LONG RANGE DEPENDENCE IN HEAVY TAILED STOCHASTIC PROCESSES

BORJANA RACHEVA-IOTOVA AND GENNADY SAMORODNITSKY

Abstract. The notion of long range dependence has traditionally been defined through a slow decay of correlations. This approach may be completely inappropriate in the case of a stochastic process with heavy tails. Yet long memory has been reported to be found in various fields where heavy tails are a standard feature of the commonly used stochastic models. Financial and communications networks data are among those often believed to exhibit long memory. We discuss alternative points of view on long range dependence that are applicable in the heavy tailed case. Such alternative approaches may be tailored for a particular applications at hand.

1. Introduction

A glance at the plot on Figure 1.1 describing the annual minima of the water level in the Nile river suggests that the process plotted there has at least 4 distinct time periods when the "level" and "drift" of the process change. This needs not, however, be necessarily taken as

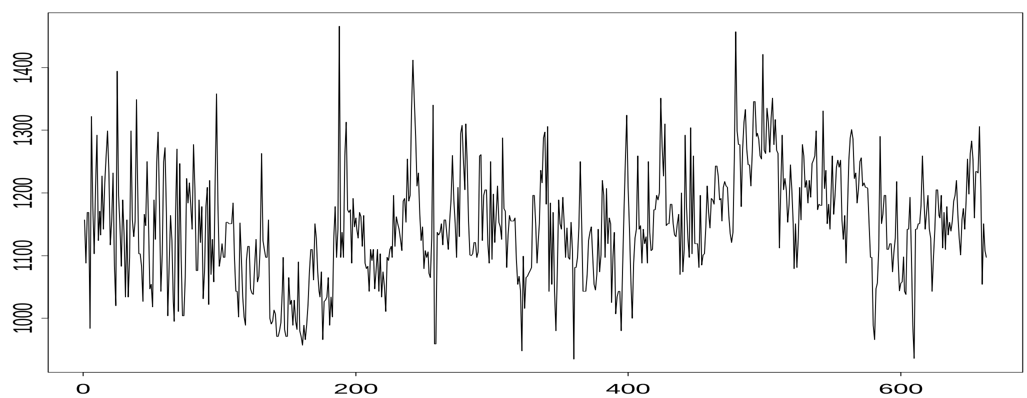

Figure 1.1. Annual minima of the water level in the Nile river for the years 622 to 1281, measured at the Roda gauge near Cairo.

an indication that a nonstationary model should be used for the annual minima of water level

AMS 1991 *subject classifications.* Primary 60G10, 62M10; secondary 60F10.

Key words and phrases. heavy tails, long range dependence, rare events, large deviations .

Samorodnitsky's research was partially supported by NSF grants DMS-0071073 and DMI-9713549 at Cornell University.

process. Even though using a nonstationary model is possible in such situations, sometimes a more *parsimonious* model is a **long memory stationary model** (or a **stationary model with long range dependence**). In fact, commonly used long memory models exhibit what Benoit Mandelbrot termed "persistence", or "Joseph effect" (referring to the long stretches of plenty and famine in Egypt of the Bible). Here is what one sees when looking at the increments of a long memory Fractional Brownian motion (also called Fractional Gaussian noise): "Nearly every sample looks like a "random noise" superimposed upon a background that performs several cycles. However, there cycles are *not* periodic, that is, *cannot* be extrapolated as the sample lengthens. In addition, one often sees an underlying trend that need not continue in the extrapolate." (Mandelbrot (1983), page 251).

The Nile river data set is a famous one; arguably, it is the data set that forced us to think about long range dependence in the first place. It was, of course, the same Mandelbrot who with co-workers (Mandelbrot (1965); Mandelbrot and Van Ness (1968) and Mandelbrot and Wallis (1968, 1969)) first realized that a long memory *stationary* Gaussian process may explain the behaviour a particular statistic (the so-called R/S statistic) suggested and applied to the Nile river data by Hurst (1951, 1955).

Today long memory models are still being used in hydrology and related areas. However, new applications have arisen, significantly in finance and communication networks. Often observations from these latter areas feature *heavy tails*, and such data sets sometimes provide extreme illustrations to the Mandelbrot remark on "spurious cycles". For example, Figure 1.2 describing the load offered by a network server, suggests that the process plotted there has at least 10 distinct time periods when the "nature" of the process changes. Once again, one should not automatically decide to use a nonstationary model because there are perfectly reasonable stationary models that have a similar behavior.

Stationary models are attractive not only because of parsimony but also because it is important to have a reasonably small class of well studied and well understood models that have wide applicability. Hence it is important to study stationary processes that can account for features we saw above; stationary processes with long range dependence.

This paper is an attempt to survey stationary models with long range dependence and heavy tails. These two features are believed to be present in various data sets of financial and communication networks origin and, hence, attracted recently much attention. Describing long range dependence in the heavy tailed case is especially challenging and most of the work is still ahead

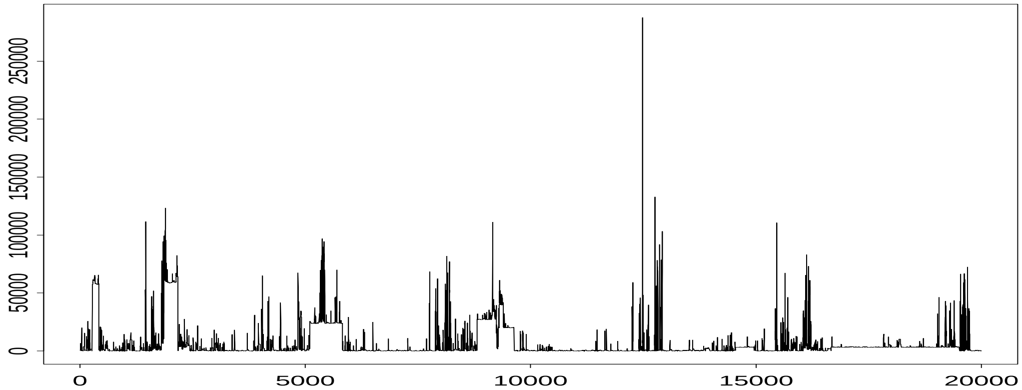

Figure 1.2. Amount of information (in bytes) sent by a server from a major telecommunication company in the middle of a workday. Time is measured in seconds.

of us. Nevertheless, it is an exciting task, and we argue that the insights we hope one will obtain are likely to be useful in other areas of stochastic modeling.

2. What is long range dependence?

The obvious way to measure the length of memory in a stochastic process is by looking at the rate at which its correlations decay with lag. Annoyingly, this requires correlations to make sense, hence finite variance needs to be assumed.

Let, therefore, X_n, $n = 0, 1, 2, \ldots$ be a stationary stochastic process with mean $\mu = EX_0$ and $0 < EX_0^2 < \infty$ (we discuss discrete time processes, but parallel formulations for stationary processes with finite variance in continuous time are entirely clear). Let $\rho_n = \text{Corr}(X_0, X_n)$, $n = 0, 1, \ldots$ be the correlation function. For most "usual" stochastic models: ARMA processes, GARCH processes, many Markov and Markov modulated processes the correlations decay exponentially fast with n; this has a number of important consequences, one of which is $\sum_{n=0}^{\infty} |\rho_n| < \infty$. This, in turn, guarantees that the variance of the partial sums $S_n = X_1 + \ldots + X_n$, $n \geq 0$ cannot grow more than linearly fast, which says, heuristically, that we do expect to see S_n to be more than about $\sqrt{n}$ away from its mean $n\mu$. What Mandelbrot realized two and half decades ago that the strange behavior of R/S statistic on the Nile river data might be explained if the variance of the partial sums could grow faster than linearly fast. As we already know this implies that

$$\sum_{n=0}^{\infty} |\rho_n| = \infty\,. \tag{2.1}$$

Hence, (2.1) is often taken as the definition of the long memory; as a definition it seems to originate with Cox (1984).

