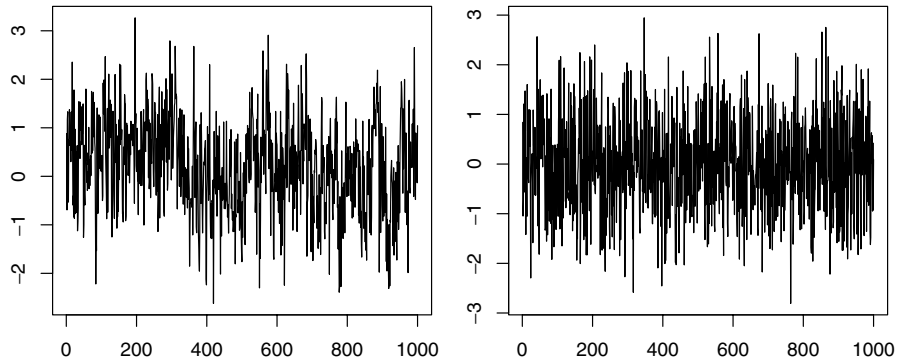

Fig. 2.3 Fractional Gaussian noise with $H = 0.8$ (left plot) and i.i.d. standard Gaussian random variables (right plot).

mean), as opposed to a much less pronounced domination of extreme observations in the right plot of Figure 2.2, where the observations are standard exponential random variables.

Joseph effect, on the other hand, is clearly visible on Figure 2.3: in the left plot, where the observations form FGN with Hurst parameter $H = 0.8$, there are long stretches of time (hundreds of observations) where the observations tend to be on one side of the true mean 0. This is, clearly, not the case on the right plot of i.i.d. normal observations. Returning momentarily to the Nile river data on Figure 2.1 we see evidence of Joseph effect there as well.

This brought the fact that memory of a certain length can make a big difference to the attention of many. The terms "long range dependent process" and "long memory" came into being; they can already be found in the early papers by Mandelbrot and co-authors. A number of surveys throughout the years helped to maintain clarity in this otherwise mysterious subject; we mention Cox [35], Rosenblatt [118], a bibliographic guide of Taqqu [139], and the monograph of Beran [14].

3

Long Memory and Non-Stationarity

It is standard in theory of stochastic processes to apply the notion of long range dependence only to stationary processes. This is not necessarily the case in certain areas of application (such as, for example, physics, or self-organizing criticality), where closely related terms (long-term correlations, $1/f$ noise, often in the context of power laws) are sometimes applied to non-stationary models, such as Brownian motion. See e.g., [15]. Because of that, it has also been suggested, e.g., in [64], to modify certain (second-order) definitions of long range dependence to apply to non-stationary processes as well.

In general, the relationship between long range dependence and non-stationarity is delicate in a number of ways. We have seen that the *Joseph effect* involves long stretches of time when the process tends to be above the mean, and long stretches of time when the process tends to be below the mean. This and related phenomena can, of course, be taken to indicate non-stationarity. Quoting a description in [90, p. 251], of a Fractional Gaussian noise with $H > 1/2$: "Nearly every sample looks like a "random noise" superimposed upon a background that performs several cycles, whichever the sample's duration. However, these cycles are *not* periodic, that is, *cannot* be extrapolated as the

sample lengthens." In application to real data, either stationary long memory models or appropriate non-stationary models can and have been used. There is, obviously, no "right" or "wrong" way to go here, beyond the principle of parsimony.

Among the first to demonstrate the difficulty of distinguishing between stationary long memory models and certain non-stationary models was Bhattacharya et al. [17] who suggested that, instead of Fractional Gaussian noise or another model with long memory, the Hurst phenomenon can be explained by a simple non-stationary model as follows. Let $Y_1, Y_2, \ldots$ be a sequence of independent identically distributed random variables with a finite variance σ^2. Let $0 < \beta < 1/2$, choose $a \geq 0$ and consider the model

$$X_i = Y_i + (a+i)^{-\beta}, \quad i = 1, 2, \ldots . \tag{3.1}$$

Clearly, the stochastic process $X_1, X_2, \ldots$ is non-stationary, for it contains a non-trivial drift. However, it is asymptotically stationary (as the time increases), and the drift can be taken to be very small to start with (by taking a to be large). This process has no memory at all, as the sequence $Y_1, Y_2, \ldots$ is i.i.d. It does, however, cause the R/S statistic to behave in the same way as if the sequence $X_1, X_2, \ldots$ were an FGN, or another long range dependent process. To see why, assume for simplicity that $a = 0$ above, and note that for this model, the numerator of the R/S statistic is bounded between

$$r_n - R_n^Y \leq \max_{0 \leq i \leq n} \left(S_i - \frac{i}{n} S_n \right) - \min_{0 \leq i \leq n} \left(S_i - \frac{i}{n} S_n \right) \leq r_n + R_n^Y ,$$

where

$$r_n = \max_{0 \leq i \leq n} \left(s_i - \frac{i}{n} s_n \right) - \min_{0 \leq i \leq n} \left(s_i - \frac{i}{n} s_n \right) ,$$

$$R_n^Y = \max_{0 \leq i \leq n} \left(S_i^Y - \frac{i}{n} S_n^Y \right) - \min_{0 \leq i \leq n} \left(S_i^Y - \frac{i}{n} S_n^Y \right) ,$$

and $S_m^Y = Y_1 + \cdots + Y_m$, $s_m = \sum_{j=1}^m j^{-\beta}$ for $m = 0, 1, 2, \ldots$. Since s_m is a sum of a decreasing sequence of numbers, we see that $\min_{0 \leq i \leq n}(s_i - \frac{i}{n} s_n) = 0$. On the other hand, the extremum in the max part of r_n is achieved at $i = \lfloor (\frac{1}{n} \sum_{j=1}^n j^{-\beta})^{-1/\beta} \rceil$, and elementary computations

show that $\max_{0\le i\le n}(s_i - \frac{i}{n}s_n) \sim C_\beta n^{1-\beta}$ with $C_\beta = \beta(1-\beta)^{1/\beta-2}$. Since R_n^Y grows as $n^{1/2}$, we immediately conclude that

$$\frac{1}{n^{1-\beta}}\left[\max_{0\le i\le n}\left(S_i - \frac{i}{n}S_n\right) - \min_{0\le i\le n}\left(S_i - \frac{i}{n}S_n\right)\right] \to C_\beta$$

in probability as $n\to\infty$. Similarly, in the denominator of the R/S statistic we have a bound

$$D_n^Y - d_n \le \left(\sum_{i=1}^n \left(X_i - \frac{1}{n}S_n\right)^2\right)^{1/2} \le D_n^Y + d_n\,,$$

where

$$D_n^Y = \left(\sum_{i=1}^n \left(Y_i - \frac{1}{n}S_n^Y\right)^2\right)^{1/2}, \quad d_n = \left(\sum_{i=1}^n \left(i^{-\beta} - \frac{1}{n}s_n\right)^2\right)^{1/2}.$$

We know that $D_n^Y/n^{1/2} \to \sigma$ a.s. as $n\to\infty$, while an elementary computation leads to $d_n/n^{1/2-1/\beta} \to C'_\beta$ with $C'_\beta = \beta^2(1-\beta)^{-2}(1-2\beta)$. Therefore,

$$n^{-1/2}\left(\sum_{i=1}^n \left(X_i - \frac{1}{n}S_n\right)^2\right)^{1/2} \to \sigma$$

a.s. and we conclude that

$$\frac{1}{n^{1-\beta}}\frac{R}{S}(X_1,\ldots,X_n) \to \frac{C_\beta}{\sigma}$$

in probability as $n\to\infty$. Therefore, for the model (3.1) the R/S statistic grows as $n^{1-\beta}$, same rate as for the FGN with $H = 1-\beta$, and so the R/S statistic cannot distinguish between these two models. Apart from fooling the R/S statistic, however, the model (3.1) is not difficult to tell apart from a stationary process with correlations decaying as in (2.9). That this can be done using *the periodogram* was quickly shown in [79].

A very important class of non-stationary models that empirically resemble long memory stationary models is that of *regime switching models*. The name is descriptive, and makes it clear where the lack of stationarity comes from. The Fractional Gaussian noise also appears to

exhibit different "regimes" (the Joseph effect), but the non-stationary regime switching models are usually those with break points, whose location changes with the sample size, in either random or non-random manner.

One class of regime switching models obtains by taking a parametric model that would be stationary if its parameters were kept constant and then changing the parameters along a sequence of non-random time points, again chosen relatively to the sample size. In [99] and [100] such a procedure was applied to the GARCH(p,q) model. Such a change can affect the mean and the variance (among many other things) of the process after break points, and to many sample statistics this will look like long memory.

To see what might happen here consider a sample $X_1,\dots,X_n$, where the observations come from r subsamples of lengths proportional to the overall sample size. That is, given fixed proportions $0<p_i<1$, $i=1,\dots,r$ with $p_1+\cdots+p_r=1$, the sample has the form:

$$X_1^{(1)},\dots,X_{[np_1]}^{(1)},X_{[np_1]+1}^{(2)},\dots,X_{[n(p_1+p_2)]}^{(2)},\dots,X_{[n(1-p_r)]}^{(r)},\dots,X_n^{(r)}, \quad (3.2)$$

where the ith subsample forms a stationary ergodic process with a finite variance, $i=1,\dots,r$. Since one of the common ways to try to detect long range dependence is by looking for a slow decay of covariances and correlations, let us check the behavior of the sample covariance on the sample (3.2). Note that for a fixed time lag m

$$\hat{R}_m(n)=\frac{1}{n}\sum_{j=1}^{n-m}(X_j-\bar{X})(X_{j+m}-\bar{X})=A_m(n)+B_m(n),$$

where $\bar{X}$ is the overall sample mean,

$$A_m(n)=\frac{1}{n}\sum_{j=1}^{n-m}X_jX_{j+m}-(\bar{X})^2,$$

and

$$B_m(n)=\frac{1}{n}\bar{X}\left(\sum_{j=1}^{m}X_j+\sum_{j=n-m+1}^{n}X_j\right)-\frac{m}{n}(\bar{X})^2.$$

Obviously $B_m(n) \to 0$ in probability as $n \to \infty$. By ergodicity, also $\bar{X} \to \sum_{i=1}^{r} p_i \mu_i$, where μ_i is the mean of the ith subsample. Finally, if I_i denotes the set of indices corresponding to the ith subsample, $i = 1, \ldots, r$, then by the same ergodicity,

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n-m} X_j X_{j+m} &= \sum_{i=1}^{r} \frac{\mathrm{Card}(I_i \cap (I_i - m))}{n} \frac{1}{\mathrm{Card}(I_i \cap (I_i - m))} \\ &\quad \times \sum_{j \in I_i \cap (I_i - m)} X_j^{(i)} X_{j+m}^{(i)} + \frac{1}{n}\sum_{i=1}^{r} \sum_{\substack{j \in \{1,\ldots,n-m\} \\ j \in I_i, j+m \in I_{i+1}}} X_j X_{j+m} \\ &\to \sum_{i=1}^{r} p_i (R_m^{(i)} + \mu_i^2),\end{aligned}$$

where $R_m^{(i)}$ is the covariance at lag m of the ith subsample. We conclude that

$$\begin{aligned}\hat{R}_m(n) &\to \sum_{i=1}^{r} p_i \big(R_m^{(i)} + \mu_i^2\big) - \left(\sum_{i=1}^{r} p_i \mu_i\right)^2 \\ &= \sum_{i=1}^{r} p_i R_m^{(i)} + \sum_{i_1=1}^{r} \sum_{i_2=i_1+1}^{r} p_{i_1} p_{i_2} \left(\mu_{i_1} - \mu_{i_2}\right)^2\end{aligned} \tag{3.3}$$

in probability as $n \to \infty$. What (3.3) indicates is that, if there is regime switching as we have described, and (some of) the mean values in different regimes are different, then the estimated from the sample covariance function will tend to stabilize, at large lags, at a positive value. This is what often observed in practice and long memory is suspected. Of course, this regime switching model is simply a deterministic way of mimicking Joseph effect (recall Figure 2.3). Various other regime switching models mimicking long range dependence are suggested in [42].

Overall, stationary long memory models have become more popular than regime switching models. An important reason for this is, undoubtedly, parsimony. Statistical goodness of fit should be and has been taken into account as well. For example, the stationary Fractional Gaussian noise fits the Nile river data very well (see [14, Chapter 10]). On the other hand, certain workload data in computer networks, often

modeled as long memory processes, can be well fit by nonstationary ARMA$(p, 1, q)$ models [131].

More generally, ARMA models (otherwise known as linear models in time series) provide further connections between stationary long memory processes and non-stationary models. A linear model is described by two functions applied to the backshift operator, the autoregressive function and moving average functions (both often polynomials); we refer the reader to [26] for details. Stationarity of the model is easier to achieve if the autoregressive function does not vanish on the unit circle in the complex plane; if the order of a root there is at least 1, stationarity is impossible — this is the so called unit root situation. On the other hand, roots of certain fractional orders allow for stationary, long memory models. These fractional models will be considered in Section 6 below. Distinguishing between non-stationary unit root models and stationary fractional models is an important problem in econometrics; see [13].

It is possible to summarize the discussion of long memory and non-stationarity by saying that the stationary long memory processes form a layer among the stationary processes that is "near the boundary" with non-stationary processes, or, alternatively, as the layer separating the non-stationary processes from the "usual" stationary processes [125]. The processes in the "layer" resemble non-stationary models (the Joseph effect), and they are unusual stationary processes to such an extent that one can talk about a phase transition. This is discussed in detail in Section 8.

4

Long Memory, Ergodic Theory, and Strong Mixing

The notion of memory in a stationary stochastic process is by definition, related to the connections between certain observations and those occurring after an amount of time has passed. If $X_1, X_2, \ldots$ is the process then the passage of time corresponds to a *shifted process*: $X_{k+1}, X_{k+2}, \ldots$, for a time shift k. In other words, the notion of memory is related to the connections between the process and its shifts. Since the process is stationary, the shifts do not change the distribution of the process. This makes various notions from ergodic theory (of measure preserving transformations on measure spaces) a very attractive language in describing memory of a stationary process. We refer the reader to [1] and [78] for the ergodic theoretical notions used in this survey.

It is convenient (but not necessary) to assume that the sample space associated with a stationary process is a space of sequences, on which shifts are naturally defined. It is even more convenient (even though, once again, not necessary) to take the sample space to be the space of two-sided sequences $\mathbf{x} = (\ldots, x_{-1}, x_0, x_1, x_2, \ldots)$ since the shifts are invertible on such spaces. Let us, therefore, assume in this section that a stationary process $\mathbf{X} = (\ldots, X_{-1}, X_0, X_1, X_2, \ldots)$ is defined as

the identity map on a probability space $(\Omega, \mathcal{F}, P)$ corresponding to such a sequence sample space (and equipped with the usual cylindrical σ-field). Let T be the left shift on Ω defined by

$$T(\ldots, x_{-1}, x_0, x_1, x_2, \ldots) = (\ldots, x_0, x_1, x_2, x_3, \ldots).$$

The basic notion in ergodic theory is that of *ergodicity* of a transformation. A transformation T is ergodic if there is no shift invariant measurable set A (i.e., a set satisfying $P(A \Delta (T^{-1}A)) = 0$) with $0 < P(A) < 1$. Equivalently, the shift transformation T is ergodic if for every measurable function $f \in L_1(\Omega, \mathcal{F}, P)$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j \mathbf{X}) = Ef(\mathbf{X}) \quad \text{a.s..} \tag{4.1}$$

Here $T^j\mathbf{X}$ is the jth shift of the process $\mathbf{X}$: $T^j(\ldots, X_{-1}, X_0, X_1, X_2, \ldots) = (\ldots, X_{j-1}, X_j, X_{j+1}, X_{j+2}, \ldots)$.

The usual terminology is to call a stationary stochastic process ergodic if the corresponding shift transformation T defined on some (equivalently, any) sequence sample space supporting the process is ergodic. It is easy to see that a stationary stochastic process is not ergodic if and only if it can be represented a non-trivial mixture of two different stationary stochastic processes:

$$\mathbf{X} = \begin{cases} \mathbf{Y} & \text{with probability } p \\ \mathbf{Z} & \text{with probability } 1-p \end{cases}, \tag{4.2}$$

where $0 < p < 1$ and $\mathbf{Y}$ and $\mathbf{Z}$ are stationary stochastic processes with different finite-dimensional distributions. Indeed, suppose that $\mathbf{X}$ is not ergodic, and take a shift invariant measurable subset A of the sequence space with $p = P(A) \in (0,1)$. Then (4.2) holds with both $\mathbf{Y}$ and $\mathbf{Z}$ being the canonical processes on the sequence space equipped with probability measures $P_1 = p^{-1}P|_A$ and $P_2 = (1-p)^{-1}P|_{A^c}$, accordingly. Since A is shift invariant, P_1 and P_2 are not affected by the shift, and so the two processes are stationary. Since A and A^c are disjoint, the two probability measures are different, and so $\mathbf{Y}$ and $\mathbf{Z}$ have different finite-dimensional distributions. Conversely, suppose that (4.2) holds. Then there is a bounded measurable function f such that $Ef(\mathbf{Y}) \neq Ef(\mathbf{Z})$,

and then by the ergodic theorem

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} f(T^j\mathbf{X}) = \begin{cases} L_1 & \text{with probability } p \\ L_2 & \text{with probability } 1-p \end{cases},$$

where L_1 and L_2 are random variables satisfying $EL_1 = Ef(\mathbf{Y})$ and $EL_2 = Ef(\mathbf{Z})$, implying that (4.1) fails, and so $\mathbf{X}$ is not ergodic.

Note that it is very natural to say that a non-ergodic stationary process $\mathbf{X}$ *has infinite memory*. Indeed, a non-ergodic process has the structure given in (4.2), and so the result of a single "coin toss" (with probabilities p and $1-p$) will be "remembered forever." Therefore, it certainly makes sense to call stationary ergodic processes "processes with finite memory," and stationary non-ergodic processes "processes with infinite memory."

It is, then, very tempting to try to find another ergodic theoretical notion, stronger than ergodicity, corresponding to stationary processes with finite and short memory. Then ergodic stationary processes that lack this stronger property will be naturally called processes with long memory. A very natural notion to try is *mixing*. Recall that a transformation T of a probability space is mixing if for every two measurable sets A and B we have $P\big(A \cap T^{-n}B\big) \to P(A)P(B)$ as $n \to \infty$, and a stationary process is called mixing if the corresponding shift transformation T is mixing. The shift transformation can, once again, be defined on any sequence sample space supporting the process.

It is obvious that a mixing stationary process is ergodic, and it is easy to construct examples of ergodic but non-mixing stationary processes. A candidate definition of a long range dependent process would then refer to an ergodic but non-mixing process.

Such a definition has not become standard, for the reasons that will be discussed below. Note, however, that the approaches to memory of a stationary process via the ergodic theoretical properties of the corresponding shift transformation are very attractive from the following point of view. Let $\mathbf{X}$ be a stationary process, and let the process $\mathbf{Y}$ be derived from the process $\mathbf{X}$ by means of a point transformation $Y_n = g(X_n)$ for all n, where $g : \mathbb{R} \to \mathbb{R}$ is a measurable function. Clearly, $\mathbf{Y}$ is also a stationary process. It is intuitively clear that the process

X "remembers at least as much" as the process **Y** does. If, in particular, g is a one-to-one map, and g^{-1} is also measurable, then this intuition says that the processes **X** and **Y** should have "the same length of memory": if one of them has long memory, then so should do the other one.

This, apparently very natural, requirement has proved to be difficult to satisfy by many of the proposed definitions of long range dependence. It is, however, automatic with ergodic theoretical-based definitions. Indeed, it follows from the definition of the ergodicity and mixing that **X** is ergodic (mixing) if and only if **Y** has this property.

It is instructive to record what the ergodic theoretical-based notions of memory discussed above mean for stationary Gaussian processes. Let **X** be a (real-valued) stationary Gaussian process with covariance function $R_k, k \geq 0$ and spectral measure F on $(-\pi, \pi]$. That is, $R_k = \int_{(-\pi,\pi]} \cos(kx)\, F(dx)$ for $k \geq 0$. Then

- the process **X** is ergodic if and only if the spectral measure F is atomless;
- the process **X** is mixing if and only if $R_k \to 0$ as $k \to \infty$;

see [34]. The requirement that the covariance function vanishes at the high lag limit has, however, proved to be insufficient when dealing with long memory for Gaussian processes. Indeed, many "unusual" phenomena have been observed for Gaussian processes whose covariance function does vanish in the limit, but sufficiently slowly, as we have already seen on the example of the Fractional Gaussian noise. Therefore, the mixing property is not believed to be sufficiently strong to say that a stationary process with this property has short memory. A stronger requirement is needed.

Several such stronger requirements have been introduced; they are collectively known under the name "strong mixing conditions" (even though one of them carries that same name separately). We refer the reader to the recent survey [24] for a detailed discussion of the notions we introduce here and their relation to the other possible conditions.

Let, once again, $\mathbf{X} = (\ldots, X_{-1}, X_0, X_1, X_2, \ldots)$ be a stationary process. Define for $n \geq 1$

$$\alpha_{\mathbf{X}}(n) = \sup \big\{ |P(A \cap B) - P(A)P(B)|, \\ A \in \sigma\big(X_k, k \leq 0\big), B \in \sigma\big(X_k, k \geq n\big) \big\}. \qquad (4.3)$$

The process $\mathbf{X}$ is called *strongly mixing* if $\alpha_{\mathbf{X}}(n) \to 0$ as $n \to \infty$. A possible connection between the strong mixing property and lack of long memory (i.e., short memory) has been observed beginning with [116]. Specifically, it turns out that strong mixing is related to the fact whether or not the partial sums of the process $\mathbf{X}$ satisfy the Functional Central Limit Theorem (2.3).