Looking over the literature on long range dependence, one realizes that the definition (2.1) has not proved to be the most popular one. By far the most widely used definition is the more concrete

$$\rho_n \sim cn^{-d} \quad \text{as } n \to \infty \text{ for some } 0 < d < 1 \text{ and } c > 0. \tag{2.2}$$

A stationary process satisfying (2.2) would have been called long range dependent of index d by Cox (1984). A weaker version of (2.2) is also sometimes mentioned; it replaces the constant c by a slowly varying function (Beran (1994)). Quite often one writes the exponent $d = 2-2H$ for some $.5 < H < 1$ (e.g. Beran (1992)), and the reasons are historical: this is the relation between the exponent H of self-similarity of a Fractional Brownian motion and the rate of decay of correlations of its increments. A misnomer, $H = 1-d/2$ is at times referred to as the self-similarity parameter even if nothing in the model is self similar.

Less common (but still used) point of view on long range dependence is to allow d in (2.2) to take any positive value or, indeed, a similar assumption of regular variation of correlations (which also allows for the slowly varying case $d = 0$). In fact, one could even draw the line between long and short memory by distinguishing between correlations decaying slower than exponentially fast and those decaying at least exponentially fast.

It is difficult to justify such importance assigned to the rate of decay of correlations (or, almost equivalently, to the rate at which the spectral density of the stationary process grows at the origin), *unless* one deals with a Gaussian model (like the Fractional Gaussian noise, the increment process of a Fractional Brownian motion), or a process that is very close to being Gaussian. Using correlations as a measure of the length of memory becomes untenable in the case of heavy tails. Specifically, let X_n, $n = 0, 1, 2, \ldots$ be a stationary stochastic process. Let F be the distribution function of X_0, and $\overline{F} = 1 - F$ the (right) distribution tail. In a tradition going back, once again, to Mandelbrot in the early 1960s (an exhaustive list of references is in Mandelbrot (1983)) heavy tails are synonymous with infinite variance of X_0. Once again, more concrete views are prevailing in literature; it is common to identify heavy tails with a particular tail behavior of $\overline{F}$. Sometimes one assumes

$$\overline{F}(x) \sim cx^{-\alpha} \quad \text{as } x \to \infty \text{ for some } 0 < \alpha < 2 \text{ and } c > 0. \tag{2.3}$$

See, for example, the recent collection Park and Willinger (2000). Note the parallels between the various definitions of heavy tails and viewing long range dependence via the rate of decay of correlations. Again, one allows sometimes any positive value of α in (2.3) (see e.g. Müller et al. (1998) or Gomez et al. (1997)). Here regular variation of the tails as opposed to power-like (*Pareto-type* is the common expression) is widely accepted. Finally, faster (but still slower than exponentially fast) rate of decay of the distribution tail is sometimes also regarded as being consistent with heavy tails. Here one does not usually go beyond the class of *subexponential distributions*; see Embrechts et al. (1997).

Obviously, one cannot use correlations to draw the line between short and long memory if the variance is infinite. Several attempts have been made to use "correlation–like" notions in that case. In the important class of stable processes notions of *covariation* and *codifference* have been introduced and their rate of decay for various classes of stationary stable processes computed; see e.g. Astrauskas et al. (1991). Obviously, such "surrogate correlations" can be expected to carry even less information than the "real" correlations do in the case when the latter are defined (although, surprisingly, codifference turns out to characterize mixing of stationary symmetric stable processes, a fact due, essentially, to Maruyama (1970), see also Gross (1994)).

Before finishing this section we remark that when talking about tails we are thinking about the *right tails*. For as long as the left tails do not interfere with the right tails, we will leave it that way. When right and left tails begin to interfere with one another, we will need to say more about the left tail and how heavy they are as well.

3. Tails and rare events

Here is an alternative point of view on long range dependence in heavy tailed processes. Most practitioners using heavy tailed models will agree that the most important feature of such processes is precisely their tails as expressed in probabilities of various *rare events*. Risk analysis, ruin probabilities, congestion and overflow analysis are just some of the key words that name such rare events in various modern applications. To be a bit more concrete here are several specific examples of rare events one usually deals with. Let, once again, X_n, $n = 0, 1, 2, \dots$ be a stationary stochastic process.

Example 3.1. For large $\lambda > 0$ the event $\{X_0 > \lambda\}$ is a rare event, whose probability is clearly related to the tails. This event is so elementary that it does not tell us anything about the memory in the process.

Example 3.2. For $k \geq 1$ and large $\lambda_0, \lambda_1, \dots, \lambda_k$ the event $\{X_0 > \lambda_0, X_1 > \lambda_1, \dots, X_k > \lambda_k\}$ is a rare event whose probability can carry very important information about the dependence in finite pieces of the process. Generally, the dependence we can measure using such rare events is a "tail dependence". However, for specific classes of heavy tailed processes (e.g. stable processes, linear processes, etc.) these events can provide even more information.

Example 3.3. For large $n \geq 1$ and a positive sequence $(\lambda_j)_{j\geq 0}$ that does not converge to zero the event $\{X_j > \lambda_j,\, j = 0, 1, \ldots, n\}$ is a rare event and its probability is a very interesting measure of the length of memory in the process. The case $\lambda_j = \lambda > 0$ for all $j \geq 0$ seems to be especially appealing.

A slight generalization of this example uses a triangular array $(\lambda_j^{(n)})_{n\geq 1, 0\leq j\leq n}$. Here the case $\lambda_j^{(n)} = \lambda^{(n)}$ for $j \leq n$ with various asymptotic rules for $(\lambda^{(n)})$ is very interesting.

Example 3.4. For $k \geq 1$ and large λ the event $\{X_1 + \ldots + X_k > \lambda\}$ is a tail event. Similarly to Example 3.2 the probability of this event can be used to clarify the "finite dimensional dependence" in the process.

Example 3.5. Suppose that the mean $\mu = EX_0$ is finite, and that the stationary process X_n, $n = 0, 1, 2, \ldots$ is ergodic. For large $n \geq 1$ and $\delta > 0$ the event $\{X_1 + \ldots + X_n > n(\mu + \delta)\}$ is a rare event whose probability measures the length of memory in the sense of a tendency of being over the mean for long stretches of time. It is, obviously, related to the tails. The effect of heavy tails is quite special, as will be discussed below.

Example 3.6. This example has a flavor similar to that of Example 3.5. Let, once again, the process X_n, $n = 0, 1, 2, \ldots$ be ergodic with a finite mean $\mu = EX_0$. Let $\delta > 0$. For large λ the event $\{X_1 + \ldots + X_n > n(\mu + \delta) + \lambda \text{ for some } n \geq 1\}$ is a rare event, whose probability is sometimes referred to as *ruin probability* in the context of risk analysis. In the queuing context various stationary quantities often have expressions of this kind for their probability tails. Adopting the risk analysis term, the ruin probability can be used to measure the length of memory; the effect of heavy tailed case is, once again, very special here.