Let, as before, $S_m = X_1 + \cdots + X_m$, $m = 0, 1, \ldots$, and define the partial sum process by (2.2).

Theorem 4.1. Assume that $\mathbf{X}$ is a zero mean strongly mixing process such that for some $\delta > 0$, $|X_0|^{2+\delta} < \infty$. Assume that $\mathrm{Var}(S_m) \to \infty$ as $m \to \infty$, and for some $K < \infty$

$$E\big|S_m\big|^{2+\delta} \leq K \left(\mathrm{Var}(S_m)\right)^{1+\delta/2} \text{ for all } m.$$

Then the properly normalized partial sum process converges weakly to the Brownian motion, i.e.,

$$\frac{1}{(\mathrm{Var}(S_n))^{1/2}} S^{(n)} \Rightarrow B \quad \text{weakly in } D[0,1],$$

where B is the standard Brownian motion. Moreover, assume, in addition, that the covariances $R_n = \mathrm{Cov}(X_0, X_n)$, $n = 0, 1, \ldots$ are summable:

$$\sum_{n=0}^{\infty} |R_n| < \infty,$$

Then the limit $\sigma_*^2 = \lim_{n\to\infty} \mathrm{Var}(S_n)/n$ exists, is finite and positive, and (2.3) holds.

See [96, 116], Proposition 34, and [23] for the last comment.

Theorem 4.1 indicates that a strongly mixing process behaves, as far as the central limit theorem is concerned, similarly to an i.i.d. sequence

and, hence, can be viewed as having short memory. Extra moment conditions involved are somewhat disappointing, and turns out that imposing "a stronger strong mixing condition" allows one to get rid of these extra conditions. For a stationary process $\mathbf{X}$ define

$$\alpha^*_{\mathbf{X}}(n) = \sup_{S,T}\sup \big\{|P(A\cap B) - P(A)P(B)|, \\ A \in \sigma\big(X_k, k\in S\big), B\in \sigma\big(X_k, k\in T\big)\big\}, \quad (4.4)$$

where the first supremum in (4.4) is taken over all subsets S and T of integers satisfying

$$\mathrm{dist}(S,T) := \min_{k_1\in S, k_2\in T} |k_1 - k_2| \ge n\,.$$

Clearly, $\alpha_{\mathbf{X}}(n) \le \alpha^*_{\mathbf{X}}(n)$. The process $\mathbf{X}$ is *interlaced strongly mixing* if $\alpha^*_{\mathbf{X}}(n) \to 0$ as $n\to\infty$. The following is a central limit theorem under this stronger mixing assumption.

Theorem 4.2. Assume that $\mathbf{X}$ is a zero mean finite variance interlaced strongly mixing process, and $\mathrm{Var}(S_m)\to\infty$ as $m\to\infty$. Then the properly normalized partial sum process converges weakly to the Brownian motion, i.e.,

$$\frac{1}{(\mathrm{Var}(S_n))^{1/2}}\, S^{(n)} \Rightarrow B \quad \text{weakly in } D[0,1],$$

where B is the standard Brownian motion. Furthermore, the limit $\sigma^2_* = \lim_{n\to\infty}\mathrm{Var}(S_n)/n$ exists, is finite and positive, and (2.3) holds.

See [110] for a proof under the so-called "interlaced ρ-mixing condition, and [22] for the equivalence of the two "interlaced conditions." In fact, Theorem 4.2 holds for a strongly mixing stationary process that satisfies $\alpha^*_{\mathbf{X}}(n) < 1$ for some $n\ge 1$.

For a stationary Gaussian process to be strongly mixing, it is necessary that its spectral measure be absolutely continuous with respect to the Lebesgue measure on $(-\pi,\pi]$. If the spectral density (i.e., the derivative of the spectral measure with respect to the Lebesgue measure on $(-\pi,\pi]$) is continuous and positive, then the process is interlaced

strong mixing (see [76] and [119]). Necessary and sufficient conditions for strong mixing of a stationary Gaussian process were later established in [62]. Explicit necessary and sufficient conditions for interlaced strong mixing of stationary Gaussian process do not appear to have been stated.

The above results explain why absence of one or another strong mixing condition (as opposed to the ergodic-theoretical mixing) is sometimes taken as the definition of long range dependence. The strong mixing properties share with the ergodic-theoretical notions of ergodicity and mixing the very desirable feature discussed above: if a process $\mathbf{Y}$ is derived from a process $\mathbf{X}$ by means of a one-to-one point transformation $Y_n = g(X_n)$ for all n, then the process $\mathbf{X}$ has long memory in the sense of lacking of one of the strong mixing properties if and only if the process $\mathbf{Y}$ does.

In spite of these attractive features of the strong mixing conditions they have not become standard as definitions of long range dependence. To some extent this is due to the fact that if one is interested not in the partial sums of a stationary process but in, say, *partial maxima*, then strong mixing conditions, while relevant, allow clustering and, hence, limits different from the ones seen for i.i.d. sequences; see [81]. More importantly, the strong mixing conditions are not easily related to the natural building blocks of many stochastic models, and are difficult to verify, with the possible exception of Gaussian processes and Markov chains. Even in the latter cases necessary and sufficient conditions are not always available, and the picture is not completely clear.

5

Second-Order Theory

By far the most popular point of view on long range dependence is through a slow decay of correlations. This is related to the original explanation of the Hurst phenomenon by Mandelbrot, discusses in Section 2, and to the simple fact that correlations are one of the easiest to understand and estimate features of a stochastic model. Clearly, such approaches to the notion of long memory are restricted to second-order stationary processes, and this is the assumption that will made throughout this section. A related, if not entirely equivalent, second-order approach is through the behavior of the spectral density of the process (assuming its existence) at the origin. These issues are discussed in this section.

The discussion in this section makes heavy use of the notions of regular and slow varying functions. We start with a brief summary of regular variation, and refer the reader to [19] for encyclopedic treatment of the subject. A measurable function $f : (0, \infty) \to \mathbb{R}$ is said to be regularly varying at infinity with exponent $\alpha \in \mathbb{R}$ if f does not vanish for large enough values of the argument and $f(ax)/f(x) \to a^\alpha$ as $x \to \infty$ for every $a > 0$. A function regularly varying at infinity with exponent $\alpha = 0$ is called slowly varying at infinity. A function f as above is said

to be regularly varying at zero with exponent $\alpha \in \mathbb{R}$ if $f(1/\cdot)$ is regularly varying at infinity with exponent $-\alpha$. Clearly, a regularly varying at infinity with exponent α function f can be represented in the form $f(x) = x^{\alpha}L(x)$, $x > 0$, where L is slowly varying at infinity. A proper subclass of slowly varying at infinity eventually positive functions is defined as follows: an eventually positive function $L : (0, \infty) \to \mathbb{R}$ is said to belong to the *Zygmund class* if

$$\begin{aligned} &\text{for every } \delta > 0 \text{ the function } g_1 = x^{\delta}L(x) \text{ is eventually increasing} \\ &\quad \text{and the function } g_2 = x^{-\delta}L(x) \text{ is eventually decreasing.} \end{aligned} \tag{5.1}$$

By Theorem 1.5.5 of [19], the Zygmund class coincides with the class of normalized slowly varying functions.

We will also apply the notion of regular variation to sequences of real numbers. A sequence $\big(a_n, n = 1, 2, \ldots\big)$ is called regular varying with exponent α if there is a regularly varying at infinity with exponent α function f such that $a_n = f(n)$ for every n.

With the above in mind, we are ready to discuss the second-order notions of memory. Let $\mathbf{X} = (X_1, X_2, \ldots)$ be a zero mean stationary stochastic process with a finite variance, $EX_1^2 = \sigma^2 \in (0, \infty)$, covariances $R_n = \mathrm{Cov}(X_1, X_{n+1})$ and correlations $\rho_n = R_n/\sigma^2$, $n = 0, 1, \ldots$. We start with an obvious computation of the variance of the partial sum $S_n = X_1 + \cdots + X_n$. We have

$$\begin{aligned} \mathrm{Var}S_n &= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{Cov}(X_i, X_j) = \sigma^2\sum_{i=1}^{n}\sum_{j=1}^{n}\rho_{|i-j|} \\ &= \sigma^2\left(n + 2\sum_{i=1}^{n-1}(n-i)\rho_i\right), \end{aligned} \tag{5.2}$$

and the behavior of the last sum is closely related to how fast the correlations of the process decay. Assume that they are summable:

$$\sum_{n=0}^{\infty}|\rho_n| < \infty\,. \tag{5.3}$$

Then by the dominated convergence theorem the limit

$$\lim_{n\to\infty}\frac{\mathrm{Var}S_n}{n} = \sigma^2\left(1 + 2\sum_{i=1}^{\infty}\rho_i\right) := \sigma_*^2 \tag{5.4}$$

exists, and is finite. However, it is possible that the limit σ_*^2 is equal to zero. Assuming that it is not equal to zero, we conclude that the variance of the partial sums of the process $\mathbf{X}$ grows linearly fast with the number of terms. On the other hand, suppose that the correlations of the process are, in fact, regularly varying:

$$\rho_n = n^{-d} L(n) \tag{5.5}$$

for $n \geq 1$, where $0 \leq d < 1$, and L is a slowly varying at infinity function. Then by Karamata's theorem (see e.g., Theorem 0.6 in [113]),

$$\mathrm{Var} S_n \sim \frac{2\sigma^2}{(1-d)(2-d)} L(n) n^{2-d} \quad \text{as } n \to \infty. \tag{5.6}$$

That is, regular variation of correlations as in (5.5) implies that the variance of the partial sums grows much faster than in the case of summable correlations. This is, of course, the case for the Fractional Gaussian Noise (with $H > 1/2$) of Section 2, whose asymptotic behavior is given by (2.9).

When the variance of the partial sums of a stationary process grows linearly fast with the number of terms, at least from that point of view, the process "is not far" from an i.i.d. sequence. Furthermore, we have seen in Section 4 that this, under certain strong mixing and moment assumptions, means that the classical invariance principle holds (modulo a different variance of the limiting Brownian motion). This is also true, for example, under the assumption of association — see [106]. On the other hand, when the variance of the partial sums grows as a regularly varying function with the exponent larger than 1 — as in (5.6) — it follows immediately from Lamperti's Theorem (see e.g., Theorem 2.1.1 in [50]) that convergence to the Brownian motion is impossible, no matter what normalization one uses, and so the invariance principle does not hold.

This is the main reason why the summability of correlations in (5.3) is often taken as the indication of short memory, and its opposite, the divergence of the series in the left-hand side of (5.3), as the definition of the long range dependence. On the other hand, it is also possible to take the rate of increase of the variance of the partial sums itself to draw the boundary. From this point of view, one could say that a

second-order stationary process has short memory if

$$\lim_{n\to\infty} \frac{\mathrm{Var} S_n}{n} < \infty\,, \tag{5.7}$$

and the infinite limit in (5.7) would then be taken as an indication of long range dependence. This is sometimes referred to as *Allen variance* short and long memory; see [64].

Of course, the summability of correlations (5.3) is not necessary for an at most linear rate of increase of the variance in (5.7). In fact, rewrite (5.2) as

$$\frac{\mathrm{Var} S_n}{n} = \sigma^2 \left(1 + 2\frac{1}{n} \sum_{j=1}^{n-1} \sum_{i=1}^{j} \rho_i \right). \tag{5.8}$$

In particular, if

$$\text{the sum } \sum_{n=0}^{K} \rho_n \quad \text{converges as } K \to \infty\,, \tag{5.9}$$

then, since the usual convergence implies the Cesaro convergence, we will still obtain (5.4), regardless of the summability of the correlations. Such situations are, clearly, possible. A simple example is $\rho_n = \sin na/na$, $n = 1, 2, \ldots$ for $0 < a < \pi$. However, even the convergence in (5.9) is not necessary for (5.7), as another simple example $\rho_n = (-1)^n/2$, $n = 1, 2, \ldots$ shows.

To get better understanding of the condition (5.7) we need to concentrate on the spectrum of the covariance function of the process. Recall that the spectral measure F is a measure on $(-\pi, \pi]$, satisfying $R_k = \int_{(-\pi,\pi]} \cos(kx)\, F(dx)$ for $k \geq 0$. Recall, further, that if the correlations are absolutely summable as in (5.3), then the spectral measure has a continuous density with respect to the Lebesgue measure on $(-\pi, \pi]$, *the spectral density*, given by

$$f(x) = \frac{\sigma^2}{2\pi} \left(1 + 2 \sum_{n=1}^{\infty} \rho_n \cos nx \right), \quad -\pi < x < \pi\,. \tag{5.10}$$

A simple computation will allow us to relate the right-hand side of (5.8) to the spectral measure. Assuming that the spectral measure does not

have atoms at zero and at π, we have for every $j \geq 1$

$$\begin{aligned}\sum_{i=1}^{j} \rho_i &= \frac{1}{\sigma^2} \int_{(-\pi,\pi]} \sum_{i=1}^{j} \rho_i \cos(ix)\, F(dx) \\ &= \frac{1}{2\sigma^2} \int_{(-\pi,\pi]} \frac{1}{\sin x} \big(\sin(j+1)x + \sin jx - \sin x\big)\, F(dx) \\ &= \frac{1}{2\sigma^2} \int_{(-\pi,\pi]} \frac{\sin(j+1)x}{\sin x}\, F(dx) + \frac{1}{2\sigma^2} \int_{(-\pi,\pi]} \frac{\sin jx}{\sin x}\, F(dx) - \frac{1}{2}.\end{aligned}$$

Furthermore, for $n \geq 1$,

$$\begin{aligned}&\sum_{j=1}^{n-1} \frac{1}{2\sigma^2} \int_{(-\pi,\pi]} \frac{\sin(j+1)x}{\sin x}\, F(dx) \\ &\quad = \frac{1}{2\sigma^2} \int_{(-\pi,\pi]} \frac{1}{\sin x} \sum_{j=1}^{n-1} \sin(j+1)x\, F(dx) \\ &\quad = \frac{1}{4\sigma^2} \int_{(-\pi,\pi]} \frac{1}{\sin^2 x} \big(\cos x + \cos 2x - \cos nx - \cos(n+1)x\big)\, F(dx)\,.\end{aligned}$$

Similarly,

$$\begin{aligned}&\sum_{j=1}^{n-1} \frac{1}{2\sigma^2} \int_{(-\pi,\pi]} \frac{\sin jx}{\sin x}\, F(dx) \\ &\qquad = \frac{1}{4\sigma^2} \int_{(-\pi,\pi]} \frac{1}{\sin^2 x} \big(1 + \cos x - \cos(n-1)x - \cos nx\big)\, F(dx)\,,\end{aligned}$$

and so

$$\frac{\mathrm{Var} S_n}{n} = a_{n-1} + 2a_n + a_{n+1} + o(1), \tag{5.11}$$

where

$$a_n = \frac{1}{2n} \int_{(-\pi,\pi]} \frac{1 - \cos nx}{x^2}\, F(dx), \quad n = 1, 2, \ldots. \tag{5.12}$$

One immediate conclusion is as follows.

Proposition 5.1. Suppose that for some $\epsilon > 0$ the spectral measure F has a density in the interval $(-\epsilon, \epsilon)$, and

$$\text{the density has a continuous at the origin version } f\,. \tag{5.13}$$

Then

$$\lim_{n\to\infty} \frac{\mathrm{Var} S_n}{n} = 2\pi\, f(0)\,.$$

Proof. Suppose first that the spectral measure has no atom at π. Then, clearly,

$$\begin{aligned} a_n &= \frac{1}{2n}\int_{-\epsilon}^{\epsilon} \frac{1-\cos nx}{x^2}\, f(x)\, dx + o(1) \\ &= \frac{f(0)}{n}\int_{0}^{\epsilon} \frac{1-\cos nx}{x^2}\, dx + o(1) \\ &= f(0)\int_{0}^{n\epsilon} \frac{1-\cos y}{y^2}\, dy + o(1) \\ &\to f(0)\int_{0}^{\infty} \frac{1-\cos y}{y^2}\, dy = \frac{\pi}{2} f(0)\,, \end{aligned}$$

and our statement follows from (5.11). Observing that adding an atom at the point π does not change the rate of growth of the variance of the partial sums, we see that the proof is complete. □

In particular, the condition

the process has a continuous at the origin spectral density (5.14)

is sometimes taken as another definition of a process with short memory. This condition is also known to be sufficient for the Central Limit Theorem for linear processes; see e.g., Corollary 5.2 in [59]. Of course, Proposition 5.1 shows that this condition is not necessary for an at most linear rate of increase of the variance as in (5.7), since it allows arbitrarily "bad" spectral measure outside of a neighborhood of zero. In fact, (5.7) can happen even if there is no neighborhood of zero where the process has a continuous at the origin spectral density as the example of the process with the spectral density $f(x) = 1 + \cos(1/x)$, $-\pi < x < \pi$, shows.

Summarizing, the assumptions (5.3), (5.7), (5.9), and (5.14) have all been used to define a short memory process in the sense of the variance of the partial sums of the process increasing at most linearly

fast. These assumptions are, of course, not equivalent. Moreover, it is also possible to have a process with a bounded from zero and infinity spectral density for which

$$0 < \liminf_{n\to\infty} \frac{\mathrm{Var} S_n}{n} < \limsup_{n\to\infty} \frac{\mathrm{Var} S_n}{n} < \infty;$$

an example is constructed in [23].

Let us now see what can cause the variance of the partial sums grow faster that linearly fast. Here is a counterpart to Proposition 5.1.

Proposition 5.2. Suppose that for some $\epsilon > 0$ the spectral measure F has a density in the interval $(-\epsilon, \epsilon)$, with a version f such that

$$f(x) = x^{-(1-d)} L_1(x), \quad 0 < x < \epsilon,$$

for some $0 < d < 1$, where L_1 is a slowly varying at zero function. Then

$$\mathrm{Var} S_n \sim \frac{4\Gamma(d)\cos(\pi d/2)}{(1-d)(2-d)} L_1(1/n)\, n^{2-d} \quad \text{as } n \to \infty.$$

Proof. As in the proof of Proposition 5.1 we may assume that the spectral measure of the process has no atom at the point π. For a_n defined in (5.12) we have

$$\begin{aligned} a_n \sim \frac{1}{n} \int_0^\epsilon \frac{1-\cos nx}{x^2} f(x)\, dx &= \frac{1}{n} \int_0^\epsilon \frac{1-\cos nx}{x^{3-d}} L_1(x)\, dx \\ &= n^{1-d} \int_0^{n\epsilon} \frac{1-\cos y}{y^{3-d}} L_1(y/n)\, dy. \end{aligned}$$

By Potter's bounds (see e.g., Proposition 0.8 in [114]) there is a finite positive constant C such that $L_1(y/n)/L_1(1/n) \le C\, y^{-d/2}$ for all $n \ge 1$ and $0 < y < n\epsilon$. By the dominated convergence theorem we obtain

$$\begin{aligned} a_n &\sim n^{1-d} L_1(1/n) \int_0^\infty \frac{1-\cos y}{y^{3-d}}\, dy \\ &= \frac{\Gamma(d)\cos(\pi d/2)}{(1-d)(2-d)} L_1(1/n)\, n^{1-d} \quad \text{as } n \to \infty. \end{aligned}$$

Now an appeal to (5.11) completes the proof. □

A related statement about the largest eigenvalue of the n-dimensional covariance matrix of the process is in [21].

Comparing the statement of Proposition 5.2 with the consequence (5.6) of the regular variation of the correlations in (5.5) we see that, for $0 < d < 1$, the assumptions

$$\rho_n \sim n^{-d} L(n) \quad \text{as } n \to \infty \tag{5.15}$$

and existence in the neighborhood of the origin of a spectral density satisfying

$$f(x) \sim x^{-(1-d)} L(1/x) \frac{\sigma^2}{2\Gamma(d)\cos(\pi d/2)} \quad \text{as } x \downarrow 0 \tag{5.16}$$

lead to the same asymptotic behavior of the variance of the partial sums.

Example 5.1. The Fractional Gaussian Noise with covariance function given by (2.8) has a spectral density given by the formula

$$f(x) = \frac{\sigma^2}{2} C(H)(1-\cos x) \sum_{j=-\infty}^{\infty} |2\pi j + x|^{-(1+2H)}, \tag{5.17}$$

where

$$C(H) = \frac{2H(1-2H)}{\Gamma(2-2H)} \frac{1}{\cos \pi H}, \quad H \neq 1/2.$$

Indeed, with f as above,

$$\begin{aligned}
\int_{(-\pi,\pi]} \cos(nx)\, f(x)\, dx &= \frac{\sigma^2}{2} C(H) \int_{-\infty}^{\infty} |x|^{-(1+2H)} (1-\cos x)\cos(nx)\, dx \\
&= \frac{\sigma^2}{2} C(H) \left[\int_0^{\infty} x^{-(1+2H)} (1-\cos(n+1)x)\, dx \right. \\
&\qquad + \int_0^{\infty} x^{-(1+2H)} (1-\cos(n-1)x)\, dx \\
&\qquad \left. -2\int_0^{\infty} x^{-(1+2H)} (1-\cos nx)\, dx \right]
\end{aligned}$$

$$= \frac{\sigma^2}{2} C(H) \int_0^\infty x^{-(1+2H)} (1 - \cos x)\, dx$$
$$\times \left[(n+1)^{2H} + |n-1|^{2H} - 2n^{2H} \right]$$
$$= \frac{\sigma^2}{2} \left[(n+1)^{2H} + |n-1|^{2H} - 2n^{2H} \right],$$

as in (2.8). Clearly, for $1/2 < H < 1$ the spectral density in (5.17) satisfies

$$f(x) \sim \frac{\sigma^2}{4} C(H)\, x^{-(2H-1)} \quad \text{as } x \downarrow 0,$$

and it is easily verified that for FGN the asymptotic behavior of correlations in (2.9) and of the spectral density at zero are related as in (5.15) and (5.16).

In fact, all three statements (5.6), (5.15), and (5.16) have been taken as definitions of long range dependence. The behavior of the partial sum variance in (5.6) is, clearly, the least demanding of these three statements. For example, a process with a spectral density equal in a neighborhood of the origin to

$$f(x) = \big(1 + \cos(1/x)\big) x^{-(1-d)} L(1/x)\, \frac{\sigma^2}{2\Gamma(d)\cos(\pi d/2)},$$

which is not regularly varying, will still satisfy (5.6). However, the relationship between (5.15) and (5.16) has been somewhat of a mystery, and in the literature one can sometimes read that these latter statements are equivalent. In fact, these statement are not equivalent; we will present examples shortly. The claims of equivalence appear to stem from several sources, one being a casual treatment of similar conditions in an influential paper of Cox [35], and the second a definition of a slowly varying function in [150], as a function satisfying (5.1), that is more restrictive than what is understood by this notion today.

We start with a positive result, that gives sufficient conditions under which the statements (5.15) and (5.16) imply each other.

Theorem 5.3. (i) Assume that the correlations are regularly varying in the sense of (5.15), and the function L belongs to the Zygmund class (i.e., satisfies (5.1)). Then the process has a spectral density that is regularly varying at zero, in the sense that (5.16) holds.

(ii) Conversely, assume that the process has a spectral density that is regularly varying at zero, in the sense that (5.16) holds, and the function L belongs to the Zygmund class. Then the correlations are regularly varying in the sense of (5.15).

Part (i) of Theorem 5.3 is in Theorem (2–6) in Chapter V of [150]. The proof of part (ii) will appear separately.

The following is an example of a situation where a spectral density is regularly varying at the origin as in (5.16), but correlations are not regularly varying as in (5.15).

Example 5.2. Let $0 < \epsilon < \pi/2$, and g a positive integrable function on $(0, \epsilon)$ satisfying (5.16). Let

$$f(x) = g(|x|)\mathbf{1}(0 < |x| < \epsilon) + g(|\pi - x|)\mathbf{1}(\pi - \epsilon < |x| < \pi)$$

for $-\pi < x < \pi$. Then f is a spectral density satisfying (5.16). Notice that

$$\begin{aligned} R_n = \int_{-\pi}^{\pi} \cos nx\, f(x)\, dx &= \int_{-\epsilon}^{\epsilon} \cos nx\, g(x)\, dx + \int_{-\epsilon}^{\epsilon} \cos n(\pi - x)\, g(x)\, dx \\ &= (1 + (-1)^n)\hat{R}_n\,, \end{aligned}$$

where $\hat{R}_n = \int \cos nx\, g(x)\, dx$. Since R_n vanishes for all odd lags n, the correlations are not regularly varying, and (5.15) fails. Examples of this sort can also be constructed by letting the spectral density "blow up" around points other than $x = \pi$.

Next is an example of a situation where the correlations are regularly varying as in (5.15), but there is no version of a spectral density that is regularly varying at the origin as in (5.16).

Example 5.3. Let us start with a spectral density g that does satisfy (5.16), and such that the correlations are also regularly varying as in (5.15) (for example, one can take the spectral density of an FGN as in Example 5.1, or any spectral density satisfying part (ii) of Theorem 5.3.) We will construct a continuous nonnegative integrable function g_1 on $(0, \pi)$ such that

$$\limsup_{x \downarrow 0} x^2 g_1(x) > 0 \tag{5.18}$$

and

$$\int_0^\pi \cos nx \, g_1(x) \, dx = o\left(\int_{-\pi}^\pi \cos nx \, g(x) \, dx \right) \tag{5.19}$$

as $n \to \infty$. Then we will set $f(x) = g(x) + g_1(|x|)$ for $-\pi < x < \pi$. It will follow from (5.18) that some pieces of f are too large near the origin to permit f to be regularly varying as in (5.16), but (5.19) and regular variation of the correlations corresponding to the density g mean that the correlations corresponding to the density f will also satisfy (5.15).

We proceed with a construction of a function g_1. Define

$$g_1(x) = 2^{2j} \quad \text{if} \quad 2^{-j} \le x \le 2^{-j} + 2^{-2^j} \quad \text{for } j = 0, 1, \ldots . \tag{5.20}$$

Clearly for $x = 2^{-j}$, $g_1(x) = x^{-2}$, so (5.18) holds. Further,

$$\begin{aligned}
\left| \int_0^\pi \cos nx \, g_1(x) \, dx \right| &\le \sum_{j=0}^\infty 2^{2j} \left| \int_{2^{-j}}^{2^{-j} + 2^{-2^j}} \cos nx \, dx \right| \\
&= \frac{2}{n} \sum_{j=0}^\infty 2^{2j} \left| \sin\left(\frac{n}{2} 2^{-2^j} \right) \cos(n(2^{-j} + 2^{-2^j - 1})) \right| \\
&\le \frac{2}{n} \sum_{j=0}^\infty 2^{2j} \left| \sin\left(\frac{n}{2} 2^{-2^j} \right) \right| \\
&= \frac{2}{n} \sum_{j \le \log_2 \log_2 n} 2^{2j} \left| \sin\left(\frac{n}{2} 2^{-2^j} \right) \right| \\
&\quad + \frac{2}{n} \sum_{j > \log_2 \log_2 n} 2^{2j} \left| \sin\left(\frac{n}{2} 2^{-2^j} \right) \right| .
\end{aligned}$$

Clearly,

$$\frac{2}{n}\sum_{j\le\log_2\log_2 n} 2^{2j}\left|\sin\left(\frac{n}{2}2^{-2^j}\right)\right| \le \frac{2}{n}\sum_{j\le\log_2\log_2 n} 2^{2j}$$
$$\le cn^{-1}\left(\log_2 n\right)^2 \text{ for some } 0<c<\infty$$

and

$$\frac{2}{n}\sum_{j>\log_2\log_2 n} 2^{2j}\left|\sin\left(\frac{n}{2}2^{-2^j}\right)\right| \le \frac{2}{n}\sum_{j>\log_2\log_2 n} 2^{2j}\left(\frac{n}{2}2^{-2^j}\right)$$
$$= \sum_{j>\log_2\log_2 n} 2^{2j}2^{-2^j} \le cn^{-1}\left(\log_2 n\right)^2 \text{ for some } 0<c<\infty$$

as well. This clearly implies (5.19).

Of course, the function g_1 in (5.20) is not continuous, but it can be easily made such by appropriately "connecting the dots" at the jump points of the function in (5.20).

If one uses the second-order approach to long range dependence in one of related, but not equivalent, ways discussed above, it would be nice to know that there is stability under point transformations discusses in Section 4. Namely, if $\mathbf{X}$ is a stationary process with a finite variance with, say, regularly varying correlations as in (5.15), and $g:\mathbb{R}\to\mathbb{R}$ is a one-to-one measurable function such that $Eg(X_i)^2<\infty$, then the process $Y_n=g(X_n)$ for $n=0,1,\ldots$ will also have a similar second-order behavior. Unfortunately, this turns out not to be the case, and the correlations of the process $\mathbf{Y}$ may turn out to decay much slower or much faster than those of the process $\mathbf{X}$. To construct examples of this type we need the notion of Hermite polynomials.

For $n\ge 0$ define a function of a real variable x by

$$H_n(x)=(-1)^n e^{x^2/2}\frac{d^n}{dx^n}e^{-x^2/2}$$

(some authors include an extra factor of $1/n!$ in the definition of H_n; see [108], which can also be consulted for more details). Then $H_0(x)=1$, $H_1(x)=x$, $H_x(x)=x^2-1$ and, in general,

$$H_n(x)=\sum_{m=0}^{[n/2]}\frac{n!}{m!(n-2m)!}(-2)^{-m}x^{n-2m}$$

for real x. That is, each H_n is a polynomial of degree n, and is called *the nth Hermite polynomial*. If X is standard normal, then

$$EH_n(X) = 0 \quad \text{for all } n \geq 1, \text{ and} \qquad \mathrm{Var}H_n(X) = n!.$$

More generally, if X and Y are jointly normal, with zero mean, unit variance, and correlation ρ, then

$$EH_n(X)H_m(Y) = \begin{cases} 0 & \text{if } n \neq m \\ \rho^n n! & \text{if } n = m \end{cases}. \tag{5.21}$$

Furthermore, the family $(H_n)_{n\geq 0}$ forms an orthogonal basis in the space $L^2(\mathbb{R}, \mu_G)$, where μ_G is the law of the standard normal random variable. That is, if X is standard normal, and $Eg(X)^2 < \infty$, then

$$g(X) = \sum_{n=0}^{\infty} \frac{a_n}{n!} H_n(X)\,, \tag{5.22}$$

where for $n \geq 0$, $a_n = E\big(H_n(X)g(X)\big)$, and the sum converges in L^2. This is the so-called *Hermite expansion* of the function g, and the smallest $n \geq 1$ such that $a_n \neq 0$ is called *the Hermite rank* of the function g.

Suppose now that $\mathbf{X}$ is a stationary Gaussian process, with zero mean and unit variance, and correlations satisfying (5.15). Let $g : \mathbb{R} \to \mathbb{R}$ be a measurable function such that $Eg(X_i)^2 < \infty$, and define a new stationary process $\mathbf{Y}$ by $Y_n = g(X_n)$ for $n = 0, 1, \ldots$. As above, we tend to think that the process $\mathbf{X}$ "remembers at least as much" as the process $\mathbf{Y}$ does and, if the function g is, additionally, one-to-one, then we expect the two processes to have the same "length of memory." On the other hand, let k be the Hermite rank of the function g. It turns out that it is possible to have a one-to-one measurable function g such that $kd > 1$ (where $0 < d < 1$ is the exponent in (5.15)), as the following example indicates.

Example 5.4. Take $a > 0$ such that

$$ae^{-a} \int_a^{\infty} xe^x e^{-x^2/2}\,dx = \int_0^a x^2 e^{-x^2/2}\,dx\,,$$

and define

$$g(x) = \begin{cases} -\frac{1}{a}x & \text{if } 0 \leq x < a \\ e^{x-a} & \text{if } x \geq a \end{cases}.$$

Set $g(x) = -g(-x)$ for $x < 0$. Clearly, the function g is odd, measurable, one-to-one, and $Eg(X)^2 < \infty$. Furthermore, by the choice of the number a, $E\big(H_n(X)g(X)\big) = 0$ for $n = 1$, and, by the fact that g is odd, the same is true for $n = 2$. Therefore, the Hermite rank of the function g is at least 3 (in fact, it is not difficult to check that in this case the rank k of g is exactly equal to 3, but one can modify the construction and obtain one-to-one functions of arbitrarily high rank). Therefore, if the exponent d in (5.15) satisfies $d > 1/3$, then we have $kd > 1$.

For a process $Y_n = g(X_n)$ for $n = 0, 1, 2, \ldots$, it follows from (5.22), (5.21) and L^2 convergence that the covariance function $R^{(Y)}$ of the process $\mathbf{Y}$ satisfies

$$R_j^{(Y)} = \sum_{n=k}^{\infty} a_n^2 \rho_j^n \sim a_k^2 \rho_j^k$$

as $j \to \infty$, where k is the Hermite rank of g. If g is a one-to-one function g satisfying $kd > 1$, then the two processes, $\mathbf{X}$ and $\mathbf{Y}$, have correlation functions decaying at vastly different rates. In particular, the process $\mathbf{Y}$ will have summable correlations, while the process $\mathbf{X}$ does not; recall that *we expected* $\mathbf{X}$ *and* $\mathbf{Y}$ *to have the same lengthy of memory as the function* g *is one-to-one*! In fact, depending on the Hermite rank of the function g, the process $\mathbf{Y}$ satisfies the Central Limit Theorem only if $kd \geq 1$, whereas in the case $kd < 1$ it satisfies a so-called *Non-Central Limit Theorem*, see [25] and [45].

Therefore, using the behavior of the correlations as a definition of long memory has the weakness that such behavior can drastically change when applying a one-to-one point map. This, unfortunately, is a problem with many alternative definitions, other than ergodic and strong mixing notions of Section 4. Still, this is a warning sign against relying too much on correlations. Incidentally, in the example we have just considered, it definitely makes sense to view the process $\mathbf{X}$ as long range dependent, since it is a centered Gaussian process, and the covariances carry full information about such processes. The same cannot be said about the transformed process $\mathbf{Y}$. In general, the covariances may carry very little information about a process unless it is similar to a Gaussian one.

6

Fractional Processes and Related Models with Long Memory

One often encounters the adjective "fractional" in the names of processes purportedly having long range dependence (the Fractional Gaussian Noise we encountered early on is an example). Partly this is due to the connotation "unusual" the adjective "fractional" carries. A deeper connection exists, however, and it goes back to the issues of stationarity and non-stationarity.

If $\mathbf{X} = (\ldots, X_{-1}, X_0, X_1, \ldots)$ is a stationary process (note that we have switched, once again, to two-sided processes, as in Section 4), then the differenced process $\mathbf{Y}$ with $Y_n = X_n - X_{n-1}$ for $n \in \mathbb{Z}$ is, clearly, also stationary. The typical notation is $\mathbf{Y} = (I - B)\mathbf{X}$, where I is the identity operator on the space of sequences $\mathbf{x} = (\ldots, x_{-1}, x_0, x_1, x_2, \ldots)$, and B is the *backward shift operator* on that space:

$$B(\ldots, x_{-1}, x_0, x_1, x_2, \ldots) = (\ldots, x_{-2}, x_{-1}, x_0, x_1, \ldots)$$

(B is the inverse of the shift operator T of Section 4). On the other hand, not every stationary process $\mathbf{Y}$ is of the form $\mathbf{Y} = (I - B)\mathbf{X}$ for some stationary process $\mathbf{X}$; e.g., a sequence of i.i.d. not identically zero random variables is not of this form. If, however, $\mathbf{Y}$ is of this form, one can write $\mathbf{X} = (I - B)^{-1}\mathbf{Y}$ and call the process $\mathbf{X}$ an integrated process

(specifically, an integrated process $\mathbf{Y}$). Obviously, if an integrated process exists, it is not uniquely determined: one can add the same random variable to each X_n, as long as doing so preserves stationarity.

It is intuitive that the differencing operator on stationary processes, $\Delta = I - B$, makes the memory in the process "less positive, more negative"; this is simply a consequence of alternating plus and minus signs attached to the same random variables. A simple illustration is obtained by considering what happens to a sequence of i.i.d. random variables under differencing. Similarly, if it is possible "to integrate" a stationary process (i.e., to apply the inverse operator $\Delta^{-1} = (I - B)^{-1}$) and obtain a stationary process, the integrated process will tend to have "more positive" memory than the original process. Long memory, when present, is usually "of the positive kind," so one can try to obtain a process with long range dependence by integrating some stationary process, and as many times as possible.

The problem is that, as we know, many "natural" stationary processes cannot be integrated even once, while preserving stationarity. It turns out, however, that sometimes one can integrate a process a fractional number of times, while preserving stationarity. This leads to a class of models known as *fractionally integrated processes*. The success of the construction depends on a definition of a fractional power of the differencing operator Δ, and the starting point is the generalized Binomial formula (or the Taylor expansion): for all real d

$$(1 - x)^d = \sum_{j=0}^{\infty} (-1)^j \binom{d}{j} x^j \,, \tag{6.1}$$

where

$$\binom{d}{j} = \frac{d(d-1)\cdots(d-j+1)}{j!} \,.$$

If d is a nonnegative integer, (6.1) is just the classical Binomial formula, and a sum with finitely many terms; otherwise it is an infinite sum, and then it can be rewritten in the form:

$$(1 - x)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} x^j \,. \tag{6.2}$$

Using the Stirling formula for the Gamma function it is easy to check that

$$\frac{\Gamma(j-d)}{\Gamma(j+1)\Gamma(-d)} \sim \frac{j^{-(d+1)}}{\Gamma(-d)} \tag{6.3}$$

as $j \to \infty$, and so the infinite series in (6.2) converges absolutely for all complex x with $|x| < 1$ (inside the unit circle), and also on the boundary of that circle if $d > 0$.

Given a stationary process $\mathbf{Y}$ we can formally define the process $\mathbf{X} = \Delta^{-d}\mathbf{Y}$ by expanding $\Delta^{-d} = (I - B)^{-d}$ for d that are not non-positive integers into powers of the backward shift operator B as in (6.2) by formally identifying the identity operator with the unity and the backshift operator B with x to obtain

$$X_n = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} Y_{n-j}\,, \tag{6.4}$$

$n = \ldots, -1, 0, 1, 2, \ldots$. If $d > 0$, we view the process $\mathbf{X}$ is an integrated process, while if $d < 0$, we view it as a differenced process $\mathbf{Y}$. We are interested in the "integrated" case, with $0 < d < 1$; if one needs to get beyond this range, one can first perform the usual "non-fractional" integration.

It is clear that, if the series in (6.4) converges in probability, then the resulting process $\mathbf{X}$ is automatically stationary. Therefore, first of all we need to make sure that the infinite series in (6.4) converges. This requires imposing restrictions on the initial process $\mathbf{Y}$.

We start with assuming that the process $\mathbf{Y}$ is a stationary zero mean finite variance process with variance σ^2 and correlation function ρ, satisfying (5.3), that is, a process with absolutely summable correlations. Denoting the jth coefficient in (6.4) by a_j we note that for $m, k \geq 1$

$$E\left(\sum_{j=m+1}^{m+k} a_j Y_{n-j}\right)^2 = \sigma^2 \sum_{j=m+1}^{m+k} a_j^2 + 2\sigma^2 \sum_{j=m+1}^{m+k} a_j \sum_{i=j+1}^{m+k} a_i \rho_{i-j}\,.$$

Since the sequence (a_j) (with $0 < d < 1$) is easily seen to be decreasing, we conclude that

$$E\left(\sum_{j=m+1}^{m+k} a_j Y_{n-j}\right)^2 \le \left(1 + 2\sum_{n=1}^{\infty} |\rho_n|\right) \sigma^2 \sum_{j=m+1}^{m+k} a_j^2 .$$

If $0 < d < 1/2$, then the sum $\sum_j a_j^2$ converges by (6.3), and so the series (6.4) converges in L^2 to a stationary process.

Under somewhat stronger assumptions than the absolute summability of the correlations of the initial process $\mathbf{Y}$, the rate of decay of the correlation function of the partially integrated process is determined by the order of partial integration, as the following proposition shows.

Proposition 6.1. Let $\mathbf{Y}$ be a stationary zero mean finite variance process with variance σ^2 and absolutely summable correlation function ρ. Let $0 < d < 1/2$ and assume that

$$\Psi_n := \sum_{m=n}^{\infty} \rho_m = o\left(n^{-(1-2d)}\right) \quad \text{as } n \to \infty. \tag{6.5}$$

Then the process $\mathbf{X}$ defined by (6.4) is a well defined zero mean stationary process whose covariance function R^* satisfies

$$R_n^* \sim \left(\sigma^2 \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} \sum_{m=-\infty}^{\infty} \rho_m\right) n^{-(1-2d)} \tag{6.6}$$

as $n \to \infty$.