The list of examples can be continued indefinitely, and we have omitted some very interesting ones. Instead, let us look at some details of the interplay between the tails, memory and rare events in the heavy tailed case, especially in the light of Examples 3.5 and 3.6. The starting point is to adopt the lenses of large deviations: *an unlikely event happens in the most likely way.* We will argue that such lenses provide a powerful way of thinking about the length of memory in a process. It is unfortunate that this idea is not made more explicit in many beautiful texts on large deviations (that also reserve the term "large deviation principle" to something else); see e.g. Deuschel and Stroock (1989) and Dembo and Zeitouni (1993). The following statement is not a rigorous mathematical statement. Nevertheless, it is often very useful as a guide and, in many ways, it captures the essence of heavy tails:

the most likely way tail related rare events happen in a heavy tailed stochastic process is because of the smallest possible number of causes.

This "smallest possible number of causes" is often equal to one.

Thus, in Example 3.4 it turns out that, if $X_1, \ldots, X_k$ are i.i.d. and heavy tailed, then

$$P(X_1 + \ldots + X_k > \lambda) \sim kP(X_1 > \lambda) \sim P(\max(X_1, \ldots, X_k) > \lambda) \quad \text{as } \lambda \to \infty. \tag{3.1}$$

That is, the sum $X_1 + \ldots + X_k$ is most likely to be very large due to one of the terms being very large. In this case the possible "causes" are simply the individual terms in the sum. The greatest generality under which (3.1) is valid is that of subexponential distributions, introduced by Chistyakov (1964). See also Chover et al. (1973), and a survey in Goldie and Klüppelberg (1998). Similarly, in Example 3.5, for every $\delta > 0$

$$P(X_1 + \ldots + X_n > n(\mu + \delta)) \sim nP(X_1 > n\delta) \quad \text{as } n \to \infty \tag{3.2}$$

for exactly the same reason as in (3.1). Indeed, one of the terms ($\equiv$ causes) in the sum $X_1 + \ldots + X_n$ has to be exceptionally large; exactly how large can be determined by realizing that the "nonexceptional" terms in that sum add up to about $n\mu$. While the domain of heavy tails over which (3.2) is valid does not extend to all subexponential distributions, it does extend to all distributions with regularly varying tails of index $\alpha > 1$; see e.g. Heyde (1968) and Nagaev (1979).

On the other hand, for distributions with "light" tails not only (3.1) and (3.2) fail, even their spirit is false. In fact, in the case of exponentially fast decaying tails the most likely way for the event $\{X_1 + \ldots + X_n > n(\mu + \delta)\}$ to happen is not because of a single cause, or a small number of causes but, rather, because most of the terms in the sum "conspire" to be a bit bigger than they would normally be. This is, in fact, the point of the classical large deviation principle.

When X_n, $n = 0, 1, 2, \ldots$ is a stationary heavy tailed stochastic process *with memory*, it is not, generally, the case that individual observations should be viewed as "causes" of rare events. The nature of such causes depends on the nature of the process and it is, sometimes, a nontrivial problem to figure out what the "right causes" are. We will see several examples below. Moreover, and this is precisely the point why we are interested in rare events, the causes, when found, typically have their effect distributed over time and it is in this way that they make the rare events happen. We argue that *this temporal distribution of the effect of the "causes" on rare events is a useful way of thinking about long range dependence.*

There are two important classes of heavy tailed processes for which progress has been made in understanding the "right causes" of certain rare events and the way the effect of these causes is distributed over time: linear processes and infinitely divisible processes. We discuss these below. Before doing so we would like to introduce another notion related to certain rare events with a potential of being useful, in a similar way, in studying long range dependence.

Certain rare events should be rather viewed as sequences of events that become more and more rare. Examples 3.3 and 3.5 are of this nature. More generally and formally, let $A_j \in \mathbb{R}^j$ be a Borel set, $j = 1, 2, \ldots$ such that

$$p_j := P\left((X_1, \ldots, X_j) \in A_j\right) \to 0 \quad \text{as } j \to \infty. \tag{3.3}$$

For $n \geq 1$ define

$$R_n = \max\left\{j - i + 1 : 1 \leq i \leq j \leq n,\ (X_i, X_{i+1} \ldots, X_j) \in A_{j-i+1}\right\}. \tag{3.4}$$

That is, R_n is the highest dimension of an A_j observed over the first n observations $X_1, \ldots, X_n$. We call R_n the functional associated with the sequence of rare events (A_j).

It is obvious that if X_n, $n = 0, 1, 2, \ldots$ is a mixing stationary process and $p_j > 0$ for an infinite sequence of j's then $R_n \to \infty$ with probability 1 as $n \to \infty$. It appears to be almost obvious that the *rate* at which R_n grows is related to the rate at which p_j decays to zero. Certain rigorous connections are, indeed, possible; other connections seem to require additional information on the process. In any case, the rate of growth of R_n is, in its own right, related to the way rare events happen and, hence, to the memory in the process.

There is a very important reason to concentrate on the probabilities of certain rare events and on functionals associated with sequences of certain rare events, instead of concentrating on correlations, when trying to understand the boundary between short memory and long memory. Such rare events and functionals are often of a direct importance on their own right, as one can see by looking at the examples above and thinking, for instance, of applications in risk analysis and congestion control. On the other hand, nobody is interested in correlations on their own right. We only study correlations hoping that they are significant for whatever application we might have at hand. Unfortunately, the information that the correlations carry is often only indirect and very limited, as anyone familiar, for example, with ARCH and GARCH models realizes.

4. Some classes of heavy tailed processes

4.1. **Linear processes.** One of the classes of heavy tailed processes we will consider is that of *heavy tailed linear processes.*

Let ε_n, $n \in \mathbb{Z}$ be iid random variables. A (two–sided) linear process with the noise sequence ε_n, $n \in \mathbb{Z}$ is defined by

$$X_n = \sum_{j=-\infty}^{\infty} \varphi_{n-j}\, \varepsilon_j\,, \quad n = 0, 1, 2, \dots\,, \tag{4.1}$$

where φ_j, $j \in \mathbb{Z}$ is a sequence of (nonrandom) coefficients. We will assume that the noise variables are heavy tailed, but how heavy the tails are will be left open at the moment. It is obvious that the linear process X_n, $n = 0, 1, 2, \dots$ is a stationary stochastic process as long as it is well defined, meaning that the sum defining it converges. The latter is an assumption on the coefficients φ_j. In particular, if $E\varepsilon_0^2 < \infty$ and $E\varepsilon_0 = 0$, then a necessary and sufficient condition for convergence of the series in (4.1) is

$$\sum_{j=-\infty}^{\infty} \varphi_j^2 < \infty; \tag{4.2}$$

a nonzero mean will require, in addition, the series $\sum_{j=-\infty}^{\infty} \varphi_j$ to converge. Frequently we will assume that the noise variables have regularly varying tails. Unless one is working with constant sign coefficients (an assumption that we will not make in this paper), it is necessary to control both right and left probability tails of the noise since, say, a negative coefficient will "translate" the left tail of the noise into the right tail of the sum in (4.1). Therefore, a typical assumption is