Proof. We have already established that $\mathbf{X}$ is a well defined zero mean stationary process with finite variance. Its covariance function is given by

$$R_n^* = \sigma^2 \lim_{M\to\infty} \sum_{i=0}^{M} \sum_{j=0}^{M} a_i a_j \rho_{n+i-j} = \sigma^2 \lim_{M\to\infty} \sum_{m=n-M}^{M} b_{n-m}^{(M)} \rho_m ,$$

where

$$b_k^{(M)} = \sum_{i=-k\vee 0}^{(M-k)\wedge M} a_i a_{i+k} .$$

Since the numbers $b_k^{(M)}$ are uniformly bounded (by $\sum_{-\infty}^{\infty} a_i^2$) and the correlations of the process $\mathbf{X}$ are absolutely summable, we can use the dominated convergence theorem to obtain

$$R_n^* = \sigma^2 \sum_{m=-\infty}^{\infty} b_{n-m}\rho_m\,,$$

with

$$b_k = \sum_{i=-k\vee 0}^{\infty} a_i a_{i+k} = b_{-k}\,.$$

It follows from (6.3) that

$$b_k \sim \frac{\Gamma(1-2d)}{\Gamma(d)\Gamma(1-d)} k^{-(1-2d)} \quad \text{as } k \to \infty. \tag{6.7}$$

Similarly, since

$$a_k - a_{k+1} = \frac{1-d}{k+1} a_k \sim \frac{1-d}{\Gamma(d)} k^{-(2-d)}\,, \tag{6.8}$$

we obtain also that

$$g_k := b_k - b_{k+1} \sim c k^{-2(1-d)} \quad \text{as } k \to \infty \tag{6.9}$$

for some $c > 0$. Clearly, the statement (6.6) will follow from (6.7) once we check that

$$\lim_{M\to\infty} \limsup_{n\to\infty} n^{1-2d} \left| \sum_{m=-\infty}^{-M} b_{n-m}\rho_m \right| = 0 \tag{6.10}$$

and

$$\lim_{M\to\infty} \limsup_{n\to\infty} n^{1-2d} \left| \sum_{m=M}^{\infty} b_{n-m}\rho_m \right| = 0\,. \tag{6.11}$$

First of all, observe that by monotonicity, for $M > 0$

$$\left| \sum_{m=-\infty}^{-M} b_{n-m}\rho_m \right| \le b_n \sum_{m=-\infty}^{-M} |\rho_m|\,,$$

and so (6.10) follows from (6.7) and summability of the correlations of the process $\mathbf{Y}$. Next, using summation by parts we see that

$$\sum_{m=M}^{\infty} b_{n-m}\rho_m = b_{n-M+1}\Psi_M + \sum_{m=M}^{\infty} g_{n-m}\Psi_m\,.$$

By (6.7), for a $c > 0$,

$$\lim_{M\to\infty} \limsup_{n\to\infty} n^{1-2d} b_{n-M+1}\Psi_M = \lim_{M\to\infty} c\Psi_M = 0\,.$$

Furthermore, write

$$\sum_{m=M}^{\infty} g_{n-m}\Psi_m = \sum_{m\le n/2} + \sum_{m>n/2} := S_n^{(1)}(M) + S_n^{(2)}(M)\,.$$

By (6.9) and the assumption (6.5) wee see that for some constant c and large n

$$\begin{aligned}\left|S_n^{(1)}(M)\right| &\le cn^{-2(1-d)} \sum_{m=M}^{[n/2]} |\Psi_m| \\ &\le cn^{-2(1-d)} \sum_{m=M}^{[n/2]} m^{-(1-2d)} \le cn^{-2(1-2d)}\,,\end{aligned}$$

and so for all $M > 0$

$$\lim_{n\to\infty} n^{1-2d} S_n^{(1)}(M) = 0\,.$$

Finally, by the assumption (6.5) we have

$$\left|S_n^{(2)}(M)\right| \le o(1) n^{-(1-2d)} \sum_{m=-\infty}^{\infty} |g_m|\,.$$

The sum in the right-hand side above is finite by (6.9) and the fact that $g_{-k} = -g_{k-1}$. Therefore, for all $M > 0$

$$\lim_{n\to\infty} n^{1-2d} S_n^{(2)}(M) = 0\,,$$

and we have checked (6.11). □

In particular, if

$$\sum_{m=-\infty}^{\infty} \rho_m \neq 0\,, \tag{6.12}$$

then the correlations ρ_n^* of the fractionally integrated process $\mathbf{X}$ satisfy $\rho_n^* \sim cn^{-(1-2d)}$ as $n \to \infty$ for some $0 < c < \infty$.

Note that it is not surprising that we could only perform above a construction of a fractionally integrated process of the order $0 < d < 1/2$. Indeed, our intuition tells us that the higher the degree of "negative dependence" in a stationary process, the easier it is to integrate it while preserving stationarity. The assumptions in the discussion preceding Proposition 6.1 (in this case, absolute summability of correlations), while preventing the process from having "too much of positive dependence," do not imply any "negative dependence" either. Therefore, the dependence in initial process $\mathbf{Y}$ can be viewed as only assumed to be "midway," between a very negatively dependent process, that can be integrated completely (of order $d = 1$), and a very positively dependent process, that cannot be integrated at all. Hence, intuitively at least, the boundary $d < 1/2$ is understandable. Processes with certain negative dependence can be integrated to a higher order, as we will see in the sequel. Such negative dependence will, in particular, imply that (6.12) breaks down.

In practice one often starts with $\mathbf{Y}$ being an i.i.d. sequence, or a stationary ARMA model (see [26]). In this case the process $\mathbf{Y}$ has exponentially fast decaying correlations, and Proposition 6.1 applies. The resulting models are typically called ARIMA models or, more explicitly, fractional ARIMA (or FARIMA, alternatively ARFIMA) models, and were originally introduced by Granger and Joyeux [57] and Hosking [65].

In the spectral domain things are even more transparent. Suppose that the original process $\mathbf{Y}$ has absolutely summable correlations, and so it has a continuous spectral density f given by (5.10). Since the series (6.4) converges in L^2, the fractionally integrated process $\mathbf{X}$ has also a spectral density, f^*, given by $f^*(x) = \left|\sum_{m=0}^{\infty} a_m e^{imx}\right|^2 f(x)$, (where (a_j) are the coefficients in (6.4)), and the infinite sum in the

expression for the density converges in $L^2\big((-\pi,\pi], f(x)dx\big)$; see e.g., Theorem 4.10.1 in [26]. Note that $\big|\sum_{m=0}^{\infty} a_m z^m\big|^2 = |1-z|^{-2d}$ for complex numbers z with $|z|<1$, and the right-hand side of this relation has a continuous extension to the part of the unit circle that is bounded away from the point $z=1$. It follows that $\big|\sum_{m=-\infty}^{\infty} a_m e^{imx}\big|^2 = |1-e^{ix}|^{-2d}$ for almost every $x \in (-\pi,\pi]$ (with respect to the Lebesgue measure), and so the integrated process $\mathbf{X}$ has a spectral density given by

$$f^*(x) = |1-e^{ix}|^{-2d} f(x), \quad x \in (-\pi,\pi]. \tag{6.13}$$

In particular,

$$f^*(x) \sim x^{-2d} f(0) = \left(\frac{\sigma^2}{2\pi} \sum_{m=-\infty}^{\infty} \rho_m\right) x^{-2d} \quad \text{as } x \downarrow 0. \tag{6.14}$$

If the correlations of the original process $\mathbf{Y}$ do not add up to zero (i.e. if (6.12) holds), then the asymptotic behavior at infinity of the correlations of the fractionally integrated process and asymptotic behavior of its spectral density at the origin correspond, once again, to each other as (5.15) corresponds to (5.16).

To what extent can one integrate a stationary process that does not have a finite second moment, and what is the effect of existing negative dependence in the original process $\mathbf{Y}$? Here is one simple situation. Let $S_n^{(Y)} = Y_1 + \cdots + Y_n$, $n = 0,1,\ldots$ be the partial sum sequence of the process $\mathbf{Y}$. The rate of growth of the partial sum sequence depends both on the memory in the stationary process $\mathbf{Y}$ and on the marginal tails of the process. Assume that for some $\theta \in (0,1)$ there is $c>0$ such that

$$E|S_n^{(Y)}| \le cn^{\theta}, \quad n = 1,2,\ldots. \tag{6.15}$$

Recall that, if $\mathbf{Y}$ is a zero mean finite variance stationary process with absolutely summable correlations (or such that the series in (5.9) converges), then (6.15) holds with $\theta = 1/2$, while certain slow decay of correlations (or a pole of the spectral density at the origin) can guarantee (6.15), but with θ larger than $1/2$; see (5.6) and Proposition 5.2. On the other hand, for the Fractional Gaussian Noise with $0<H<1/2$, the relation (6.15) holds with θ smaller than $1/2$; see (2.10).

Proposition 6.2. Let a stationary process $\mathbf{Y}$ be such that (6.15) holds for some $0 < \theta < 1$. Then for any $0 < d < 1 - \theta$ the series (6.4) converges in L^1 and the resulting process $\mathbf{X}$ is a well defined stationary process.

Proof. We may consider the sum (6.4) for $n = 0$, and we may also reverse the time in the process $\mathbf{Y}$ noting that, marginally, the partial sums of the time reversed process have the same law as those of the original process. Using summation by parts we see that for $m, k \geq 1$,

$$\sum_{j=m+1}^{m+k} a_j Y_j = \sum_{j=m+1}^{m+k} (a_j - a_{j+1}) S_j^{(Y)} + a_{m+k} S_{m+k}^{(Y)} - a_{m+1} S_m^{(Y)}.$$

The assumption $d < 1 - \theta$ together with (6.15) and (6.3) shows that $a_n S_n^{(Y)} \to 0$ in L^1 as $n \to \infty$. Therefore, the last two terms in the above relation converge to zero in L^1 as $m \to \infty$ uniformly in k. Similarly, for some $c > 0$ we have by (6.15) and (6.8),

$$E \left| \sum_{j=m+1}^{m+k} (a_j - a_{j+1}) S_j^{(Y)} \right| \leq c \sum_{j=m+1}^{\infty} j^{-(2-d)} j^{\theta} \to 0$$

as $m \to \infty$ because $d < 1 - \theta$. This shows the L^1 convergence, and stationarity is obvious. □

It is interesting to note that, if $\mathbf{Y}$ is a sequence of i.i.d. zero mean random variables in the domain of attraction of an α-stable law with $1 < \alpha < 2$, then (6.15) holds with any $\theta < 1/\alpha$ (see e.g., [53]), and so by Proposition 6.2 such sequences can be integrated up to the order $1 - 1/\alpha$. I.i.d. sequences with even fatter tails (e.g., in the domain of attraction of an α-stable law with $0 < \alpha \leq 1$) cannot be fractionally integrated at all! However, assuming appropriate negative dependence, even "very fat tailed" stationary processes can be integrated up to some order (with the series (6.4) also converging in an appropriately weaker sense).

It is clear that much of the previous discussion depends on little more than the asymptotic order of the magnitude of the coefficients in

the infinite series (6.4) and their differences. The specific choice arising from fractional differencing is attractive both because of its intuitive meaning and because of parsimony arising from dependence on a single parameter $0 < d < 1$. An example of a different moving average with regularly varying coefficients is in [11], where the rate of decay of coefficients is shown to affect the rate of growth of the R/S statistic.

Fractionally integrated models, especially FARIMA models, have found numerous applications in economics and econometrics; two examples are [36] and [55]. In this area of application one would like to combine fractional integration with the so-called *clustering of volatility*, or *conditional heteroscedasticity*. The standard (but non-fractional) model with clustering of volatility is the Generalized AutoregRessive Conditionally Heteroscedastic (or GARCH) process, introduced in [51] in its original (non-generalized) form and generalized by Bollerslev [20]. A possible way of introducing clustering of volatility into a fractionally integrated model is to start with a process $\mathbf{Y}$ in (6.4) that has the clustering of volatility property, for example with a GARCH process. This approach is suggested in [61]. Even earlier on, an alternative model was suggested by Baillie et al. [12]. This model directly combines fractional differencing/integration with the recursion for computation of the conditional variance of each subsequent observation, and has become known as a Fractionally Integrated GARCH (or FIGARCH) model. This model has proved difficult to analyze; even existence of a stationary version of the model that has desired properties is an issue. Recent progress has been made in [46]; see also [149].

7

Self-Similar Processes

Recall that a stochastic process $\mathbf{Y} = (Y(t), t \geq 0)$ is called self-similar if there is H such that for all $c > 0$ one has

$$(Y(ct), t \geq 0) \stackrel{\mathrm{d}}{=} (c^H Y(t), t \geq 0)\,.$$

The number H is alternatively referred to as the exponent of self-similarity, the scaling exponent, or the Hurst exponent. If $X_i = Y(i) - Y(i-1)$, $i = 1, 2, \ldots$ is the increment process of $\mathbf{Y}$, then the partial sum process $S_n = X_1 + \cdots + X_n$, $n = 1, 2, \ldots$, clearly satisfies for $n \geq 1$,

$$S_n = Y(n) - Y(0) \stackrel{\mathrm{d}}{=} n^H \big(Y(1) - Y(0)\big) = n^H S_1\,. \tag{7.1}$$

If the process $\mathbf{Y}$ also has stationary increments, then the process $\mathbf{X} = (X_1, X_2, \ldots)$ is stationary, and then (7.1) shows that the scaling exponent H determines the *distributional* order of magnitude of the partial sum process of the stationary model $\mathbf{X}$.

We have seen in Section 2 that the success of the Fractional Gaussian Noise in explaining the Hurst phenomenon is, at least in part, related to the fact that this stationary process is the increment process of the Fractional Brownian motion, a self-similar process, with the scaling exponent in the range $0 < H < 1$. Recall also that the correlations

of the Fractional Gaussian Noise are summable when $0 < H \leq 1/2$, and not summable when $1/2 < H < 1$ (cf. (2.9)), while the spectral density of the Fractional Gaussian Noise, given by (5.17), is continuous at the origin when $0 < H \leq 1/2$, while diverging at the origin when $1/2 < H < 1$. It is, therefore, attractive to consider the class of stationary models given as increments of general self-similar processes with stationary increments, and to call these stationary processes long range dependent if the scaling exponent is large enough; see e.g., [14] and [140].

This program has the advantage of being applicable to stationary processes with or without finite second moment. The boundary between short and long memory is, further, given by a single number — a certain critical value of the scaling exponent. This last feature is also a drawback of the approach: a single number does not usually represent well the dependence structure of a stochastic process, despite the example of certain Gaussian models. Another drawback of this approach is that a reasonably limited family of the models is thus considered — the increments of self-similar stationary increments processes. To overcome this one can distinguish between *exactly self-similar* models as above, and those that are only self-similar in a certain asymptotic sense; see e.g., [85]. This class of models has become subject of intense research since it was pointed out in [82] that Ethernet traffic data have features strikingly in common with certain models of this type and a logically attractive explanation of the connection of the network traffic to self-similarity was offered in [146]. Models arising from self-similar processes have also been used in risk theory (see e.g., [97]) and finance (see e.g., [33]).

Attractiveness of using the increment processes of self-similar processes with stationary increments as "canonical" models with shorter or longer types of memory is particularly obvious because such processes turn out to be *the only possible weak limit* in a common class of limiting procedures. Specifically, let $(U(t), t \geq 0)$ be a stochastic process, and $a_n \uparrow \infty$ be a sequence of positive numbers. If

$$\left(\frac{1}{a_n} U(nt), t \geq 0\right) \Rightarrow \big(Y(t), t \geq 0\big) \tag{7.2}$$

in terms of finite-dimensional distributions, and the limiting process $\mathbf{Y}$ is non-degenerate in the sense that $P(Y(t) \neq 0) > 0$ for all $t > 0$, then $\mathbf{Y}$ is H-self-similar for some $H > 0$ (and the sequence (a_n) is automatically regularly varying with exponent H). This was proved (in a slightly different form) by Lamperti [80], and is often referred to as the Lamperti theorem. Since great many of the limiting results in probability theory and its applications can be formulated in the form (7.2), it is not surprising the self-similar models are ubiquitous. If, in addition, the process $\mathbf{U}$ in (7.2) has stationary increments (as often happens in applications), the limiting process $\mathbf{Y}$ will have stationary increments as well. Furthermore, the "type" of memory the increments of the process $\mathbf{U}$ have, often translates into the "type" of memory that the increments of the limiting self-similar process $\mathbf{Y}$ possess. For example, the strong mixing properties in Theorems 4.1 and 4.2 above is the sort of short range dependence that guarantees that any memory completely disappears in the limit, which in both theorems is a Brownian motion, that has independent increments. On the other hand, there are examples of processes $\mathbf{U}$ whose memory is so strong that it persists in the limit; see e.g., [45] and [136]. Then the limiting self-similar process $\mathbf{Y}$ is not a Brownian motion; some of the possible limiting processes are discussed below, and their increments can be strongly dependent. The results of the type (7.2) where the limiting process $\mathbf{Y}$ is different from the Brownian motion are often referred to as non-central limit theorems.

Many facts on self-similar processes can be found in Chapter 7 of [128] and in a recent book of Embrechts and Maejima [50].

Let $\mathbf{Y} = (Y(t), t \geq 0)$ be a self-similar process with stationary increments (commonly abbreviated to an SSSI process). There are restrictions on the feasible values of the scaling exponent H. It is immediate that the only self-similar process with $H < 0$ is the trivial zero process $Y(t) = 0$ a.s. for each $t \geq 0$. The value $H = 0$ of the scaling exponent does allow some non-trivial SSSI processes (the process for which $Y(t_j), j = 1, \ldots, k$ are i.i.d. for any $t_1, \ldots, t_k$ and $k = 1, 2, \ldots$ is an example), but, assuming that $\mathbf{Y}$ has a measurable version, leaves only the constant process $Y(t) = Y(1)$ a.s. for each $t \geq 0$ as a possibility [143]. In modeling one assumes a positive scaling exponent H, as we will

do from now on. This assumption, clearly, means that $Y(0) = 0$ a.s. We will assume in the sequel that we are not dealing with the trivial zero process.

Further restrictions on the value of the scaling exponent of an SSSI process $\mathbf{Y}$ are related to finiteness of the marginal moments of the process. For example, suppose that for some $0 < \gamma < 1$ we have $E|Y(1)|^\gamma < \infty$. The assumption that $Y(1) \neq 0$ with positive probability implies that for n large enough, on a set of positive probability, at least 2 of the variables in a finite stationary sequence $(Y(1), Y(2) - Y(1), \ldots, Y(n) - Y(n-1))$ are different from zero at the same time. Then by the self-similarity and stationarity of the increments

$$\begin{aligned} n^{\gamma H} E|Y(1)|^\gamma = E|Y(n)|^\gamma &= E\left|\sum_{j=1}^{n} \big(Y(j) - Y(j-1)\big)\right|^\gamma \\ &< \sum_{j=1}^{n} |Y(j) - Y(j-1)|^\gamma = nE|Y(1)|^\gamma , \end{aligned}$$

which implies that $H < 1/\gamma$. In particular, the finite mean assumption $E|Y(1)| < \infty$ implies that also $E|Y(1)|^\gamma < \infty$ for all $0 < \gamma < 1$, and so we must have $H < 1/\gamma$ for all such γ which is, clearly, equivalent to $H \leq 1$. Summarizing,

$$\begin{cases} H < \frac{1}{\gamma} & \text{if } E|Y(1)|^\gamma < \infty \quad \text{for } 0 < \gamma < 1 \\ H \leq 1 & \text{if } E|Y(1)| < \infty. \end{cases} \tag{7.3}$$

In fact, the only SSSI process with a finite mean for which $H = 1$ is *the straight line process* for which $Y(t) = tY(1)$ a.s. for every $t > 0$, as the following argument (due to [143]) shows. By self-similarity, $Y(n)/n \stackrel{\mathrm{d}}{=} Y(1)$ for all $n \geq 1$. By the stationarity of the increments and ergodic theorem

$$\frac{Y(n)}{n} = \frac{1}{n}\sum_{j=1}^{n} \big(Y(j) - Y(j-1)\big) \to E(Y(1)|\mathcal{I})$$

with probability 1, where $\mathcal{I}$ is the invariant σ-field for the increment process $\mathbf{X}$. Therefore, $Y(1) \stackrel{\mathrm{d}}{=} E\big(Y(1)\big|\mathcal{I}\big)$. As Smit showed in [130],

this implies that $Y(1)$ is measurable with respect to the completion of $\mathcal{I}$, and so $Y(1) = Y(n) - Y(n-1)$ a.s. for all $n \geq 1$, implying that $Y(t) = tY(1)$ a.s. for every $t = 1, 2, \ldots$. Now one can use self-similarity to extend this relation first to $t = 1, 1/2, 1/3, \ldots$, then to all rational $t > 0$ and, finally, by the continuity in probability, which all SSSI processes with $H > 0$, clearly, possess, to all $t > 0$.

Non-trivial finite mean SSSI models exist, therefore, only for $0 < H < 1$, and we will, correspondingly, restrict ourselves to that range when the mean is finite. Since self-similarity forces $EY(n) = n^H EY(1)$, while stationarity of the increments implies $EY(n) = nEY(1)$, the non-trivial finite mean SSSI models must have zero mean.

Suppose that $\mathbf{Y}$ is a zero mean *finite variance* SSSI process with $0 < H < 1$. Denoting $\sigma^2 = EY(1)^2$, we immediately see that for all $s, t \geq 0$

$$\mathrm{Cov}\big(Y(s), Y(t)\big) = \frac{\sigma^2}{2}\left[t^{2H} + s^{2H} - |t-s|^{2H}\right], \tag{7.4}$$

so the self-similarity and stationarity of the increments uniquely determine the correlation function of any such process, which is then also the correlation function of the Fractional Brownian motion introduced in Section 2. It turns out that for any $0 < H < 1$ the expression in the right-hand side of (7.4) is, in fact, nonnegative definite and, hence, a legitimate covariance function. This can be demonstrated by simply exhibiting a Gaussian process whose covariance function is given by the right-hand side of (7.4).

Let $(B(t), t \in \mathbb{R})$ be the standard Brownian motion. Choose a real number $1/2 < \gamma < 1$, $\gamma \neq 3/2 - H$, and define a stochastic process by

$$B_H(t) = \frac{\sigma}{C(H,\gamma)} \int_{-\infty}^{\infty} \left(\int_x^{\infty} (v-x)^{-\gamma} \big(|v|^{H+\gamma-3/2} - |v-t|^{H+\gamma-3/2} \big) dv \right) B(dx), \tag{7.5}$$

where

$$C(H,\gamma) = \left(\int_{-\infty}^{\infty} \left(\int_x^{\infty} (v-x)^{-\gamma} (|v|^{H+\gamma-3/2} - |v-1|^{H+\gamma-3/2}) dv \right)^2 dx \right)^{1/2}.$$

This is a well defined centered Gaussian process, and its covariance function is easily checked to be given by the right-hand side of (7.4) (independently of γ). Since that characteristic function has the property $\mathrm{Cov}\big(Y(cs),Y(ct)\big) = c^{2H}\mathrm{Cov}\big(Y(s),Y(t)\big)$ for all $c>0$, we conclude that a centered Gaussian process with that characteristic function is, in fact, self-similar with exponent H. Since (7.4) is equivalent to the incremental variance statement $E\big(Y(t)-Y(s)\big)^2 = \sigma^2|t-s|^{2H}$, it implies, for a Gaussian process, stationarity of the increments as well. Therefore, we have constructed in (7.5) a Fractional Brownian motion, which is then the only SSSI Gaussian process. For $H=1/2$ and $0<\gamma<1$ (7.5) gives different representations of the standard Brownian motion.

Other important finite variance SSSI processes, different from the Fractional Brownian motion, can be represented as multiple Wiener-Itô integrals with respect to the Brownian motion; we refer the reader to [88] or Section 1.1.2 in [108] for basic information on the multiple integrals. For $k=1,2,\ldots$ and $1/2<\gamma<1/2+1/(2k)$, $H+k\gamma\neq 1+k/2$, define

$$Y^{(k)}(t) = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} Q_t^{(k)}(x_1,\ldots,x_k)\,B(dx_1)\cdots B(dx_k) \tag{7.6}$$

for $t\geq 0$, where $(B(t),t\in\mathbb{R})$ is still the standard Brownian motion, and the kernel $Q_t^{(k)}$ is defined by

$$\begin{aligned} Q_t^{(k)}(x_1,\ldots,x_k) = \int_{\max\{x_1,\ldots,x_k\}}^{\infty} &\prod_{j=1}^{k}(v-x_j)^{-\gamma} \\ &\times\big(|v|^{H+k\gamma-1-k/2} - |v-t|^{H+k\gamma-1-k/2}\big)dv\,. \end{aligned} \tag{7.7}$$

This process is mentioned in [104], following a similar process introduced in [117] (for $k=2$). It is, obviously, a generalization of the Fractional Brownian motion in (7.5). If the latter can be viewed as a linear functional of the sample paths of the Brownian motion, the process in (7.6) can be viewed as a polynomial functional of order k of these sample paths. The fact that the process $\mathbf{Y}$ in (7.6) is well defined follows from the fact that

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} Q_t^{(k)}(x_1,\ldots,x_k)^2\,dx_1\cdots dx_k<\infty \quad \text{for all } t\geq 0; \tag{7.8}$$

the verification is standard, if somewhat tedious. It can be easily checked that, in addition, the kernel $Q_t^{(k)}$ has also the following properties. For all $0 \le s < t$ and $c > 0$

$$Q_t^{(k)}(x_1,\dots,x_k) - Q_s^{(k)}(x_1,\dots,x_k) = Q_{t-s}^{(k)}(x_1 - s,\dots,x_k - s), \quad (7.9)$$

and

$$Q_{ct}^{(k)}(cx_1,\dots,cx_k) = c^{H-k/2} Q_t^{(k)}(x_1,\dots,x_k) \quad (7.10)$$

for almost all $(x_1,\dots,x_k)$. Every stochastic process given in the form (7.6) with the functions $(Q_t^{(k)}, t \ge 0)$ satisfying (7.9), has stationary increments, and every stochastic process given in the form (7.6) with the functions $(Q_t^{(k)}, t \ge 0)$ satisfying (7.10) is self-similar with exponent of self-similarity H. Both statements are heuristically obvious when one makes the appropriate change of variable in the defining multiple integral in (7.6), and uses the stationary of the increments of the Brownian motion, and its self-similarity with exponent $1/2$. This argument can be made precise by approximating the kernel $Q_t^{(k)}$ by simple symmetric kernels.

Therefore, the stochastic process defined in (7.6) with the kernel given by (7.7) is SSSI, with the scaling exponent H. As all other multiple Wiener-Itô integrals with respect to the Brownian motion, it has finite moments of all orders. It shares with the Fractional Brownian motion its correlation function, but is not a Gaussian process if $k \ge 2$.

It is not difficult to check that the properties (7.9) and (7.10) (with $k = 1$) of the kernel $Q_t^{(1)}(x)$ in the representation (7.5) of the Fractional Brownian motion imply that for $0 < H < 1$, $H \ne 1/2$,

$$\begin{aligned} Q_t^{(1)}(x) \;=\; g_t(c_1,c_2;H;x) &:= c_1\Big[\big((t-x)_+\big)^{H-1/2} - \big((-x)_+\big)^{H-1/2}\Big] \\ &\quad + c_2\Big[\big((t-x)_-\big)^{H-1/2} - \big((-x)_-\big)^{H-1/2}\Big], \end{aligned} \quad (7.11)$$

where $a_+ := \max(a,0)$ is the positive part of a real number a, and $a_- := \max(-a,0)$ is its negative part, and 0^a is interpreted as 0 for all $a \in \mathbb{R}$. Here $c_i = c_i(H,\gamma)$, $i = 1,2$, are real numbers; in fact, one can start with choosing c_1 and c_2 in such a way that $Q_1^{(1)}(-1) = g_1(c_1,c_2;H;-1)$ and $Q_1^{(1)}(2) = g_1(c_1,c_2;H;2)$, and then show that the equality extends to

all $t \geq 0$ and $x \neq 0, t$. For $H = 1/2$ a similar argument shows that

$$Q_t^{(1)}(x) = g_t(c_1, c_2; 1/2; x) := c_1 \mathbf{1}_{[0,t]}(x) + c_2\big(\log|t-x| - \log|x|\big), \tag{7.12}$$

once again for some real $c_i = c_i(\gamma)$, $i = 1, 2$.

In fact, alternative representations of the Fractional Brownian motion (up to a multiplicative constant) are obtained via

$$B_H(t) = \int_{-\infty}^{\infty} g_t(c_1, c_2; H; x)\, B(dx), \quad t \geq 0$$

for *arbitrary* real c_1, c_2. These are the so called *moving average representations* of the Fractional Brownian motion, originating with [92]; see Section 7.2.1 in [128].

Moving average representations of the Fractional Brownian motion different from the representation (7.5) can themselves be extended to SSSI processes represented by multiple Wiener-Itô integrals. For example, take $c_2 = 0$ in (7.11). In the case $1/2 < H < 1$ one can rewrite the resulting expression in an equivalent form, and extend it, leading to a family of processes

$$Y^{(k)}(t) = \int_{-\infty}^{t} \cdots \int_{-\infty}^{t} \left(\int_0^t \prod_{j=1}^{k} \big((v - x_j)_+\big)^{-(1/2+(1-H)/k)} dv \right) B(dx_1) \cdots B(dx_k), \quad t \geq 0, \tag{7.13}$$

introduced in [137]; it appeared as a limit in a "non-central limit theorem" in [138] and, in a more general situation, in [133] (see also [9]). In the case $0 < H < 1/2$ a similar procedure leads to a family of processes

$$Y^{(k)}(t) = \int_{-\infty}^{t} \cdots \int_{-\infty}^{t} \left(\int_t^{\infty} \prod_{j=1}^{k} (v - x_j)^{-(1/2+(1-H)/k)} dv - \mathbf{1}\big(\max(x_1, \ldots, x_k) < 0\big) \int_0^{\infty} \prod_{j=1}^{k} (v - x_j)^{-(1/2+(1-H)/k)} dv \right) B(dx_1) \cdots B(dx_k), \quad t \geq 0. \tag{7.14}$$

One can check that in both cases the kernels in the multiple integrals satisfy (7.8), (7.9), and (7.10) and, hence, the processes defined in (7.13) and (7.14) are SSSI processes, with the corresponding scaling exponent H. In fact, for $k \geq 2$ the process given in (7.14) is well defined for all $0 < H < 1$.

The SSSI processes with representations as in (7.6), (7.13), and (7.14) are examples of such processes in *the* kth *Gaussian chaos*, in the terminology of [144]. If, for $k \geq 1$, $\mathbf{Y}^{(k)}$ has the representation (7.6), with the kernels $Q_t^{(k)}$ satisfying, for each k, (7.8), (7.9), and (7.10), then for any sequence of constants (a_k) such that

$$\sum_{k=1}^{\infty} a_k^2 \, k! \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} Q_t^{(k)}(x_1, \ldots, x_k)^2 \, dx_1 \cdots dx_k < \infty \,,$$

the new process

$$\begin{aligned} Y(t) &= \sum_{k=1}^{\infty} a_k Y_k(t) \\ &= \sum_{k=1}^{\infty} a_k \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} Q_t^{(k)}(x_1, \ldots, x_k) \, B(dx_1) \cdots B(dx_k) \end{aligned} \tag{7.15}$$

is a well defined second-order stochastic process (see [108]), and the previous argument using stationarity of the increments of the Brownian motion and its self-similarity implies that this process is SSSI, with exponent H of self-similarity. Of course, this process is no longer, in general, a polynomial-like functional of the Brownian motion.

Yet more representations of the Fractional Brownian motion exist; see e.g., [39] and [107]. We mention only one more, as it has an impact on our discussion of long range dependence. Let $(B^{(j)}(t), t \geq 0)$ for $j = 1, 2$ be independent standard Brownian motions, and extend $\mathbf{B}^{(1)}$ to the entire real line as an even function, and $\mathbf{B}^{(2)}$ as an odd function. For $0 < H < 1$ the following is, up to a multiplicative constant, a representation of the Fractional Brownian motion:

$$B_H(t) = \int_{-\infty}^{\infty} \frac{e^{itx} - 1}{ix} |x|^{-(H-1/2)} \, \tilde{B}(dx), \quad t \geq 0 \,, \tag{7.16}$$

where $\tilde{B}(t) = B^{(1)}(t) + iB^{(2)}(t), t \in \mathbb{R}$. This is the so called *harmonizable representation* of the Fractional Brownian motion; its origins go

back to [75] and [148]. The harmonizable representation also has a natural extension to a SSSI process in the kth Gaussian chaos for $k = 2, 3, \ldots$, appearing (as a limit) in [45], who used the techniques introduced in [44]. The obtained processes have been showed by Taqqu [138] to coincide, in the case $1/2 < H < 1$, with those defined in (7.13) via a moving average representation.

The above discussion presented a large number of SSSI processes with a finite variance and exponent of self-similarity $H \in (0, 1)$; even more can be obtained by replacing some of the repeated Brownian motion in the multiple integrals by independent copies of a Brownian motion, or by more generally correlated Brownian motions. The increments of each one of them form a stationary process, whose covariance function coincides with that of the Fractional Gaussian Noise given in (2.8). One often says that these stationary processes have long range dependence if $1/2 < H < 1$ (and short memory if $0 < H \leq 1/2$); see e.g., [14], [50] and [145]. This is, of course, entirely understandable from the point of view of the rate of decay of correlations, as in (5.15), or from the point of view of the behavior of the spectral density at the origin, as in (5.16), or from the point of view of the rate of increase of the variance of the partial sums, as in (5.6). While no further justification seems to be necessary for the Fractional Gaussian noise, the increment process of the Fractional Brownian motion; for the other models the second-order measures provides, of course, only partial information. A further important point is the distributional rate of growth of the partial sums of the increment processes: for $1/2 < H < 1$ this rate of growth is above what is allowed for a central limit theorem and convergence to a Brownian motion.

Notice that, if $\mathbf{Y}$ is an SSSI process with an exponent H of self-similarity, and A is a random variable independent of $\mathbf{Y}$, then the process

$$Z(t) = AY(t), \quad t \geq 0 \tag{7.17}$$

is also an SSSI process with the same scaling exponent H is $\mathbf{Y}$. If $\mathbf{Y}$ is a finite variance process, and A also has a finite variance, then the resulting SSSI process in (7.17) will have a finite variance as well. In particular, its increment process will have exactly the same second-order

properties as the Fractional Gaussian noise, and it will have the same distributional rate of growth of the partial sums as the latter. However, apart from a small number of degenerate situations, the increments of $\mathbf{Z}$ will be non-ergodic and, hence, will arguably have infinite memory, *regardless of the value of the scaling exponent* H.

This is, of course, not the situation with the SSSI processes with a chaos representation as in (7.15). The increment processes of the latter are always ergodic because of the property (7.9) and ergodicity of Bernoulli shifts applied to the increments of the Brownian motion (see Section 1.4 in [78]). However the example of the process (7.17) emphasizes the limited amount information provided by the scaling exponent alone.

The most common *infinite variance* SSSI processes are α-stable processes, $0 < \alpha < 2$; we refer the reader to [128] for information on such models. For an α-stable random variable X one has power-like tails: $P(|X| > x) \sim cx^{-\alpha}$ as $x \to \infty$ for some $c > 0$; this implies that the mean is finite in the case $1 < \alpha < 2$, and infinite if $0 < \alpha \leq 1$. We will consider here *symmetric* α-stable (SαS) SSSI processes.

There are many similarities between finite variance Gaussian SSSI models and processes related to them, and SαS and related SSSI processes, the most important of which is the fact that both arise in a number of natural limit theorems. In the finite mean case $1 < \alpha < 2$ the exponent of self-similarity of any SαS SSSI process is still restricted (to avoid trivialities) to the range $0 < H < 1$, while in the infinite mean case $0 < \alpha \leq 1$, the tail behavior of the marginal distributions restricts the scaling exponent of a SαS SSSI process to the range $0 < H \leq 1/\alpha$. In a significant departure from the Gaussian case, where the exponent of self-similarity determines the correlation function and, hence, the law of the SSSI process (up to a multiplicative constant), for every feasible pair (α, H) of the index of stability and scaling exponent, there are generally *many different* SαS SSSI models; some of them will be discussed below. The only exception is the case $0 < \alpha < 1$, $H = 1/\alpha$, which corresponds to a single process, the SαS Lévy motion; see [127]. It is common to use the increments of certain SαS SSSI processes as canonical heavy tailed models with long range dependence.

A SαS Lévy process (motion) $(Y(t), t \geq 0)$ is the heavy tailed equivalent of the Brownian motion, a process with stationary and independent increments. It is self-similar, and its scaling exponent is $H = 1/\alpha$. No other (symmetric) Lévy processes are self-similar; see [129]. When deciding which SαS SSSI processes should be said to have long range dependent increments, the Gaussian case has been often taken as guidance. This means using the SαS Lévy motion as the benchmark, and viewing the processes with $H > 1/\alpha$ as long range dependent, see e.g., [2], [77] and [132]. Unfortunately, this range of the exponent of self-similarity is only possible when $\alpha > 1$. A number of limit theorems in which α-stable SSSI processes appear have been established, most of them apply to partial sums of linear infinite order moving average processes; in situations where long memory is believed to be present the exponent of self-similarity of the resulting SSSI process turned out to be in the range $H > 1/\alpha$. See [7], [73], [87] or [83]. A continuous-time version for a shot noise model is in [56].

Let $\mathbf{Y}$ be an SαS SSSI process, and $X_n = Y(n) - Y(n-1)$ for $n = 1, 2, \ldots$ be its stationary increment process. Since α-stable processes with $0 < \alpha < 2$ have infinite variance, it is impossible to relate the case $H > 1/\alpha$ to a slow decay of correlations of $\mathbf{X}$. However, when $H > 1/\alpha$, the partial sums $S_n = X_1 + \cdots + X_n = Y(n)$, $n = 1, 2, \ldots$, grow distributionally at the rate n^H, larger that the "usual" rate of $n^{1/\alpha}$ (times a slowly varying function) associated with the heavy tailed version of the Functional Central Limit Theorem, where the limit is the α-stable Lévy motion; see e.g., [49] (for interesting topological difficulties that may arise see [10]). While for $1 < \alpha < 2$ the benchmark $H = 1/\alpha$ is in the middle of the feasible range $0 < H < 1$ of the exponent of self-similarity, in the case $0 < \alpha \leq 1$ it is its *right endpoint*. This, of course, means that the rate of $n^{1/\alpha}$ is the fastest possible rate at which the partial sums of the increment process of an SαS SSSI process with such an index of stability α can grow; in fact, the partial sums of *any* stationary SαS process with $0 < \alpha \leq 1$ can grow at most at the rate of $n^{1/\alpha}$. However, it means that, according to the rule $H > 1/\alpha$, no SαS SSSI process can have long range dependent increments. This is, clearly, unfortunate.

Most of SαS SSSI processes discussed in the extensive literature on the subject are constructed with, once again, guidance from the

Gaussian case. One starts, typically, with a representation of the Fractional Brownian motion and modifies it appropriately. The best known SαS self-similar process with stationary increments originates with the moving average representation (see (7.11)) of the latter. Let $(L(t), t \in \mathbb{R})$ be an SαS Lévy motion. Choose $0 < H < 1$ (with $H \neq 1/\alpha$ if $1 < \alpha < 2$) and define for real c_1, c_2

$$Y(t) = \int_{-\infty}^{\infty} Q_t(x)\, L(dx)\,, \quad t \geq 0\,, \tag{7.18}$$

where

$$\begin{aligned} Q_t(x) = c_1 \Big[\big((t-x)_+\big)^{H-1/\alpha} - \big((-x)_+\big)^{H-1/\alpha} \Big] \\ + c_2 \Big[\big((t-x)_-\big)^{H-1/\alpha} - \big((-x)_-\big)^{H-1/\alpha} \Big]\,. \end{aligned} \tag{7.19}$$

This is a well defined SαS process. The kernel defined by (7.19) satisfies, for $0 \leq s < t$ and $c > 0$

$$Q_t(x) - Q_s(x) = Q_{t-s}(x-s) \tag{7.20}$$

and

$$Q_{ct}(cx) = c^{H-1/\alpha} Q_t(x) \tag{7.21}$$

for almost all $x \in \mathbb{R}$. It is clear that the intuition we used in the Gaussian case still works here: the properties (7.20) and (7.21) of the kernel imply that the process in (7.18) has stationary increments and is self-similar (and, once again, the argument can be made precise). The SSSI process defined by (7.18) with the kernel given by (7.19) is called *Linear Fractional Stable Motion*. This process originates in [86] and [141]. It is a general phenomenon that the stable integrals are much more "rigid" than similarly looking Gausssian integrals. Whereas any choice of the constants c_1 and c_2 in (7.11) produces, up to a multiplicative constant, a representation of the same Fractional Brownian motion, different pairs (c_1, c_2) in (7.19) will produce *different* Linear Fractional Stable Motions, unless these parameters are proportional (see [30] and [126]).

If $1 < \alpha < 2$ and $H = 1/\alpha$, the process corresponding to the Linear Fractional Stable Motion is given in the form (7.18) with the kernel

$$Q_t^{(1)}(x) = c_1 \mathbf{1}_{[0,t]}(x) + c_2 \big(\log|t-x| - \log|x| \big)\,, \tag{7.22}$$

for real c_1, c_2, which is, of course, identical to (7.12). When $c_2 = 0$ this gives back the SαS Lévy motion, for $c_1 = 0$ we obtain the so called *Log-fractional Stable Motion* introduced by Kasahara et al. in [73]; it is easily seen not to have independent increments.

In the case $1 < \alpha < 2$ it is also possible to start with the kernel in a representation of the Fractional Brownian motion as given in (7.5). The corresponding kernel in the α-stable case is

$$Q_t(x) = \int_x^\infty (v-x)^{-\gamma} \big(|v|^{H+\gamma-1-1/\alpha} - |v-t|^{H+\gamma-1-1/\alpha} \big) dv$$

for $0 < H < 1$ and $1/\alpha < \gamma < 1$, $\gamma \neq 1 + 1/\alpha - H$. This kernel satisfies (7.20) and (7.21) and, hence, can be shown to coincide with (7.19) when $H \neq 1/\alpha$, and with (7.22) when $H = 1/\alpha$, in both cases for some c_1 and c_2 depending on γ.

One can also start with the harmonizable representation (7.16) of the Fractional Brownian motion and extend it to the α-stable case, $0 < \alpha < 2$. This is usually done by starting with a complex-valued *isotropic* SαS Lévy motion $\tilde{M}$ (see [128] for details) and defining

$$Y(t) = \mathcal{R}e \int_{-\infty}^{\infty} \frac{e^{itx} - 1}{ix} |x|^{-(H-1+1/\alpha)} \tilde{M}(dx), \quad t \geq 0\,. \tag{7.23}$$

This process is often referred to as *Harmonizable Fractional Stable Motion*, and it was introduced in [30]. The Harmonizable Fractional Stable Motion is *a different process* from the Linear Fractional Stable Motion; in certain cases this follows from [31], more generally for $1 < \alpha < 2$ this is in [30], and in full generality with $0 < \alpha < 2$ it is in Chapter 7 of [128]. In fact, the stationary increment process of the latter is a mixing stationary process (this is implicit in [28] and explicit in [135]), while the increment process of the former is not even ergodic (this statement is in [28], and it also follows from the fact that real *stationary* harmonizable processes have a representation as mixtures of stationary Gaussian processes, discovered by Marcus and Pisier [95]).