$$\begin{cases} P(|\varepsilon_0| > \lambda) = L(\lambda)\, \lambda^{-\alpha}\,, \\ \lim_{\lambda\to\infty} \dfrac{P(\varepsilon_0 > \lambda)}{P(|\varepsilon_0| > \lambda)} = p, \quad \lim_{\lambda\to\infty} \dfrac{P(\varepsilon_0 < -\lambda)}{P(|\varepsilon_0| > \lambda)} = q\,, \end{cases} \tag{4.3}$$

as $\lambda \to \infty$, for some $\alpha \geq 0$ and $0 < p = 1 - q \leq 1$. Here L is a slowly varying (at infinity) function. If $\alpha > 2$ we are in the case of finite variance, but for $\alpha \leq 2$ the precise condition for convergence in (4.1) depends on the slowly varying function, and can be stated through the three series theorem. In particular,

$$\sum_{j=-\infty}^{\infty} |\varphi_j|^{\alpha-\epsilon} < \infty \tag{4.4}$$

for some $\epsilon > 0$ is a sufficient condition for convergence if $0 < \alpha \leq 1$ or if $1 < \alpha \leq 2$ and $E\varepsilon_0 = 0$; a nonzero mean in the latter case will also require, as before, the series $\sum_{j=-\infty}^{\infty} \varphi_j$ to converge.

A rich source of information on linear processes in Brockwell and Davis (1991). This book covers, mostly, the L^2 case. For more information on the infinite variance case see, for example, Cline (1983, 1985) and Mikosch and Samorodnitsky (2000b).

Heavy tailed linear processes are attractive to us because, in this case, the potential "causes" of rare events appear to be evident: those are the individual noise variables ε_n, $n \in \mathbb{Z}$. This intuition has been born out in a number of situations, as will be seen below.

4.2. **Infinitely divisible processes.** A stochastic process X_n, $n = 0, 1, 2, \ldots$ is *infinitely divisible* if for any $k = 1, 2, \ldots$ there is a stochastic process $Y_n^{(k)}$, $n = 0, 1, 2, \ldots$ such that the finite dimensional distributions of X_n, $n = 0, 1, 2, \ldots$ and of $\sum_{i=1}^k Y_n^{(k,i)}$, $n = 0, 1, 2, \ldots$ coincide. Here for $i = 1, \ldots, k$, the processes $Y_n^{(k,i)}$, $n = 0, 1, 2, \ldots$, are iid copies of $Y_n^{(k)}$, $n = 0, 1, 2, \ldots$. Many important classes of stochastic processes are, in fact, infinitely divisible. All Gaussian processes, and all stable processes in particular, are infinitely divisible. In general, an infinitely divisible process will have two independent components, a Gaussian one and a non-Gaussian one. Since we are interested in heavy tails, for a vast majority of applications the Gaussian component will have only a negligible effect on the probabilities of rare events we consider. Therefore, we will only consider infinitely divisible processes without a Gaussian component. Such processes have a characteristic function of the form

$$E \exp\left\{ \sum_{n=0}^{\infty} \theta_n X_n \right\} \tag{4.5}$$

$$= \exp\left\{ \int_{\mathbb{R}^\infty} \left(\exp\left\{ \sum_{n=0}^{\infty} \theta_n x_n \right\} - 1 - i \sum_{n=0}^{\infty} \theta_n x_n \mathbf{1}(|x_n| \le 1) \right) \nu(d\mathbf{x}) + i \sum_{n=0}^{\infty} \theta_n b_n \right\}$$

for all θ_n, $n = 0, 1, 2, \ldots$ only finitely many of which are different from zero. Here ν is a σ-finite measure on $\mathbb{R}^\infty$ equipped with the product σ-field (the *Lévy measure of the process*) and b_n, $n = 0, 1, 2, \ldots$ is a constant vector in $\mathbb{R}^\infty$.

The Lévy measure of an infinitely divisible process is its most important feature. Often an infinitely divisible process is given in the form of a stochastic integral with respect to an infinitely divisible random measure. In that case there is a natural way to relate the Lévy measure of the process to the basic characteristics of such an integral.

Unlike the linear processes in the previous subsection, it is less obvious what are the potential "causes" of rare events when one deals with infinitely divisible processes as above. There is, however, a point of view on infinitely divisible processes that turns out to be useful here. To be able to see the essence better and not to get bogged in the technical details, let us consider, first, a particular case, when

$$\int_{\mathbb{R}^\infty} x_n \mathbf{1}(|x_n| \le 1) \nu(d\mathbf{x}) < \infty \quad \text{for all } n = 0, 1, 2 \ldots. \tag{4.6}$$

In that case one can rewrite (4.5) in the form

$$E \exp\left\{\sum_{n=0}^{\infty}\theta_n X_n\right\} = \exp\left\{\int_{\mathbb{R}^\infty}\left(\exp\left\{\sum_{n=0}^{\infty}\theta_n x_n\right\}-1\right)\nu(d\mathbf{x}) + i\sum_{n=0}^{\infty}\theta_n b_n'\right\} \tag{4.7}$$

with $b_n' = b_n - \int_{\mathbb{R}^\infty} x_n \mathbf{1}(|x_n|\le 1)\nu(d\mathbf{x})$ for $n\ge 0$.

Let M be a Poisson random measure on $\mathbb{R}^\infty$ with mean measure ν. It is easy to check that the process $\int_{\mathbb{R}^\infty} x_n M(d\mathbf{x}) - b_n'$ for $n \ge 0$ is well defined and has characteristic function given by (4.7). That is, one can represent the process X_n, $n = 0,1,2,\dots$ in the sense of equality of finite dimensional distributions in the form

$$X_n = \int_{\mathbb{R}^\infty} x_n\, M(d\mathbf{x}) - b_n', \quad n = 0,1,2,\dots . \tag{4.8}$$

If $(\mathbf{z}^{(j)} = (z_n^{(j)}, n\ge 0),\ j = 1,2,\dots)$ is a (measurable) enumeration of the points of the random measure M, then (4.8) means that the process X_n, $n = 0,1,2,\dots$ is the sum of $(\mathbf{z}^{(j)})$, $j = 1,2,\dots$ (shifted by the sequence (b_n')). This "discrete" structure of infinitely divisible processes makes the potential "causes" of certain rare events visible, and it is precisely the Poisson points $\big((\mathbf{z}^{(j)}),\ j = 1,2,\dots\big)$ that turn out to be such "causes".

Even if the assumption (4.6) does not hold, then a representation similar to (4.8) can still be written, but this time an appropriate centering is required to make the Poisson integral to converge. The important point is that the discrete structure is still here, and the potential causes of rare events are still visible.

There are various ways of summing up the Poisson points to get an infinitely divisible process. A very general description is in Rosiński (1989, 1990). Sometimes it is convenient to order the Poisson points according to the value of a particular test functional. If the process is originally given in the form of a stochastic integral with respect to an infinitely divisible random measure, then one can have a more concrete structure of the Poisson points, hence better understanding of the possible causes of rare events.

The literature on infinitely divisible processes is rich. The framework preferred by many authors is that of infinitely divisible probability laws on Banach (or other nice) spaces. See for example Araujo and Giné (1980) and Linde (1986). A very general treatment of stochastic integrals with respect to infinitely divisible random measures as well as representations of infinitely divisible processes as such stochastic integrals is in Rajput and Rosiński (1989).