The above classes of SαS SSSI processes can be viewed as linear functionals of SαS Lévy motions in their integral representations. Analogously to the Gaussian case, new SSSI models can be constructed as

polynomial-type functions of SαS Lévy motions, as multiple stochastic integrals, i.e., as stochastic processes of the form:

$$Y^{(k)}(t) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} Q_t^{(k)}(x_1, \ldots, x_k)\, L(dx_1) \cdots L(dx_k) \tag{7.24}$$

for $t \geq 0$, where $(L(t), t \in \mathbb{R})$ is an SαS Lévy motion. The conditions on the kernel $Q_t^{(k)}$ for the integral in (7.24) to exist are "more complicated" in the stable case than in the Gaussian case. A sufficient condition for integrability is

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left|Q_t^{(k)}(x_1, \ldots, x_k)\right|^{\alpha} \left(\log_+ \frac{|Q_t^{(k)}(x_1, \ldots, x_k)|^{\alpha}}{\psi(x_1) \cdots \psi(x_k)} \right)^{k-1} dx_1 \cdots dx_k < \infty, \quad t \geq 0, \tag{7.25}$$

for a strictly positive probability density ψ on $\mathbb{R}$, where $\log_+ x = \log x$ for $x > 1$, and $= 0$ otherwise. See [121]. Once the process is well defined, if the kernel $Q_t^{(k)}$ also satisfies the condition (7.9) and the following analog of (7.10): for all $c > 0$

$$Q_{ct}^{(k)}(cx_1, \ldots, cx_k) = c^{H-k/\alpha} Q_t^{(k)}(x_1, \ldots, x_k) \tag{7.26}$$

for almost all $(x_1, \ldots, x_k)$, then the process defined in (7.24) is SSSI, with exponent H of self-similarity.

A model with these properties was introduced in [134], and it is an SαS version of the process in (7.13). Assume that $1 < \alpha < 2$ and $H \in (1/\alpha, 1)$. Then the choice

$$Q_t^{(k)}(x_1, \ldots, x_k) = \int_0^t \prod_{j=1}^{k} \big((v - x_j)_+\big)^{-(1/\alpha + (1-H)/k)} dv \tag{7.27}$$

leads to a well defined SSSI process, which is, of course, a direct generalization of the multiple Wiener-Itô integral process in (7.13). This process appears as a limit in the "non-central limit theorem" setting (in the case $k = 2$), as shown in [6]. Similarly, the multiple Wiener-Itô integral process in (7.14) can be generalized to the α-stable multiple

integral situation as well, resulting in another SSSI model with

$$Q_t^{(k)}(x_1,\ldots,x_k)$$
$$= \mathbf{1}\big(\max(x_1,\ldots,x_k) < t\big) \int_t^\infty \prod_{j=1}^k (v - x_j)^{-(1/\alpha+(1-H)/k)} dv$$
$$- \mathbf{1}\big(\max(x_1,\ldots,x_k) < 0\big) \int_0^\infty \prod_{j=1}^k (v - x_j)^{-(1/\alpha+(1-H)/k)} dv\,. \quad (7.28)$$

For single integrals this is well defined only if $H < 1/\alpha$, but for $k \geq 2$ this process is well defined for all $0 < H < 1$ (for both (7.27) and (7.28) the condition (7.25) can be verified with, for example, ψ being the standard Cauchy density). We are not aware of limit theorems in which the process with the kernel as in (7.28) appears as the limit.

There is no doubt that other SSSI processes in the form of a finite order symmetric α-stable chaos can be defined, for example by extending yet other SαS processes, or by a direct analogy with the Gaussian case. Using the recipe (7.15) one can construct even more models (even though necessary and sufficient conditions on the sequence (a_k) for the series in (7.15) to converge when each SSSI process $\mathbf{Y}^{(k)}$ is in the kth SαS chaos do not seem to be known, it is obvious that the series will converge if the coefficients are "small enough"). Similarly to the Gaussian case, one can also replace some of the repeated SαS Lévy motions in a k-tiple integrals by the components of a k-dimensional SαS Lévy motion (see [129]). Unlike the Gaussian case, this last exercise can even be performed on certain SαS SSSI processes by, for example, integrating each one of the two parts in (7.19) or (7.22) with respect to different components of a bivariate SαS Lévy motion.

This provides for an even greater variety of SSSI processes with infinite variance, based on stochastic integrals with respect to SαS Lévy processes with $0 < \alpha < 2$, than of the finite variance models we considered above. In the case $1 < \alpha < 2$ the value $H = 1/\alpha$ is considered to be "the critical value" for exponent of self-similarity, with the range $(1/\alpha, 1)$ of H corresponding to long memory of the increment process. This, of course, cannot be justified any longer by looking at the change in the behavior of the covariance function of the increments, which is not defined now. Certain substitutes have been used, mostly for SSSI

processes that are themselves symmetric α-stable. Two such substitutes have appeared in literature, the *covariation* and the *codifference*.

Let X_1 and X_2 be jointly SαS random variables. Their joint characteristic function can be written in the form:

$$E^{i(\theta_1 X_1+\theta_2 X_2)} = \exp\left\{-\int_{S_2} |\theta_1 s_1 + \theta_2 s_2|^\alpha \Gamma(d\mathbf{s})\right\} \tag{7.29}$$

for real θ_1, θ_2, where S_2 is the unit circle, and Γ is a uniquely determined finite symmetric measure on the unit circle, the so called *spectral measure* of (X_1, X_2) see [128]. If $1 < \alpha < 2$, one defines the covariation of X_1 and X_2 by

$$[X_1, X_2]_\alpha = \int_{S_2} s_1 s_2^{\langle \alpha-1\rangle} \Gamma(d\mathbf{s}),$$

where for real a, b, the notation $a^{\langle b\rangle}$ stands for the signed power $|a|^b \mathrm{sign}(a)$. The covariation can also be defined for $\alpha = 1$, but it appears to be less useful in that case; an extension to the case $0 < \alpha < 1$ is only partially possible, and requires restrictions on the spectral measure of (X_1, X_2). The covariation is not, generally, symmetric in its arguments. It reduces to half the covariance in the Gaussian case $\alpha = 2$, if one chooses to write the characteristic function as (7.29) in that case (the spectral measure is not uniquely defined if $\alpha = 2$.) The notion of covariation was introduced in [101]. If $(X_n, n = 1, 2, \ldots)$ is a stationary SαS process, its covariation function can be defined via $\gamma(k) = [X_{n+k}, X_n]_\alpha$ for $k = 0, 1, 2, \ldots$ (but changing the order of the arguments will lead, in general, to a different function).

The codifference can be defined for any random variables and stochastic processes. For a random vector (X_1, X_2) we define

$$\tau(X_1, X_2) = \log Ee^{i(X_1 - X_2)} - \log Ee^{iX_1} - \log Ee^{-iX_2},$$

(where we take the continuous branch of the logarithm equal to zero at point 1). The codifference is equal to the covariance for a Gaussian vector (X_1, X_2). The term "codifference" appeared first in [74], but related notions had been used many times before. For a stationary process $(X_n, n = 1, 2, \ldots)$ its codifference function is defined by $\tau(k) = \tau(X_{n+k}, X_n)$ for $k = 0, 1, 2 \ldots$.

Both covariation and codifference are equal to zero in the case of independence, but the converse is not true. On the other hand, zero covariation does imply a kind of Banach space orthogonality, the so called *James orthogonality*; see [29]. Furthermore, it is possible to characterize independence of infinitely divisible random vectors based on codifference (see [123]), and for a certain class of stationary infinitely processes (including stable processes) mixing is equivalent to convergence the codifference function to zero at large lags (see [122]).

Suppose $\mathbf{Y}$ is an SSSI process, that is SαS with $1 < \alpha < 2$. In order to understand if there is a significant change in the properties of the increment process $X_n = Y(n) - Y(n-1),\ n = 1, 2, \ldots$ at the value $H = 1/\alpha$ of the exponent of self-similarity, one can try to see if the behavior of either the covariation function or codifference function changes significantly at that point.

The behavior of the codifference function of the increment process of the Linear Fractional Stable Motions as in (7.18) with (7.19), was considered in [8]. They discovered that

$$\tau(k) \sim \begin{cases} Ct^{-(1+\frac{1}{\alpha}-H)} & \text{if } 0 < H < 1 - \frac{1}{\alpha(\alpha-1)} \\ Ct^{-\alpha(1-H)} & \text{if } 1 - \frac{1}{\alpha(\alpha-1)} < H < 1, \ \ H \neq \frac{1}{\alpha} \end{cases}$$

as $t \to \infty$, where C is a constant depending on the parameters in (7.19). Here a change in the rate of decay of the codifference function does occur, but at the point $1 - 1(\alpha(\alpha - 1))$, and not at the point $1/\alpha$. A similar computation was performed by Kokoszka and Taqqu [74] for the FARIMA model with SαS noise, and the results for both the covariation function and the codifference function were similar (in the case $1 < \alpha < 2$) to the above.

It is not easy to evaluate the evidence provided by the behavior of the covariation and codifference functions. One expects it to be smaller than that provided by the covariances for the second-order stationary processes. However, even this available evidence does not necessarily point to a particular importance of the point $H = 1/\alpha$ when deciding whether or not the increments of an SαS SSSI processes are long range dependent or not.

In fact, the ergodic theory seems to provide a better guidance to the memory of the increment process than the rate of decay the covariation

and codifference functions. Recall that the increments of the Harmonizable Fractional Stable Motions are not ergodic, while the increments of the Linear Fractional Stable Motions are mixing. This already says that the memory of the latter is shorter than that of the former, and this is regardless of the value of the scaling exponent H. In fact, in studying the memory of any stationary SαS process the ergodic theory enters the picture in yet another, even more informative way, as will be seen in the next section.

8

Long Range Dependence as a Phase Transition

A different point of view on long range dependence was suggested by Samorodnitsky [124]. Suppose that we are given a family of shift-invariant probability measures $\big(P_\theta, \theta \in \Theta\big)$ on $\mathbb{R}^{\mathbb{Z}}$; that is, each P_θ describes the finite-dimensional distributions of a two-sided (for convenience) stationary stochastic process $\mathbf{X} = (\ldots, X_{-1}, X_0, X_1, X_2, \ldots)$. Assume that, as θ varies over the parameter space Θ, the one-dimensional marginal distributions of the process do not change significantly. This requirements allows, for example, a change in scale, or other changes not relevant for the application of interest. We do not usually want to allow a serious change in the marginal tails, for instance loss/gain of a finite variance. For example, $\big(P_\theta, \theta \in \Theta\big)$ might describe a family of correlations functions of unit variance stationary Gaussian processes, or a family of coefficients of an infinite moving average model. A subset Θ_0 of the parameter space corresponds to the choices of the parameters under which the process $\mathbf{X}$ is a sequence of i.i.d. random variables; sometimes Θ_0 is a singleton.

Let $\phi = \phi(\mathbf{X})$ be a measurable functional of the process; typical examples are the sequence of the partial sums, for which $\phi : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\infty}$, $\phi_n(\mathbf{x}) = \sum_{j=1}^{n} x_j$, and the sequence of the partial maxima, for which also $\phi : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\infty}$, but this time $\phi_n(\mathbf{x}) = \max_{j=1}^{n} x_j$. The behavior of

this functional will, in general, be different under different probability measures $P_\theta, \theta \in \Theta$.

Suppose that there is a partition $\Theta = \Theta_1 \cup \Theta_2$ of the parameter space into disjoint parts, with $\Theta_0 \subset \Theta_1$, such that the behavior of the functional ϕ changes significantly as the parameter crosses the boundary between Θ_1 and Θ_2. This means that, as long as the parameter stays within Θ_1, the behavior of the functional ϕ does not change much, and it remains the same as in the i.i.d. case $\theta \in \Theta_0$, perhaps "up to a multiplicative constant." Once the parameter θ crosses the boundary into Θ_2, there is a change "in the order of magnitude" in the behavior of the functional ϕ. Moreover, there continue to be significant changes as the parameter θ moves within Θ_2.

Under these conditions we view the part Θ_1 of the parameter space as corresponding to short memory models, and the part Θ_2 of the parameter space as corresponding to long memory models. From this point of view, the boundary between Θ_1 and Θ_2 is the boundary between short and long memory, and it is natural to regard the appearance of long range dependence as a phase transition.

This approach has drawbacks. It is "an approach," and not a rigorous definition. It ties the notion of long range dependence to a particular functional and, perhaps, to a particular aspect of the behavior of that functional.

This, however, appears to be inevitable. One of the reasons it has been so difficult to define long range dependence is that one has tried to give a single definition to what is, really, a series of phenomena. Moreover, studying the change in behavior of a functional relevant in applications is, arguably, more important than trying to find a single critical parameter. If one adopts this point of view on long range dependence, the problem reduces to that of finding critical boundaries. It will undoubtedly turn out that for many models there will be more than one such boundary.

A functional of major interest, and the one that historically generated most interest, is the sequence of the partial sums of the process. When considering stationary processes with a finite second moment, the second-order approach to long range dependence concentrates on the behavior of the variances of the partial sums.

If $\big(P_\theta, \theta \in \Theta\big)$ is the family of laws of all stationary stochastic processes with marginal variance equal to (say) 1, then for every $\theta \in \Theta_0$ (corresponding to the law of a sequence of i.i.d. random variables) the variance of the partial sums grows linearly with the sample size. It is natural then to define Θ_1 to be that subset of Θ such that, under the law P_θ with $\theta \in \Theta_1$, the variance of the partial sums grows at most linearly fast with the sample size. Then Θ_2, the complement of Θ_1 in Θ, is the collection of the laws of finite second-order stationary processes that will be considered as long range dependent, and the notion relies strictly on the variance of the partial sums growing, at least along a subsequence, faster than linearly fast. From this point of view, the various alternative notions of a short memory process discussed in Section 5: (5.3), (5.7), (5.9), or (5.14) are entirely reasonable when viewed as *sufficient conditions* for the law of the stochastic process to be one of P_θ with $\theta \in \Theta_1$, hence of short memory, but it is quite a bit less reasonable to view the failure of one of these conditions as an indication of long memory.

Similarly, the conditions (5.5), (5.6) or those of Proposition 5.2 can very reasonably be viewed as *sufficient conditions* for long memory, but their absence should not be viewed as an indication of short memory.

It is not by any means obvious that the change from situation where the variance of the partial sums grows at most linearly fast with the sample size, to the situation where this is not the case, is, by itself, important enough to justify calling this change a passage from short memory to long memory. The exception is, of course, the Gaussian case. If $\big(P_\theta, \theta \in \Theta\big)$ is the family of laws of all stationary *Gaussian* processes with marginal variance 1, then the analogous to the above partition of the parameter space Θ into Θ_1 and Θ_2 is a natural basis for distinction between short and long memory.

Still concentrating on the behavior of the partial sums of a stationary second-order process, one can partition the parameter space Θ into two parts, depending on whether or not the partial sums satisfy the invariance principle with convergence to the Brownian motion. From this point of view, the strong mixing conditions in Theorems 4.1 and 4.2 should be viewed as sufficient conditions for short memory, but their

absence should not be regarded as an indication of long memory. The Fractional Gaussian noise with *any* $H \neq 1/2$ (and not only $H > 1/2$) will then be considered to be long range dependent.

In principle, basing the decision on whether or not a given stationary process has long memory on the order of magnitude of the partial sums of the process is possible also for infinite variance processes. Even more specifically, suppose $\big(P_\theta, \theta \in \Theta\big)$ is the family of laws of all stationary stochastic processes whose one-dimensional marginal distributions are in the domain of attraction of an α-stable law with $0 < \alpha < 2$. Then for every $\theta \in \Theta_0$ (corresponding to the law of a sequence of i.i.d. random variables in the α-stable domain of attraction), the partial sums satisfy an invariance principle with convergence to an α-stable Lévy motion, and one can define a partition of the parameter space Θ into Θ_1 and Θ_2 depending on whether such an invariance principle still holds (an example is in [6] or [72]). From this point of view, the stationary increment process of any α-stable SSSI process, other than the strictly stable Lévy motion, has long memory *regardless of the value of the Hurst exponent* H, which can even be equal to $1/\alpha$ if $1 < \alpha < 2$; see the example of the Log-fractional Stable Motion in (7.22).

As the marginal tails of a stationary process become heavier, concentrating on the partial sums of the process, particularly on their rate of growth, to draw the boundary between short and long memory becomes less useful. The following proposition (whose proof will appear elsewhere) shows that, when the marginal tails are sufficiently heavy, the partial sums cannot grow faster than those of an i.i.d. sequence. For simplicity we state it in the symmetric case, but the statement holds in a much greater generality.

Proposition 8.1. Let X be a symmetric random variable such that $E|X|^\beta = \infty$ for some $0 < \beta < 1$. Let $\mathbf{X} = (X_1, X_2, \ldots)$ be a stochastic process with each $X_i \stackrel{\mathrm{d}}{=} X$, and let $\mathbf{Y} = (Y_1, Y_2, \ldots)$ be a sequence of independent copies of X. Let $a_n \uparrow \infty$ be a sequence of positive numbers such that

$$\limsup_{n\to\infty} \frac{a_{n+1}}{a_n} < \infty\,.$$

If

$$\limsup_{n\to\infty} P(|X_1 + X_2 + \cdots + X_n| > a_n) > 0 \tag{8.1}$$

then also

$$\limsup_{n\to\infty} P(|Y_1 + Y_2 + \cdots + Y_n| > a_n) > 0\,. \tag{8.2}$$

When the marginal tails of a stationary process are heavy, extreme values are, often, important. The partial maxima of the process are a natural functional to use in this case in order to draw the boundary between short and long memory. In the case of stationary SαS processes such boundary was found in [124].

A stationary SαS process has an integral representation

$$X_n = \int_E f_n(x)\, M(dx)\,, \quad n = 1, 2, \ldots\,, \tag{8.3}$$

where M is an SαS random measure on a standard Borel space $(E, \mathcal{E})$ with a σ-finite control measure m. The functions f_n, $n = 1, 2, \ldots$ can be chosen to be of the form:

$$f_n(x) = a_n(x) \left(\frac{dm \circ \phi^{n-1}}{dm}(x)\right)^{1/\alpha} f \circ \phi^{n-1}(x), \quad x \in E\,, \tag{8.4}$$

for $n = 1, 2, \ldots$, where $\phi : E \to E$ is a measurable non-singular map (i.e., a one-to-one map with both ϕ and ϕ^{-1} measurable, mapping the control measure m into an equivalent measure),

$$a_n(x) = \prod_{j=0}^{n-1} u \circ \phi^j(x), \quad x \in E\,,$$

for $n = 0, 1, 2, \ldots$, with $u : E \to \{-1, 1\}$ a measurable function and $f \in L^\alpha(m)$. See [128] and [120] for the details.

A basic fact from ergodic theory is the existence of the Hopf decomposition of the set E with respect to the *flow* $(\phi^n, n = 0, 1, 2, \ldots)$: a decomposition of E into a disjoint (modulo a null set with respect to m) union $E = C \cup D$, such that C and D are measurable ϕ-invariant

sets, and *the flow is conservative on C and dissipative on D*; we refer the reader to [78] for the details. This allows us to write

$$X_n = \int_C f_n(x)\, M(dx) + \int_D f_n(x)\, M(dx) := X_n^C + X_n^D\,, \qquad (8.5)$$

$n = 1, 2, \ldots$, a unique in law decomposition of a stationary symmetric α-stable process into a sum of two independent such processes, one of which is *generated by a conservative flow*, and the other is *generated by a dissipative flow*. The i.i.d. SαS sequence is generated by a dissipative flow (i.e., the component $\mathbf{X}^C$ in the decomposition (8.5) vanishes). See [120].

Let

$$M_n = \max_{j=1,2,\ldots,n} |X_j|, \quad n = 1, 2, \ldots, \qquad (8.6)$$

be the sequence of the partial maxima. The following result was proved in [124]: if the component $\mathbf{X}^D$ generated by a dissipative flow in the decomposition (8.5) does not vanish, then

$$n^{-1/\alpha} M_n \Rightarrow C\, Z_\alpha\,, \qquad (8.7)$$

where C is a finite positive constant, and Z_α is the standard Frechét extreme value random variable with the distribution

$$P\big(Z_\alpha \le z\big) = e^{-z^{-\alpha}}, \quad z > 0\,. \qquad (8.8)$$

If, on the other hand, the component $\mathbf{X}^D$ generated by a dissipative flow vanishes, then

$$n^{-1/\alpha} M_n \to 0 \qquad (8.9)$$

in probability as $n \to \infty$.

This is a phase transition that qualifies as a change from short memory to long memory. Let us call stationary SαS processes with only one nondegenerate component in the decomposition (8.5) *single component processes*. We parametrize the family of laws of single component stationary SαS processes (with the scale fixed to, say, 1) by a space $(E, \mathcal{E}, m)$, a *flow* $(\phi^n, n = 0, 1, 2, \ldots)$, a function $f \in L^\alpha(m)$, and

a *cocycle* $(a_n, n = 0, 1, 2, \ldots)$. The collection of these parameters forms the parameter space Θ. Then the set Θ_1 of parameters for which the flow is dissipative corresponds to short memory processes, while the set Θ_2 of parameters for which the flow is conservative corresponds to short memory processes, and the boundary between the two is the phase transition boundary.

It is interesting that *the partial maxima grow the fastest in the short memory case* (including the i.i.d. case). In particular, if a stationary SαS process has both a nondegenerate dissipative component and a nondegenerate conservative component in (8.5), then the long range dependent conservative component will be hidden by the faster growing maximum of the short memory dissipative component. Therefore, if we use the same parameters as above to parametrize the family of laws of *all* stationary SαS processes with the same scale, then the phase transition becomes less interesting, because in this case the short memory part Θ_1 of the parameter space becomes the set of the parameters in which the flow has a non-vanishing dissipative component. This will allow for short memory processes with a nondegenerate long memory component generated by a conservative flow. This is an indication that for certain functionals of a stationary process it is important to choose the parametrization carefully.

In the light of the present discussion let us revisit the question of short or long memory in the increments of SSSI SαS processes considered in Section 7. If $\mathbf{Y}$ is the Linear Fractional Stable Motion defined in (7.18) and (7.19), then its increment process is a single component process generated by a dissipative flow (see [120]), hence a short memory process, regardless of the value of index of stability α and Hurst exponent H. Similarly, if $\mathbf{Y}$ is the Harmonizable Fractional Stable Motion defined in (7.23), then its increment process is a single component process generated by a conservative flow (see once again [120]), hence a long memory process, once again regardless of the values of α and H. In fact, since the increment process of the Harmonizable Fractional Stable Motion is not ergodic, we can view this process as having infinite memory. On the other hand, Cohen and Samorodnitsky [32] constructed a family of SSSI SαS process for which the increment process is also a single component process generated by a conservative flow, but this time

the process is ergodic (even mixing). Thus we can view it as having finite but long memory.

It is not difficult to see that this classification of the memory of the increments of SSSI SαS processes is more informative than making a distinction based on a single critical value of the Hurst exponent or on decay rates of covariance substitutes such as covariation and codifference.

It is interesting that for stationary *Gaussian* processes the change in the rate of increase of the partial maxima occurs "much later," from the point of view of the rate of decay of correlations, than what is needed to change the rate of increase of the partial sums as discussed above. For example, if the condition

$$R_n \log n \to 0 \quad \text{as } n \to \infty$$

is satisfied by the covariance function of a stationary Gaussian process (i.e., if the the correlations decay faster than *logarithmically* fast), then the partial maxima of the process increase at the same rate as in the i.i.d. case (and even the constants are preserved!); see [16].

More examples of a phase transition of the type discussed above are given by infinite moving average models. These are models of the form:

$$X_n = \sum_{j=-\infty}^{\infty} \varphi_{n-j}\, \varepsilon_j\,, \quad n = 1, 2, \ldots, \tag{8.10}$$

where $(\varepsilon_n, n = \ldots, -1, 0, 1, 2, \ldots)$ are i.i.d. *noise variables*, and (φ_n) are deterministic coefficients; the latter, clearly, have to satisfy certain conditions for the series to converge and the process to be well defined. We will assume that the noise variables have a finite mean, in which case the absolute summability of the coefficients

$$\sum_{j=-\infty}^{\infty} |\varphi_j| < \infty \tag{8.11}$$

guarantees that the series in (8.10) converges (absolutely) a.s., but weaker assumptions (depending on the noise distribution) will suffice for a.s. convergence as well. Obviously, if the moving average process (8.10) is well defined, it is automatically a stationary process. It is

also known as *a linear process* (as is the ARMA model discussed in Section 3, which is a special case of the infinite moving average model).

Fixing the distribution of the noise variables, the law of the stationary moving average process is determined by its sequence of coefficients, so in this case the parameter space Θ is the collection of sequences (φ_n) for which the series (8.10) converges. Unless the noise variables are Gaussian or α-stable, different choices of the parameter will affect the one-dimensional marginal distributions of the moving average process by more than a scale factor. Still, it makes sense to restrict the parameter space in an appropriate way to keep the marginal distributions from varying too much.

The part of the parameter space Θ defined by (8.11) is sometimes viewed as leading to short memory models, often under the additional assumption

$$\sum_{j=-\infty}^{\infty} \varphi_j \neq 0\,, \tag{8.12}$$

see e.g., Section 13.2 of [26]. (Intuitively, the case of the zero sum in (8.12) corresponds to negative memory, not dissimilar with the Fractional Gaussian Noise with $0 < H < 1/2$.) Obviously, the laws of i.i.d. sequences belong to this part of the parameter space. Is there a phase transition that occurs when (8.11) breaks down?

Assume first that the noise variables have a finite variance $\sigma^2 > 0$. Then the necessary and sufficient condition for convergence of the series in (8.10) is

$$S_\varphi := \sum_{j=-\infty}^{\infty} \varphi_j^2 < \infty\,. \tag{8.13}$$

For the purpose of normalization of the marginal distributions one can define the parameter space Θ to consist of the sequences of the coefficients satisfying $S_\varphi = 1$.

As before, a reasonable partition of Θ into two parts, corresponding to short and long memory processes, is obtained depending on whether or not the partial sums of the moving average process satisfy the invariance principle with convergence to the Brownian motion.

Moving average processes with absolutely summable coefficient have this property (see [60]). However, also every parameter point satisfying, for example, condition (5.37) in [59] guarantees such an invariance principle, hence a short memory process, and this condition may be satisfied even when (8.11) fails. From this point of view, the summability of the coefficients (8.11) is a sufficient condition for a short memory process, but its failure does not automatically imply a long memory moving average.

Still concentrating on the partial sums of a stationary process, important aspects of their behavior are related to large deviations. Suppose that the noise random variables satisfy

$$Ee^{\lambda\varepsilon_0} < \infty \tag{8.14}$$

for $\lambda \in (-\epsilon, \epsilon)$, some $\epsilon > 0$, i.e., have exponentially fast decaying tails.

Let $\mathbf{X} = (X_1, X_2, \ldots)$ be a stationary process. We say that the large deviation principle holds for the sample averages of the process for some *speed sequence* $b_n \uparrow \infty$ and upper and lower rate function $I_u(\cdot)$ and $I_l(\cdot)$, respectively,

$$\begin{aligned} -\inf_{x\in A^\circ} I_l(x) &\le \liminf_{n\to\infty} \frac{1}{b_n} \log P\left(\frac{X_1 + \cdots + X_n}{n} \in A\right) \\ &\le \limsup_{n\to\infty} \frac{1}{b_n} \log P\left(\frac{X_1 + \cdots + X_n}{n} \in A\right) \\ &\le -\inf_{x\in \bar{A}} I_u(x) \end{aligned} \tag{8.15}$$

for every Borel set A, where A° and $\bar{A}$ denote the interior and closure of a set A, correspondingly. Detailed accounts of large deviations are in [40] and [41]. The speed sequence has the single most important role in the large deviation principle (8.15). For i.i.d. sequences satisfying (8.14), the classical Cramer theorem says that the large deviation principle holds with the speed $b_n = n$, $n = 1, 2, \ldots$.

Returning to infinite moving averages with the noise variables satisfying the exponential tail condition (8.14), define a partition of the parameter space Θ (consisting of the sequences of the coefficients satisfying $S_\varphi = 1$ in (8.13)) into parts Θ_1 and Θ_2 by declaring Θ_1 to be that set of parameters θ for which the large deviation principle holds under

P_θ with the speed $b_n \equiv n$, and Θ_2, corresponding to the long memory moving averages, to be its complement. Similar partitions can be created based on the functional version of the large deviation principle, known to hold in the i.i.d. case with the linear speed by the Mogulskii theorem ([102] or Theorem 5.1.2 in [40], assuming that (8.14) holds for all $\lambda \in \mathbb{R}$), or on other versions of the large deviation principle.

The fact that, under the assumptions (8.11) and (8.12), the moving average process satisfies the large deviation principle with the speed $b_n \equiv n$ is well established, albeit under a variety of tail assumptions on the noise variables: see [27] and [69]. However, such large deviation principle also holds under the weaker assumption (5.14); see [43] or [147]. (All these authors establish their large deviation principles at different levels of generality, but always covering the simplest one-dimensional version formulated in (8.15)). From this point of view, the assumptions (8.11) and (8.12) are sufficient for a short memory linear process, but not necessary.

Certain situations where the large deviation principle with a linear rate no longer holds were presented in [54]. Specifically, they assumed that the coefficients (φ_n) are *balanced regularly varying*: there is a regularly varying at infinity with exponent $-\beta, 1/2 < \beta \leq 1$, function $\psi : [0,\infty) \to [0,\infty)$ and $0 \leq p \leq 1$, such that

$$\lim_{n\to\infty} \frac{\phi_n}{\psi(n)} = p \quad \text{and} \quad \lim_{n\to\infty} \frac{\phi_{-n}}{\psi(n)} = 1 - p. \tag{8.16}$$

If $\beta = 1$, assume further that (8.11) fails.

Denote $\Psi_n = \sum_{1\leq i\leq n} \psi(i)$, $n = 1, 2, \ldots$. It turns that, under the balanced regular variation assumption (8.16), the moving average process satisfies the large deviation principle with the speed $b_n = n/\Psi_n^2$, $n = 1, 2, \ldots$. Observe that (by Karamata's theorem) this speed sequence is regular varying with exponent $2\beta - 1$. Even in the case $\beta = 1$ the speed sequence has the form $b_n = nL_n$, with a slowly varying function L converging to zero, and so it grows strictly slower that linearly fast. More general versions of the large deviation principle also exhibit similar behavior; see [54].

From this point of view, moving average processes with coefficients satisfying (8.16) are long range dependent.

Several other large deviation-related ways to look at short and long memory exist, the better known ones of which are related to *ruin probabilities* and *long strange segments*; a recent extensive account of the former is in [5], the latter have been analyzed since [52]. We will look at the behavior of the long strange segments, defined as follows. For a Borel set A and $n = 1, 2, \ldots$ let

$$R_n(A) = \sup\left\{ j - i : 0 \le i < j \le n, \frac{X_{i+1} + \cdots + X_j}{j - i} \in A \right\}, \quad (8.17)$$

(defined to be equal to zero if the supremum is taken over the empty set). If the closure of A does not contain the mean of the stationary process (which we have assumed to be equal to zero), then the long segments over which the sample mean belongs to A are "strange" because the law of large numbers seems to break down there. For the i.i.d. sequence under the finite exponential moment assumption (8.14),

$$\frac{1}{\inf_{x \in A^\circ} I_l(x)} \le \liminf_{n\to\infty} \frac{R_n(A)}{\log n} \le \limsup_{n\to\infty} \frac{R_n(A)}{\log n} \le \frac{1}{\inf_{x \in \bar{A}} I_u(x)} \quad (8.18)$$

with probability 1, where I_l and I_u are the rate functions in the large deviation principle (8.15); see Theorem 3.2.1 in [40].

Let Θ_1 be that part of the parameter space Θ (still consisting of the sequences of the coefficients satisfying $S_\varphi = 1$ in (8.13)) where the length of the long strange segments grows at the logarithmic rate, as in (8.18), and Θ_2 to be its complement. Since the statement (8.18) is set-dependent, one can restrict the test set A to the form $A = \{x : |x - EX| > \theta\}$ for small $\theta > 0$. It was shown in [54] that, under the assumptions (8.11) and (8.12), the logarithmic rate of increase still holds for the long strange segments (but with a generally different rate functions in (8.18)), whereas under the assumption (8.16) of the balanced regular variation assumption, the long strange segments grow at a strictly faster rate: the rate is now $h(\log n)$, where the function $h : (0, \infty) \to (0, \infty)$ satisfies

$$\frac{h(s)}{\left[\Psi\left(\lfloor h(s) \rfloor\right)\right]^2} \to 1 \quad \text{as } s \to \infty .$$

Note that the function h is regularly varying with exponent $(2\beta - 1)^{-1}$. Therefore, also from the point of view of the long strange segments, the assumptions (8.11) and (8.12) are sufficient for a short memory of the moving average process, while the balanced regular variation assumption (8.16) is sufficient for long memory.

In summary, the change from short to long memory in infinite moving average processes with finite exponential moments as in (8.14) is of a phase transition nature. Only sufficient conditions for being on either side of the boundary are known at the moment; future research will undoubtedly tell us more about the description of the boundary in terms of the coefficients in the model.

We conclude by briefly looking at related phase transitions for infinite moving averages, where the noise variables do not have finite exponential moments. Suppose that the noise variables have, in fact, balanced regularly varying tails; this is a notion slightly more general than the balanced power tails in (2.5). Specifically, assume that

$$\begin{cases} P(|\varepsilon_0| > \lambda) = L(\lambda)\, \lambda^{-\alpha}\,, \\ \lim_{\lambda\to\infty} \dfrac{P(\varepsilon_0 > \lambda)}{P(|\varepsilon_0| > \lambda)} = p_\varepsilon, \quad \lim_{\lambda\to\infty} \dfrac{P(\varepsilon_0 < -\lambda)}{P(|\varepsilon_0| > \lambda)} = q_\varepsilon\,, \end{cases} \tag{8.19}$$

as $\lambda \to \infty$, for some $\alpha > 1$ and $0 < p_\varepsilon = 1 - q_\varepsilon \leq 1$. Here L is a slowly varying function at infinity. Note that, if $\alpha \leq 2$, then the assumption of the finite variance of the noise variables may fail, and it certainly does fail if $\alpha < 2$. If $\alpha > 2$, then the variance of the noise variables is still finite, and the square summability condition (8.13) is still the necessary and sufficient condition for the linear process (8.10) to be well defined; in the case $1 < \alpha \leq 2$ a sufficient condition is

$$\sum_{j=-\infty}^{\infty} |\varphi_j|^{\alpha-\epsilon} < \infty \quad \text{for some } \epsilon > 0. \tag{8.20}$$

In both cases the resulting moving average process is a stationary process whose one-dimensional marginal tails are proportional to the tails of the noise variables; see [98].

Once again, let Θ be the parameter space appropriate to the situation at hand: this will be the space of the coefficients satisfying (8.13)

is $\alpha > 2$, or the space of the coefficients satisfying (8.20) if $1 < \alpha \le 2$. Keeping the marginal distributions of the process from varying too much as the parameter changes is desirable; with "power-like" tails it is often a good idea to control the tails: the normalization $\sum_j |\varphi_j|^\alpha = 1$ achieves that (see [98]), and it can be taken as a part of the description of the parameter space.

Full large deviation principles have not been extended for stochastic processes with "power-like" tails far beyond the i.i.d. case (see [66] and [105] for what happens in that case), so we will only consider the long strange segments. It turns out that the absolutely summable coefficients satisfying (8.11) still belong to that part Θ_1 of the parameter space such that, under the law P_θ with $\theta \in \Theta_1$ the long strange segments of the moving average process behave as in the i.i.d. case, while certain balanced regularly varying coefficients belong to that part Θ_2 of the parameter space such that, under the law P_θ with $\theta \in \Theta_2$ the long strange segments of the moving average process behave in a drastically different way. We need to modify the assumption of the balanced regularly varying coefficients in (8.16) as follows: assume that (8.16) is satisfied with $1/2 < \beta < 1$ if $\alpha > 2$, and with $1/\alpha < \beta < 1$ if $1 < \alpha \le 2$.

We consider long strange intervals defined in (8.17), with test sets of the form $A = (\theta, \infty)$ for $\theta > 0$ (these are "strange" because of the assumption of zero mean). Let F denote the distribution function of the noise random variable $|\varepsilon_0|$, and for $n \ge 1$ define

$$a_n = \left(\frac{1}{1 - F} \right)^{\leftarrow} (n) , \tag{8.21}$$

where for a nondecreasing function U, $U^{\leftarrow}(y) = \inf\{s : U(s) \ge y\}$, $y > 0$, is the left continuous inverse of U. Clearly (a_n) is regularly varying at infinity with exponent $1/\alpha$. It was proved by Mansfield et al. [94] that under the absolutely summability assumption (8.11),

$$a_n^{-1} R_n\big((\theta, \infty)\big) \Rightarrow \frac{M(\varphi)}{\theta} Z_\alpha ,$$

where Z_α is the standard Frechét random variable defined in (8.8), and

$$M(\varphi) = \left(p_\varepsilon M_+(\varphi)^\alpha + q_\varepsilon M_-(\varphi)^\alpha \right)^{1/\alpha} ,$$

with

$$M_+(\varphi) = \max\left\{ \sup_{-\infty<k<\infty} \left(\sum_{j=-\infty}^{k} \varphi_j \right)_+ , \sup_{-\infty<k<\infty} \left(\sum_{j=k}^{\infty} \varphi_j \right)_+ \right\}$$

and

$$M_-(\varphi) = \max\left\{ \sup_{-\infty<k<\infty} \left(\sum_{j=-\infty}^{k} \varphi_j \right)_- , \sup_{-\infty<k<\infty} \left(\sum_{j=k}^{\infty} \varphi_j \right)_- \right\}.$$

Therefore, under the assumption (8.11), the length of the long strange segments grows at the rate a_n (which is regularly varying at infinity with exponent $1/\alpha$); this rate is the same as the rate of growth of the long strange segments for i.i.d. sequences with the same marginal tails. Note that the assumption (8.12) is not needed here.

On the other hand, under the assumption (8.16) of the balanced regularly variation of the coefficients (modified to the present case of the "power-like" tails), it was shown in [112] that

$$\tilde{a}_n^{-1} R_n\big((\theta,\infty)\big) \Rightarrow \frac{p_\varepsilon^{1/(\alpha\beta)}\big(p^{1/\beta}+q^{1/\beta}\big)}{(1-\beta)^{1/\beta}\theta^{1/\beta}} Z_{\alpha\beta}\,,$$

where $Z_{\alpha\beta}$ is, once again, the standard Frechét random variable, but this time with the exponent equal to $\alpha\beta$, and the sequence $(\tilde{a}_n)$ satisfies

$$\frac{\psi(\tilde{a}_n)}{a_n} \to 1 \quad \text{as } n \to \infty.$$

Therefore, under the (modified) assumption of the balanced regularly variation of the coefficients, the length of the long strange segments grows at the rate $\tilde{a}_n$, which is regularly varying at infinity with exponent $1/(\alpha\beta)$ and, hence, faster than in the case of absolutely summable coefficients.

We conclude that the behavior of long strange segments in the infinite moving average processes with noise variables with balanced regularly varying tails exhibits a phase transition (similar to that in the case of linear processes with the "light-tailed" noise variables) that may qualify as a phase transition that separates between short memory and long memory processes. The assumption of absolute summability of the

coefficients is sufficient for a short memory process, while the assumption of a certain regular variation of the coefficients is sufficient for a long memory process.

In summary, it appears that connecting the notion of long range dependence to certain types of phase transitions is promising. It fits well with our intuition of the term "long memory" describing a model that is out of the ordinary. Furthermore, it allows us to concentrate on the behavior of really important functionals. Much remains to be done to clarify both possible types of such phase transitions and the relevant boundaries for concrete families of stochastic processes.

Acknowledgments

Samorodnitsky's research was partially supported by an NSA grant MSPF-05G-049 and an ARO grant W911NF-07-1-0078 at Cornell University.

References

[1] J. Aaronson, "An introduction to infinite ergodic theory," in volume 50 *of Mathematical Surveys and Monographs*, Providence: American Mathematical Society, 1997.

[2] P. Abry, L. Delbeke, and P. Flandrin, "Wavelet based estimator for the self-similarity parameter of α-stable processes," in *International Conference on Signal Processing, 16th World Computer Congress*, 2000.

[3] E. Alvarez-Lacalle, B. Dorow, J.-P. Eckmann, and E. Moses, "Hierarchical structures induce long-range dynamical correlations in written texts," *PNAS*, vol. 103, pp. 7956–7961, 2006.

[4] V. Anh, J. Angulo, and M. Ruiz-Medina, "Possible long-range dependence in fractional random fields," *Journal of Statistical Planning and Inference*, vol. 80, pp. 95–110, 1999.

[5] S. Asmussen, *Ruin Probabilities*. World Scientific Publishing Company, 2000.

[6] A. Astrauskas, "Limit theorems for quadratic forms of linear processes," *Lithuanian Mathematical Journal*, vol. 23, pp. 355–361, 1983.

[7] A. Astrauskas, "Limit theorems for sums of linearly generated random variables," *Lithuanian Mathematical Journal*, vol. 23, pp. 127–134, 1983.

[8] A. Astrauskas, J. Levy, and M. S. Taqqu, "The asymptotic dependence structure of the linear fractional Lévy motion," *Lietuvos Matematikos Rinkinys (Lithuanian Mathematical Journal)*, vol. 31, pp. 1–28, 1991.

[9] F. Avram and M. S. Taqqu, "Noncentral limit theorems and Appell polynomials," *The Annals of Probability*, vol. 15, pp. 767–775, 1987.

[10] F. Avram and M. S. Taqqu, "Weak convergence of sums of moving averages in the α-stable domain of attraction," *The Annals of Probability*, vol. 20, pp. 483–503, 1992.

[11] F. Avram and M. S. Taqqu, "Robustness of the R/S statistic for fractional stable noises," *Statistical Inference for Stochastic Processes*, vol. 3, pp. 69–83, 2000.

[12] R. Baillie, T. Bollerslev, and H. Mikkelsen, "Fractionally integrated generalized autoregressive conditional heteroscedasticity," *Journal of Econometrics*, vol. 74, pp. 3–30, 1996.

[13] J. Barkoulas, W. Labys, and J. Onochie, "Long memory in future prices," *The Financial Review*, vol. 34, pp. 91–100, 1999.

[14] J. Beran, *Statistics for Long–Memory Processes*. New York: Chapman and Hall, 1994.

[15] K. Berg-Sørensen and H. Flyvberg, "The colour of thermal noise in classical Brownian motion: A feasibility study of direct experimental observation," *New Journal of Physics*, vol. 7, pp. 38–47, 2005.

[16] S. Berman, "Limit theorems for the maximum term in stationary sequencies," *Annals of Mathematical Statistics*, vol. 35, pp. 502–516, 1964.

[17] R. Bhattacharya, V. Gupta, and E. Waymire, "The Hurst effect under trends," *Journal of Applied Probability*, vol. 20, pp. 649–662, 1983.

[18] P. Billingsley, *Convergence of Probability Measures*. New York: Wiley, Second ed., 1999.

[19] N. Bingham, C. Goldie, and J. Teugels, *Regular Variation*. Cambridge University Press, 1987.

[20] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, pp. 307–327, 1986.

[21] A. Böttcher and J. Virtanen, "Norms of Toeplitz matrices with Fisher-Hartwig symbols," *SIAM Journal of Matrix Analysis and Applications*, vol. 29, pp. 660–671, 2007.