An important and reasonably well understood class of infinitely divisible processes is that of *α-stable processes.* The latter are characterized by the following scaling property of their Lévy measure:

$$\nu(rA) = r^{-\alpha}\nu(A) \quad \text{for all mesurable } A \in \mathbb{R}^\infty \text{ and } r > 0. \tag{4.9}$$

Here α is a parameter with the range $0 < \alpha < 2$. See Samorodnitsky and Taqqu (1994) for information on stable processes; the structure of stationary stable processes has been elucidated by J. Rosinski; see e.g. Rosiński (1998).

5. Rare events, associated functionals and long range dependence

Suppose that we are considering a parametric family of laws of a stationary stochastic process X_n, $n = 0, 1, 2, \ldots$. Let Ξ be the (generally, infinite-dimensional) parameter space. We are interested in significant changes ("phase transitions") in the rate of decay of probabilities of certain rare events and/or in the rate of growth of the functionals associated with sequences of rare events that may occur when the parameter ξ crosses the boundary between a subset Ξ_1 of Ξ and its complement. We argue that *certain phase transitions of this kind can be viewed as transitions between short and long range dependence.*

It is clear that it is not useful to view *every* significant change in, say, probabilities of rare events as an indication of interesting and important things happening to the memory of the process. Other factors may be in play as well, most significantly related to the heaviness of the tails. If, for example, one of the components of parameter $\xi \in \Xi$ governs how heavy the tails of X_0 are, one can very easily induce a very significant change in the probabilities of certain rare events by simply changing that particular component of the parameter without doing anything to the memory of the process. In the examples in the sequel we will be careful to look for phase transitions that do not involve changing how heavy the tails are.

We will see several examples of such phase transitions indicating a shift from short to long memory below. We present some known results; these are quite scarce. When appropriate, we supplement those with conjectures. In other cases we have performed numerical studies to try to guess whether a phase transition occurs and, if so, of what kind.

5.1. **Unusual sample mean and long strange segments for heavy tailed linear processes.** Here we consider the sequence of rare events of the Example 3.5 $A_n = \{X_1 + \ldots + X_n > n(\mu + \delta)\}$ (for a fixed $\delta > 0$) and the corresponding associated functional

$$R_n = \max\left\{ j - i + 1 : 1 \le i \le j \le n,\ \frac{X_i + X_{i+1} + \ldots + X_j}{j - i + 1} > \mu + \delta \right\}. \tag{5.1}$$

We will keep the distribution of the noise variables ε_n, $n \in \mathbb{Z}$ in the heavy tailed linear processes of Subsection 4.1 fixed; it is assumed to have the regular variation property (4.3) with $\alpha > 1$. In particular, the parameter α which is responsible for the heaviness of the tails is kept fixed. We will also assume that the $E\varepsilon_0 = 0$. In this case the parameter space is

(5.2)
$$\Xi = \Big\{\boldsymbol{\varphi} = (\dots, \varphi_{-1}, \varphi_0, \varphi_1, \varphi_2, \dots) \in \mathbb{R}^{\mathbb{Z}}, \ \text{ satisfying (4.2) if } \alpha > 2 \text{ or (4.4) if } 1 < \alpha \leq 2.\Big\}$$

Let $\Xi_1 \subset \Xi$ be the set of all sequences $\boldsymbol{\varphi} \in \mathbb{R}^{\mathbb{Z}}$ satisfying

$$\sum_{j=-\infty}^{\infty} |\varphi_j| < \infty \,. \tag{5.3}$$

Note that the set Ξ_1 contains the parameter sequence $\varphi_j = \mathbf{1}(j = 0)$, $j \in \mathbb{Z}$, in which case the linear process is an iid sequence.

It turns out that for any value of the parameters in Ξ_1 the functionals R_n defined by (5.1) grow at the same rate, i.e. at the same rate as for an iid sequence with the same marginal tails. This has been established in Mansfield et al. (1999). Specifically, let F be the distribution function of the noise random variable ε_0 and define the usual quantile sequence

$$a_n = \left(\frac{1}{1-F}\right)^{\leftarrow}(n)\,. \tag{5.4}$$

Here for a function U on $[0, \infty)$, $U^{\leftarrow}$ denotes its generalized inverse

$$U^{\leftarrow}(y) = \inf\{s :\ U(s) \geq y\}\,.$$

Note that, by (4.3), the sequence (a_n) is regularly varying at infinity with exponent $1/\alpha$. See Resnick (1987) for more information on regular varying tails and their quantile functions.

For $\beta > 0$ let Z_β be a Fréchet random variable with

$$P(Z_\beta > z) = \exp\left\{-z^{-\beta}\right\}, \quad z > 0\,. \tag{5.5}$$

Assume (5.3). Then the numbers

$$\begin{cases} M_+(\boldsymbol{\varphi}) = \max\left\{\sup_{-\infty<k<\infty}\left(\sum_{j=-\infty}^{k}\varphi_j\right)_+ ,\ \sup_{-\infty<k<\infty}\left(\sum_{j=k}^{\infty}\varphi_j\right)_+\right\} \\ M_-(\boldsymbol{\varphi}) = \max\left\{\sup_{-\infty<k<\infty}\left(\sum_{j=-\infty}^{k}\varphi_j\right)_- ,\ \sup_{-\infty<k<\infty}\left(\sum_{j=k}^{\infty}\varphi_j\right)_-\right\} \end{cases}, \tag{5.6}$$

are, obviously, finite. Then

$$a_n^{-1}R_n \Rightarrow \delta^{-1}\Big(pM_+(\boldsymbol{\varphi})^\alpha + qM_-(\boldsymbol{\varphi})^\alpha\Big)^{1/\alpha} Z_\alpha \ \ \text{(weakly) as } n \to \infty\,, \tag{5.7}$$

once again as long as (5.3) holds. Here p and q are the tail weights in (4.3). See Theorem 2.1 in Mansfield et al. (1999).

What happens if $\boldsymbol{\varphi} \in \Xi_1^c$ (i.e. if (5.3) fails)? It is not known whether, in this case, R_n *always* grows at the rate faster than a_n, that is whether the sequence (of the laws of) $(a_n^{-1}R_n, n = 1, 2, \dots)$

is not tight. However, the following is known. Assume that the coefficients (φ_j) are themselves regularly varying and balanced. That is, there is a function $\varphi:\ [0,\infty)\to[0,\infty)$ such that

$$\varphi(t) = L_2(t)\, t^{-h} \tag{5.8}$$

as $t\to\infty$ and such that

$$\lim_{j\to\infty}\frac{\varphi_j}{\varphi(j)} = c_+, \quad \lim_{j\to\infty}\frac{\varphi_{-j}}{\varphi(j)} = c_-\,, \tag{5.9}$$

for some $c_+, c_- \geq 0$, at least one of which is positive. Here

$$h > \max\{\frac{1}{\alpha}, \frac{1}{2}\} \tag{5.10}$$

and L_2 is a slowly varying function. Clearly any such parameter vector $\boldsymbol{\varphi}$ is in Ξ_1^c. Define

$$b_n = \left(\frac{1}{\varphi}\right)^{\leftarrow}(a_n)\,, \tag{5.11}$$

$n\geq 1$ and note that sequence (b_n) is regularly varying with exponent $1/(\alpha h)$. Then

$$b_n^{-1}R_n \Rightarrow p^{1/\alpha h}\Big((1-h)\delta\Big)^{-1/h}\Big(c_+^{1/h}+c_-^{1/h}\Big)Z_{\alpha h} \ \ \text{(weakly) as } n\to\infty\,. \tag{5.12}$$

See Theorem 2.1 in Rachev and Samorodnitsky (2001).