[22] R. Bradley, "Equivalent mixing conditions for random fields," *Annals of Probability*, vol. 21, pp. 1921–1926, 1993.

[23] R. Bradley, "On the growth of variances in a central limit theorem for strongly mixing sequences," *Bernoulli*, vol. 5, pp. 67–80, 1999.

[24] R. Bradley, "Basic properties of strong mixing conditions. A survey and some open questions," *Probability Surveys*, vol. 2, pp. 107–144, 2005.

[25] P. Breuer and P. Major, "Central limit theorems for non-linear functionals of Gaussian fields," *Journal of Multivariate Analysis*, vol. 13, pp. 425–441, 1983.

[26] P. Brockwell and R. Davis, *Time Series: Theory and Methods*. New York: Springer-Verlag, 1987.

[27] R. Burton and H. Dehling, "Large deviations for some weakly dependent random processes," *Statistics and Probability Letters*, vol. 9, pp. 397–401, 1990.

[28] S. Cambanis, C. Hardin, Jr., and A. Weron, "Ergodic properties of stationary stable processes," *Stochastic Processes and Their Applications*, vol. 24, pp. 1–18, 1987.

[29] S. Cambanis, C. Hardin, Jr., and A. Weron, "Innovations and Wold decompositions of stable sequences," *Probabilistic Theory of Relative Fields*, vol. 79, pp. 1–27, 1988.

[30] S. Cambanis and M. Maejima, "Two classes of self-similar stable processes with stationary increments," *Stochastic Processes and Their Applications*, vol. 32, pp. 305–329, 1989.
[31] S. Cambanis and A. Soltani, "Prediction of stable processes: Spectral and moving average representation," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 66, pp. 593–612, 1984.
[32] S. Cohen and G. Samorodnitsky, "Random rewards, fractional Brownian local times and stable self-similar processes," *Annals of Applied Probability*, vol. 16, pp. 1432–1461, 2006.
[33] R. Cont, "Long range dependence in financial markets," in *Fractals in Engineering*, (E. Lutton and J. Vehel, eds.), pp. 159–180, Springer, 2005.
[34] I. Cornfeld, S. Fomin, and Y. G. Sinai, *Ergodic Theory*. Springer-Verlag, 1982.
[35] D. Cox, "Long-range dependence: A review," in *Statistics: An Appraisal*, (H. David and H. David, eds.), pp. 55–74, Iowa State University Press, 1984.
[36] N. Crato and P. Rothman, "Fractional integration analysis of long-run behavior for US macroeconomic time series," *Economics Letters*, vol. 45, pp. 287–291, 1994.
[37] D. Daley and R. Vesilo, "Long range dependence of point processes, with queueing examples," *Stochastic Processes and Their Applications*, vol. 70, pp. 265–282, 1997.
[38] Y. Davydov and V. Egorov, "On convergence of empirical processes," Technical report, IRMA, Lille, vol. 64, no. V, 2005.
[39] L. Decreusefond and A. Üstünel, "Stochastic analysis of the Fractional Brownian motion," *Potential Analysis*, vol. 10, pp. 177–214, 1999.
[40] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Boston: Jones and Bartlett Publishers, 1993.
[41] J.-D. Deuschel and D. Stroock, *Large Deviations*. Boston: Academic Press, 1989.
[42] F. Diebold and A. Inoue, "Long memory and regime switching," *Journal of Econometrics*, vol. 105, pp. 131–159, 2001.
[43] H. Djellout and A. Guillin, "Large and moderate deviations for moving average processes," *Annales de la Faculté des Sciences de Toulouse*, vol. X, pp. 23–31, 2001.
[44] R. Dobrushin, "Gaussian and their subordinated self-similar random generalized fields," *Annals of Probability*, vol. 7, pp. 1–28, 1979.
[45] R. Dobrushin and P. Major, "Non-central limit theorems for non-linear functions of Gaussian fields," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 50, pp. 27–52, 1979.
[46] R. Douc, F. Roueff, and P. Soullier, "On the existence of some ARCH(∞) processes," *Stochastic Processes and Their Applications*, To appear, 2006.
[47] R. Dudley, "The sizes of compact subsets of Hilbert space and continuity of Gaussian processes," *Journal of Functional Analysis*, vol. 1, pp. 290–330, 1967.
[48] R. Durrett, *Probability: Theory and Examples*. Belmont, California: Duxbury Press, Second ed., 1996.
[49] R. Durrett and S. Resnick, "Functional limit theorems for dependent variables," *Annals of Probability*, vol. 6, pp. 829–846, 1978.

[50] P. Embrechts and M. Maejima, *Selfsimilar Processes*. Princeton and Oxford: Princeton University Press, 2002.
[51] R. Engle, "Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation," *Econometrica*, vol. 50, pp. 987–1008, 1982.
[52] P. Erdós and A. Rényi, "On a new law of large numbers," *Journal d'Analyse Mathématique*, vol. 22, pp. 103–111, 1970.
[53] W. Feller, *An Introduction to Probability Theory and its Applications*. Vol. 2, New York: Wiley, Second ed., 1971.
[54] S. Ghosh and G. Samorodnitsky, "Large deviations and memory in moving average processes," Technical report, 2007.
[55] L. Gil-Alana, "A fractionally integrated model for the Spanish real GDP," *Economics Bulletin*, vol. 3, pp. 1–6, 2004.
[56] L. Giraitis and D. Surgailis, "On shot noise processes attracted to fractional Lévy motion," in *Stable Processes and Related Topics*, (S. Cambanis, G. Samorodnitsky, and M. Taqqu, eds.), pp. 261–273, Boston: Birkhäuser, 1991. (Volume 25 of *Progress in Probability*).
[57] C. Granger and R. Joyeux, "An introduction to long-memory time series and fractional differencing," *Journal of Time Series Analysis*, vol. 1, pp. 15–30, 1980.
[58] D. Guégan, "How can we define the concept of long memory? An econometric survey," *Econometric reviews*, vol. 24, pp. 113–149, 2005.
[59] P. Hall and C. Heyde, *Martingale Limit Theory and Its Applications*. New York: Academic Press, 1980.
[60] E. J. Hannah, "The central limit theorem for time series regression," *Stochastic Processes and Their Applications*, vol. 9, pp. 281–289, 1979.
[61] M. Hauser and R. Kunst, "Fractionally integrated models with ARCH errors: With an application to the Swiss 1-month Euromarket interest rate," *Review of Quantitative Finance and Accounting*, vol. 10, pp. 95–113, 1998.
[62] H. Helson and D. Sarason, "Past and future," *Mathematica Scandinavica*, vol. 21, pp. 5–16, 1967.
[63] C. Heyde, "On the modes of long-range dependence," *Journal of Applied Probability*, vol. 39, pp. 882–888, 2002.
[64] C. Heyde and Y. Yang, "On defining long range dependence," *Journal of Applied Probability*, vol. 34, pp. 939–944, 1997.
[65] J. Hosking, "Fractional differencing," *Biometrika*, vol. 68, pp. 165–176, 1981.
[66] H. Hult, F. Lindskog, T. Mikosch, and G. Samorodnitsky, "Functional large deviations for multivariate regularly varying random walks," *Annals of Applied Probability*, vol. 15, pp. 2651–2680, 2005.
[67] H. Hurst, "Long-term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers*, vol. 116, pp. 770–808, 1951.
[68] H. Hurst, "Methods of using long-term storage in reservoirs," *Proceedings of the Institution of Civil Engineers, Part I*, pp. 519–577, 1955.
[69] T. Jiang, M. Rao, and X. Wang, "Large deviations for moving average processes," *Stochastic Processes and Their Applications*, vol. 59, pp. 309–320, 1995.

[70] T. Karagiannis, M. Molle, and M. Faloutsos, "Long-range dependence — Ten years of Internet traffic modeling," *IEEE Internet Computing*, vol. 8, pp. 57–64, 2004.

[71] D. Karmeshu and A. Krishnamachari, "Sequence variability and long-range dependence in DNA: An information theoretic perspective," in *Neural Information Processing*, pp. 1354–1361, Berlin: Springer, 2004. 3316 of Lecture Notes in Computer Science.

[72] Y. Kasahara and M. Maejima, "Weighted sums of i.i.d. random variables attracted to integrals of stable processes," *Probability Theory Related Fields*, vol. 78, pp. 75–96, 1988.

[73] Y. Kasahara, M. Maejima, and W. Vervaat, "Log-fractional stable processes," *Stochastic Processes and their Applications*, vol. 30, pp. 329–339, 1988.

[74] P. Kokoszka and M. Taqqu, "Infinite variance stable ARMA processes," *Journal of Time Series Analysis*, vol. 15, pp. 203–220, 1994.

[75] A. Kolmogorov, "Wienersche Spiralen und einige andere interessante kurven in Hilbertschen raum," *Computes Rendus (Doklady) Academic Sciences USSR (N.S.)*, vol. 26, pp. 115–118, 1940.

[76] A. Kolmogorov and Y. Rozanov, "On a strong mixing condition for a stationary Gaussian process," *Teoria Veroyatnostei i Primeneniya*, vol. 5, pp. 222–227, 1960.

[77] M. Kotulska, "Natural fluctuations of an elctropore show Fractional Lévy Stable motion," *Biophysical Journal*, vol. 92, pp. 2412–2421, 2007.

[78] U. Krengel, *Ergodic Theorems*. Berlin, New York: De Gruyter, 1985.

[79] H. Künsch, "Discrimination between monotonic trends and long-range dependence," *Journal of Applied Probability*, vol. 23, pp. 1025–1030, 1986.

[80] J. Lamperti, "Semi-stable stochastic processes," *Transaction of the American Mathematical Society*, vol. 104, pp. 62–78, 1962.

[81] M. Leadbetter, G. Lindgren, and H. Rootzén, *Extremes and Related Properties of Random Sequences and Processes*. New York: Springer Verlag, 1983.

[82] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic," *ACM/SIGCOMM Computer Communications Review*, pp. 183–193, 1993.

[83] J. Levy and M. Taqqu, "Renewal reward processes with heavy-tailed inter-renewal inter-renewal times and heavy-tailed rewards," *Bernoulli*, vol. 6, pp. 23–44, 2000.

[84] A. Lo, "Fat tails, long memory, and the stock market since the 1960s," *Economic Notes,* Banca Monte dei Paschi di Siena, 2001.

[85] J. Lopez-Ardao, C. Lopez-Garcia, A. Suarez-Gonzales, M. Fernandes-Vega, and R. Rodriguez-Rubio, "On the use of self-similar processes in network simulation," *ACM Transactions on Modeling and Computer Simulation*, vol. 10, pp. 125–151, 2000.

[86] M. Maejima, "On a class of self-similar processes," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 62, pp. 235–245, 1983.

[87] M. Maejima, "A self-similar process with nowhere bounded sample paths," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 65, pp. 115–119, 1983.

[88] P. Major, *Multiple Wiener-Itô Integrals*. New York: Springer Verlag, 1981. (Volume 849 of *Springer Lecture Notes in Mathematics*).

[89] B. Mandelbrot, "Une classe de processus stochastiques homothetiques a soi; application a loi climatologique de H. E. Hurst," *Comptes Rendus Academic Sciences Paris*, vol. 240, pp. 3274–3277, 1965.

[90] B. Mandelbrot, *The Fractal Geometry of Nature*. San Francisco: W. H. Freeman and Co., 1983.

[91] B. Mandelbrot and M. Taqqu, "Robust R/S analysis of long-run serial correlation," in *Proceedings of the 42nd Session of the International Statistical Institute*, pp. 69–104, Manila: Bulletin of the I.S.I., 1979.

[92] B. Mandelbrot and J. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, pp. 422–437, 1968.

[93] B. Mandelbrot and J. Wallis, "Noah, Joseph and operational hydrology," *Water Resources Research*, vol. 4, pp. 909–918, 1968.

[94] P. Mansfield, S. Rachev, and G. Samorodnitsky, "Long strange segments of a stochastic process and long range dependence," *Annals of Applied Probability*, vol. 11, pp. 878–921, 2001.

[95] M. Marcus and G. Pisier, "Characterization of almost surely continuous p-stable random Fourier series and strongly stationary processes," *Acta Mathematica*, vol. 152, pp. 245–301, 1984.

[96] F. Merlevéde, M. Peligrad, and S. Utev, "Recent advances in invariance principles for stationary sequences," *Probability Surveys*, vol. 3, pp. 1–36, 2006.

[97] Z. Michna, "Self-similar processes in collective risk theory," *Journal of Applied Mathematics and Stochastic Analysis*, vol. 11, pp. 429–448, 1998.

[98] T. Mikosch and G. Samorodnitsky, "The supremum of a negative drift random walk with dependent heavy–tailed steps," *Annals of Applied Probability*, vol. 10, pp. 1025–1064, 2000.

[99] T. Mikosch and C. Stărică, "Change of structure in financial time series, long range dependence and the GARCH model," IWI–preprint 99-5-06 of the University of Groningen, 1999.

[100] T. Mikosch and C. Stărică, "Is it long memory or just changing variance we see in financial returns?," *Bulletin of EU & US Inflation and Microeconomic Analysis*, vol. 68, pp. 46–68, 2000.

[101] G. Miller, "Properties of certain symmetric stable distributions," *Journal of Multivariate Analysis*, vol. 8, pp. 346–360, 1978.

[102] A. Mogulski, "Large deviations for trajectories of multidimensional random walks," *Theory of Probability and its Applications*, vol. 21, pp. 300–315, 1976.

[103] P. Moran, "On the range of cumulative sums," *Annals of the Institute of Statistical Mathematics*, vol. 16, pp. 109–112, 1964.

[104] T. Mori and H. Oodaira, "The law of the iterated logarithm for self-similar processes represented by multiple Wiener integrals," *Probability Theory Related Fields*, vol. 71, pp. 367–391, 1986.

[105] S. Nagaev, "Large deviations of sums of independent random variables," *Annals of Probability*, vol. 7, pp. 745–789, 1979.

[106] C. Newman and A. Wright, "An invariance principle for certain dependent sequences," *Annals of Probability*, vol. 9, pp. 671–675, 1981.

[107] I. Norros, E. Valkeila, and J. Virtamo, "An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motion," *Bernoulli*, vol. 5, pp. 571–588, 1999.

[108] D. Nualart, *The Mallavin Calculus and Related Topics*. New York: Springer, 1995.

[109] S. Painter, "Long-range dependence in the subsurface: Empirical evidence and simulation methods," Invited paper at the American Geophysical Union 1998 Fall Meeting, 1998.

[110] M. Peligrad, "Maximum of partial sums and an invariance principle for a class of weak dependent random variables," *Proceedings of American Mathematical Society*, vol. 126, pp. 1181–1189, 1998.

[111] P. Phillips, "Time series regression with a unit root," *Econometrica*, vol. 55, pp. 277–301, 1987.

[112] S. Rachev and G. Samorodnitsky, "Long strange segments in a long range dependent moving average," *Stochastic Processes and Their Applications*, vol. 93, pp. 119–148, 2001.

[113] S. Resnick, "Point processes, regular variation and weak convergence," *Advances in Applied Probability*, vol. 18, pp. 66–138, 1986.

[114] S. Resnick, *Extreme Values, Regular Variation and Point Processes*. New York: Springer-Verlag, 1987.

[115] P. Robinson, ed., *Time Series with Long Memory*. Advanced Texts in Econometrics. Oxford University Press, 2003.

[116] M. Rosenblatt, "A central limit theorem and a strong mixing condition," *Proceedings of the National Academy of Sciences*, vol. 42, pp. 43–47, 1956.

[117] M. Rosenblatt, "Some limit theorems for partial sums of quadratic forms in stationary Gaussian variables," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 49, pp. 125–132, 1979.

[118] M. Rosenblatt, "Stochastic processes with short-range and long-range dependence," in *Statistics: An Appraisal*, (H. David and H. David, eds.), pp. 509–520, Iowa State University Press, 1984.

[119] M. Rosenblatt, *Stationary Sequences and Random Fields*. Boston: Birkhauser, 1985.

[120] J. Rosiński, "On the structure of stationary stable processes," *The Annals of Probability*, vol. 23, pp. 1163–1187, 1995.

[121] J. Rosiński, G. Samorodnitsky, and M. Taqqu, "Sample path properties of stochastic processes represented as multiple stable integrals," *Journal of Multivariate Analysis*, vol. 37, pp. 115–134, 1991.

[122] J. Rosiński and T. Żak, "Simple conditions for mixing of infinitely divisible processes," *Stochastic Processes and Their Applications*, vol. 61, pp. 277–288, 1996.

[123] J. Rosiński and T. Żak, "The equivalence of ergodicity and weak mixing for infinitely divisible processes," *Journal of Theoretical Probability*, vol. 10, pp. 73–86, 1997.

[124] G. Samorodnitsky, "Extreme value theory, ergodic theory, and the boundary between short memory and long memory for stationary stable processes," *Annals of Probability*, vol. 32, pp. 1438–1468, 2004.

[125] G. Samorodnitsky, "Long memory and self-similar processes," *Annales de la Faculté des Sciences de Toulouse*, vol. 15, pp. 107–123, 2006.

[126] G. Samorodnitsky and M. Taqqu, "The various linear fractional Lévy motions," in *Probability, Statistics and Mathematics: Papers in Honor of Samuel Karlin*, (K. A. T. W. Anderson and D. Iglehart, eds.), pp. 261–270, Boston: Academic Press, 1989.

[127] G. Samorodnitsky and M. Taqqu, "$(1/\alpha)$-self-similar processes with stationary increments," *Journal of Multivariate Analysis*, vol. 35, pp. 308–313, 1990.

[128] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes*. New York: Chapman and Hall, 1994.

[129] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1999.

[130] J. Smit, "Solution to Problem 130," *Statistica Neerlandica*, vol. 37, p. 87, 1983.

[131] A. Stegeman, "Non-stationarity versus long-range dependence in computer network traffic measurements," Technical report, University of Groningen, 2001.

[132] S. Stoev and M. Taqqu, "Simulation methods for linear fractional stable motions and FARIMA using the Fast Fourier Transform," *Fractals*, vol. 1, pp. 95–121, 2004.

[133] D. Surgailis, "Convergence of sums of nonlinear functions of moving averages to selfsimilar processes," *Soviet Mathematique Doklady*, vol. 23, pp. 247–250, 1981.

[134] D. Surgailis, "On L^2 and non-L^2 multiple stochastic integration," in *Stochastic Differential Systems*, (Lecture Notes in Control and Information Sciences), pp. 212–226, New York: Springer-Verlag, 1981.

[135] D. Surgailis, J. Rosiński, V. Mandrekar, and S. Cambanis, "Stable mixed moving averages," *Probability Theory Related Fields*, vol. 97, pp. 543–558, 1993.

[136] M. Taqqu, "Weak convergence to fractional Brownian motion and to the Rosenblatt process," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 31, pp. 287–302, 1975.

[137] M. Taqqu, "A representation for self-similar processes," *Stochastic Processes and their Applications*, vol. 7, pp. 55–64, 1978.

[138] M. Taqqu, "Convergence of integrated processes of arbitrary Hermite rank," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 50, pp. 53–83, 1979.

[139] M. Taqqu, "A bibliographical guide to self-similar processes and long-range dependence," in *Dependence in Probability and Statistics*, (E. Eberlein and M. Taqqu, eds.), pp. 137–162, Boston: Birkhäuser, 1986.

[140] M. Taqqu, "Random processes with long-range dependence and high variability," *Journal of Geophysical Research*, vol. 92, pp. 9683–9686, 1987.

[141] M. Taqqu and R. Wolpert, "Infinite variance self-similar processes subordinate to a Poisson measure," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 62, pp. 53–72, 1983.

[142] C. Varotsos and D. Kirk-Davidoff, "Long-memory processes in global ozone and temperature variations at the region 60° S–60° N," *Atmospheric Chemistry and Physics*, vol. 6, pp. 4093–4100, 2006.

[143] W. Vervaat, "Sample paths of self-similar processes with stationary increments," *The Annals of Probability*, vol. 13, pp. 1–27, 1985.

[144] N. Wiener, "The homogeneous chaos," *American Journal of Mathematics*, vol. 60, pp. 897–936, 1938.

[145] W. Willinger, V. Paxson, and M. Taqqu, "Self-similarity and heavy tails: Structural modeling of network traffic," in *A Practical Guide to Heavy Tails*, (R. Adler, R. Feldman, and M. Taqqu, eds.), pp. 27–54, Boston: Birkhäuser, 1998.

[146] W. Willinger, M. Taqqu, R. Sherman, and D. Wilson, "Self–similarity through high variability: Statistical analysis of ethernet LAN traffic at the source level (extended version)," *IEEE/ACM Transactions on Networking*, vol. 5, pp. 71–96, 1997.

[147] L. Wu, "On large deviations for moving average processes," in *Probability, Finance and Insurance*, pp. 15–49, River Edge, NJ: World Scientific Publishers, 2004.

[148] A. Yaglom, "Correlation theory of processes with stationary random increments of order n," *Matematica Sbornik*, vol. 37, pp. 141–196, English translation in *American Mathematical Society of Translations Series 2*, vol. 8, pp. 87-141, 1958, 1955.

[149] P. Zaffaroni, "Stationarity and memory of ARCH(∞) models," *Econometric Theory*, vol. 20, pp. 147–160, 2004.

[150] A. Zygmund, *Trigonometric Series*. Vol. I,II, Cambridge: Cambridge University Press, 1968.