Since b_n grows faster than a_n does, under the assumptions (5.9) the sequence R_n does grow faster than a_n and, hence, faster than in the iid case and, more generally, faster than it is the case for any $\boldsymbol{\varphi}\in\Xi_1$.

Both results (5.7) and (5.12) are, in the final analysis, a consequence of change in the temporal distribution of the effect of the individual "causes": exceptionally large or exceptionally small values of the noise variables (ε_m). In fact, the contribution of each individual noise variable ε_m to the sum $X_i + X_{i+1} + \ldots + X_j$ in (5.1) is $\varepsilon_m \sum_{d=i-m}^{j-m}\varphi_d$. The intuition of heavy tailed large deviations says that it is a single ε_m that is most likely to be responsible for a large value of R_n. Therefore, one would expect that for large x_n

$$P(R_n > x_n) \sim P\Bigg(\text{for some } m=\ldots,-1,0,1,\ldots\ \left(\sum_{d=i-m}^{j-m}\varphi_d\right)\varepsilon_m > (j-i+1)(\mu+\delta) \\ \text{for some } 1\leq i\leq j\leq n,\ j-i+1\geq x_n\Bigg)\,. \tag{5.13}$$

This turns out to be valid. Moreover, this intuition allows one, in both cases (i.e. under (5.3) and under (5.9)) to select the right rate of growth for x_n in (5.13), which is equivalent to selecting the appropriate normalization to R_n.

It is a bit surprising that less is known about the apparently easier problem of identifying the rate of decay of probabilities $p_n = P\big(X_1 + \ldots + X_n > (\mu+\delta)n\big)$ for $\delta > 0$ as $n \to \infty$. It has been checked that under the assumption

$$\sum_{j=-\infty}^{\infty} j|\varphi_j| < \infty \tag{5.14}$$

which defined a proper subset of Ξ_1,

$$p_n \sim n^{-(\alpha-1)} L(n) \delta^{-\alpha} \left(p \left(\sum_{j=-\infty}^{\infty} \varphi_j \right)_+^{\alpha} + q \left(\sum_{j=-\infty}^{\infty} \varphi_j \right)_-^{\alpha} \right) \quad \text{as } n \to \infty, \tag{5.15}$$

where p, q and L are defined in (4.3), and one assumes that $q > 0$ if $\sum_{j=-\infty}^{\infty} \varphi_j < 0$. See Lemma A.5 in Mikosch and Samorodnitsky (2000b). It looks very plausible that (5.15) holds for every parameter $\boldsymbol{\varphi} \in \Xi_1$. The logic of large deviations indicates that, under the assumptions (5.9), p_n is regularly varying with exponent $-(\alpha h - 1)$ at infinity, but nobody has presented a rigorous proof so far.

5.2. **Ruin probability for heavy tailed linear processes.** In this subsection we consider the rare event in the Example 3.6, $A = \{X_1 + \ldots + X_n > n(\mu+\delta) + \lambda \text{ for some } n \geq 1\}$, when $\delta > 0$ is fixed and λ is large. Unfortunately, the result for the entire set Ξ_1 is not available here. However, there is a result for the subset of Ξ_1 defined by (5.14). In the latter case, the probability of the event A (commonly referred to as the ruin probability) satisfies

$$P(A) \sim \frac{p M_+^{(1)}(\boldsymbol{\varphi})^{\alpha} + q M_-^{(1)}(\boldsymbol{\varphi})^{\alpha}}{\delta(\alpha-1)} \lambda^{-(\alpha-1)} L(\lambda) \quad \text{as } \lambda \to \infty\,, \tag{5.16}$$

where

$$M_+^{(1)}(\boldsymbol{\varphi}) = \sup_{-\infty < k < \infty} \left(\sum_{j=-\infty}^{k} \varphi_j \right)_+ , \quad M_-^{(1)}(\boldsymbol{\varphi}) = \sup_{-\infty < k < \infty} \left(\sum_{j=-\infty}^{k} \varphi_j \right)_- , \tag{5.17}$$

compare with (5.6). See Theorem 2.1 in Mikosch and Samorodnitsky (2000b). We conjecture that (5.16) holds whenever $\boldsymbol{\varphi} \in \Xi_1$. Once again, a good way to think of the asymptotic behavior of the ruin probability is to think about the most likely way the "ruin" can happen. Realizing that the ruin is, most likely, due to a single "extraordinary" value of a noise variable ε_m, one would expect that

$$P(A) \sim \sum_{m=-\infty}^{\infty} P\left(\left(\sum_{d=1-m}^{n-m} \varphi_d \right) \varepsilon_m > n\delta + \lambda \ \text{ for some } n \geq 1 \right) . \tag{5.18}$$

Once again, this turns out to be valid (at least, under the assumption (5.14)).

The problem of the behaviour of the ruin probability for $\xi \in \Xi_1^c$ has not, to the best of our knowledge, been treated. One can pursue the logic of large deviations, leading to (5.18). This leads us to conjecture that, under the assumptions (5.9), $P(A)$ is, as a function of λ, regularly varying with exponent $-(\alpha h - 1)$ at infinity.

Based on the above discussion (admittedly, some part of it is "hard" results, and another part is conjectures) one can argue that a significant change occurs for heavy tailed linear processes as parameter θ crosses the boundary between Ξ_1 and its complement. Not only the order of magnitude of the probabilities of certain rare events, and of certain functionals associated with sequences of certain rare events, appears to change at that boundary but another interesting phenomenon seems to happen. Various orders of magnitude do not change as the parameter varies inside of Ξ_1; not only these orders of magnitude do change at the boundary but, also, they may keep changing as the parameter varies outside of Ξ_1.

It is important to make a remark at this moment. It does appear that one should, in fact, look at the behavior of a family of related rare events, or a family of sequences of related rare events, if one wants to see what precisely happens at a boundary. For example, the assumptions (5.9) do not cover the entire Ξ_1^c. We conjecture, however, that important changes happen when one moves from Ξ_1 into Ξ_1^c and not, necessarily, into the subset of Ξ_1^c defined by (5.9). It is likely that, in order to see these changes, one should look not only, say, at the event $A_n = \{X_1 + \ldots + X_n > n(\mu + \delta)\}$ but also at some related rare events, for example at the event $B_n = \{|X_1| + \ldots + |X_n| > n(\mu_1 + \delta)\}$, with $\mu_1 = E|X_1|$.

It is also interesting to mention that, in the case $\alpha > 2$, the condition (5.3) also implies the absolute summability of correlations (i.e. (2.1) fails).

5.3. **Rare events for stationary stable processes.** The situation regarding "phase transitions" for general stationary heavy tailed infinitely divisible processes of Subsection 4.2 has been investigated even less than it is the case with the heavy tailed linear processes. There are several reasons for this, including relatively complicated structure of stationary infinitely divisible processes and its very involved parameter space, which is a space of measures. Most of the known results are for stable processes, whose structure is better understood. We present here the results for a subclass of stationary stable processes, where we will be able to see a "phase transition".

Specifically, let X_n, $n = 0, 1, 2, \ldots$ be the linear fractional symmetric α-stable noise, $1 < \alpha < 2$. For a fixed α the law of the process has an important parameter $H \in (0, 1)$. That is,

$$X_n = \int_{\mathbb{R}} f_n(x)\, M(dx), \quad n = 0, 1, 2, \ldots , \tag{5.19}$$

where M is a symmetric α-stable random measure on the real line with the Lebesgue control measure, and

$$f(x) = a\left((-x)_+^{H-1/\alpha} - (-x-1)_+^{H-1/\alpha}\right) + b\left((-x)_-^{H-1/\alpha} - (-x-1)_-^{H-1/\alpha}\right) \tag{5.20}$$

if $H \in (0,1)$, $H \neq 1/\alpha$. Here a and b are real numbers not simultaneously equal to zero. For $H = 1/\alpha$ one has two choices,

$$f(x) = a\mathbf{1}\Big([-1,0]\Big)(x) \tag{5.21}$$

and

$$f(x) = a(\ln|x| - \ln|x+1|)\,. \tag{5.22}$$

In the latter two cases a is a real number different from zero. The resulting symmetric α-stable process in (5.19) is an ergodic stationary process. It is the increment process of the linear fractional symmetric α-stable motion if $H \neq 1/\alpha$, an iid sequence ($\equiv$ the increment process of the symmetric α-stable Lévy motion) under (5.21), and the increment process of the log- fractional symmetric α-stable motion under (5.22). All of these processes are H-self-similar with stationary increments. We refer the reader to Samorodnitsky and Taqqu (1994) for information on stable processes, their integral representations and on self-similar processes. The parameter space Ξ is, then, the collection of all triples (H,a,b) with $H \in (0,1)$, $H \neq 1/\alpha$, and a,b real, $a^2+b^2>0$, together with the triples (H,a,i) with $H = 1/\alpha$, a real, different from zero, and $i = 1,2$, depending on the choice between (5.21) and (5.22). Let Ξ_1 be the subset of Ξ corresponding to $0 < H < 1/\alpha$.

We consider, once again, the rare event in the Example 3.6, $A = \{X_1 + \ldots + X_n > n(\mu+\delta)+\lambda$ for some $n \geq 1\}$, when $\delta > 0$ is fixed and λ is large. Of course $\mu = 0$ here. Then

$$P(A) \sim \begin{cases} \frac{K}{\delta}\lambda^{-(\alpha-1)} & \text{if } \ 0 < H < 1/\alpha \ \text{ or under (5.21)} \\ \frac{K}{\delta}\,\lambda^{-(\alpha-1)}(\log\lambda)^{\alpha} & \text{under (5.22)} \\ \frac{K}{\delta^{\alpha H}}\lambda^{-\alpha(1-H)} & \text{if } \ 1/\alpha < H < 1 \end{cases} \tag{5.23}$$

as $\lambda \to \infty$. Here K is a finite positive constant that depends on α, H, a and b, but not on δ. See Proposition 4.4 in Mikosch and Samorodnitsky (2000a).

Observe that the order of magnitude of the ruin probability remains the same as H varies in $(0,1/\alpha)$. Furthermore, this order of magnitude is the same as under independence. On the other hand, as H varies in the interval $(1/\alpha, H)$, the order of magnitude of the ruin probability is greater than that in the case of independence and, furthermore, *this order of magnitude changes with* H. As we argued earlier, this gives us a reason to say that the range $H \in (0,1/\alpha)$ corresponds to short memory, and the range $H \in (1/\alpha,1)$ corresponds to long memory. It is interesting that, in this case, the boundary $H = 1/\alpha$ contains two points, corresponding to (5.21) and to (5.22), and it makes sense to view the latter as corresponding to long memory, while the former is the independent case.

Here is how the intuition of large deviations works here. As mentioned in Subsection 4.2, the process X_n, $n = 0, 1, 2, \ldots$ can be represented as a sum of Poisson points. In the symmetric stable case this can be done as follows. One can write (in terms of equality of finite dimensional distributions) the process given by (5.19) in the form

$$X_n = C_\alpha^{1/\alpha} \sum_{j=1}^{\infty} \varepsilon_j \, \Gamma_j^{-1/\alpha} \, g(V_j)^{-1/\alpha} \, f_n(V_j), \quad n = 0, 1, 2, \ldots, \tag{5.24}$$

where C_α is a finite positive constant that depends only on α, g a strictly positive measurable function such that $\int_{\mathbb{R}} g(x)dx = 1$, $(\varepsilon_n)_{n\geq 1}$ is an iid sequence of Rademacher variables ($P(\varepsilon_n = -1) = P(\varepsilon_n = 1) = 1/2$), $(\Gamma_n)_{n\geq 1}$ are the points of a unit rate Poisson process on $(0, \infty)$, and $(V_n)_{n\geq 1}$ is an iid sequence of real valued random variables with common density g. Moreover, the three sequences are mutually independent. See Samorodnitsky and Taqqu (1994), Section 3.10.

Rewriting

$$P(A) = P\left(C_\alpha^{1/\alpha} \sup_{n\geq 1} \left(\sum_{j=1}^{\infty} \varepsilon_j \Gamma_j^{-1/\alpha} g(V_j)^{-1/\alpha} \sum_{k=1}^{n} f_k(V_j) - n\delta \right) > u \right),$$

the intuition of rare events says that it is a single one of the Poisson points (in the function space) $\left(\varepsilon_j \Gamma_j^{-1/\alpha} g(V_j)^{-1/\alpha} \sum_{k=1}^{n} f_k(V_j), \, n = 1, 2, \ldots\right)$ that is most likely to cause the ruin. This intuition translates into

$$P(A) \sim \sum_{j=1}^{\infty} P\left(C_\alpha^{1/\alpha} \Gamma_j^{-1/\alpha} g(V_j)^{-1/\alpha} \sup_{n\geq 1} \left(\varepsilon_j \sum_{k=1}^{n} f_k(V_j) - n\delta \right) > u \right) \tag{5.25}$$

as $\lambda \to \infty$. It is the equivalence (5.25) that allows one to understand the change in the way the effect of these Poisson points is distributed over time as the parameter H crosses the boundary $1/\alpha$.

Interestingly, the probabilities of the rare events of the Example 3.5 $A_n = \{X_1 + \ldots + X_n > n(\mu + \delta)\}$ do not indicate anything interesting happening at the point $H = 1/\alpha$. In fact, since the processes under considerations are the increments of H-self-similar processes,

$$p_n = P\Big(X_1 + \ldots + X_n > \delta n\Big) = P(n^H X_1 > \delta n) \sim \text{ const}\, \delta^{-\alpha} n^{-\alpha(1-H)}$$

as $n \to \infty$. Hence the order of magnitude of p_n changes "ordinarily" as H crosses the boundary $1/\alpha$. As mentioned at the end of Subsection 5.2, one should, probably, look at certain related rare events as well. The behavior of the associated functionals in (5.1) does not seem to have been studied so far.

5.4. **High dimensional joint tails for a linear process with stable innovations.** We conclude this paper with a simulation study of a situation in which no analytical results are yet available. Consider a heavy tailed linear process (4.1). For a fixed $\lambda > 0$ we consider the probability of the event $A_n = \{X_j > \lambda,\, j = 0, \ldots, n\}$, when n is large. We are within the framework of Example 3.3. The discussion above makes it possible to conjecture that there is a phase transition at the boundary between the set Ξ_1 in (5.3) and its complement in the set Ξ in (5.2). To check this conjecture we ran a simulation of 10^7 realizations of a linear process with symmetric α-stable innovations with different α. We estimated both the probability $P(A_n)$ as a function of n and the rate of growth of the associated functional

$$R_n = \max\Big\{ j - i + 1 : 1 \le i \le j \le n,\ \min(X_i, \ldots, X_j) > \lambda \Big\}. \tag{5.26}$$

We simulated first an AR(1) process with $\varphi_j = 0$ for $j \ne 0$ or 1, $\varphi_0 = 1$ and varying φ_1. This choice of coefficients is, clearly, in Ξ_1^c. Then we simulated a linear process with $\varphi_j = 0$ for $j < 0$ and $\varphi_j = (1+j)^{-.8}$ for $j \ge 0$ (and $\alpha > 1/.8$). This choice of parameters is in the set Ξ_1.

While a simulation study of this type cannot provide a definite answer, it seems to indicate that for the AR(1) process the probabilities $P(A_n)$ decay exponentially fast with n. We plotted in Figure 5.1 the ratio $-\big(\log P(A_n)\big)/n$ over the range of n for λ in the set $\{.1, .2, .3, .4\}$ for the AR(1) process with $\alpha = 1.5$ and $\varphi_1 = .5$. Notice how the curves become horizontal.

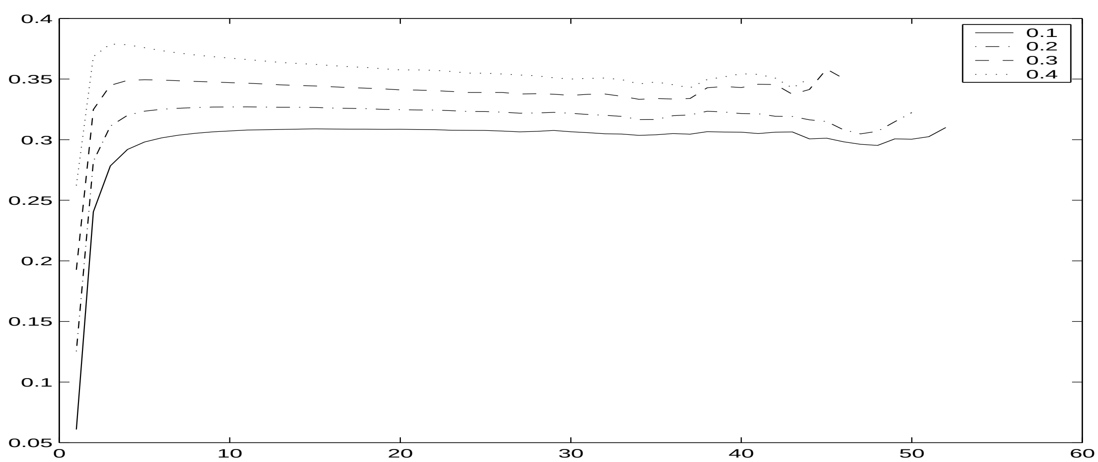

Figure 5.1. The ratio $-\big(\log P(A_n)\big)/n$ for the AR(1) process with $\alpha = 1.5$ and $\varphi_1 = .5$.

In comparison, our simulations seem to indicate that for the linear process with $\varphi_j = (1+j)^{-.8}$, $j \geq 0$ the probabilities $P(A_n)$ decay hyperbolically fast with n. We plotted in Figure 5.2 $P(A_n)$ against n in the log log scale, for the case $\alpha = 1.5$. Here we use λ in the set $\{.1,\ 1,\ 5,\ 40\}$. Notice how linear the plots are. Finally, we present a plot of $(\log R_n)/\log n$ for the long memory

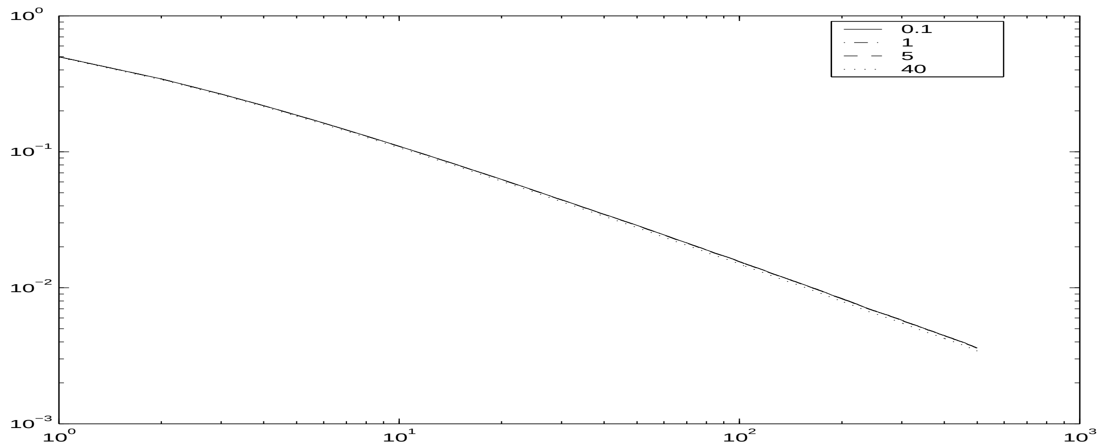

Figure 5.2. A plot of $P(A_n)$ against n for a linear process with $\alpha = 1.5$ and $\varphi_j = (1+j)^{-.8}$, $j \geq 0$. Log log scale.

process with $\alpha = 1.5$ and $\lambda \in \{.1, .2, .5,\ 1\}$. Our intuition tells us that in that case R_n should grow polynomially fast with n, and the simulation appears to bear this out.

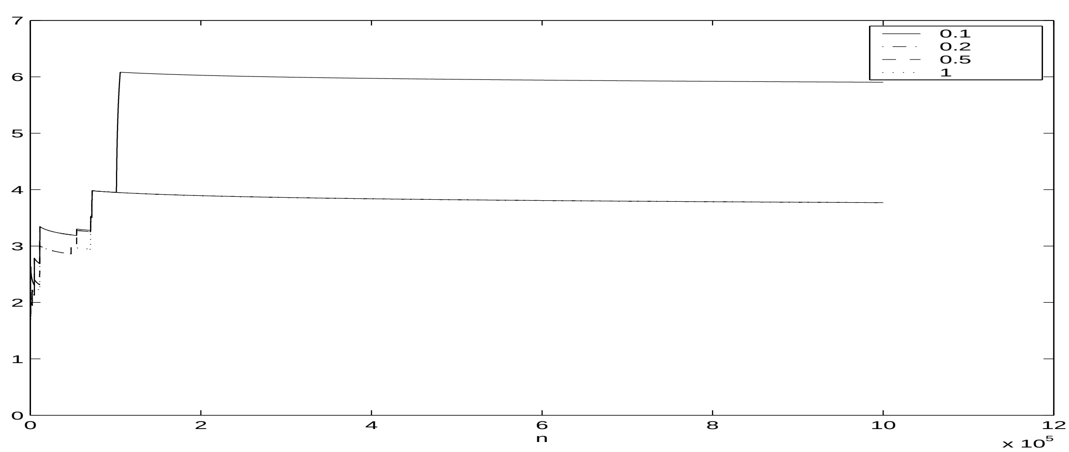

Figure 5.3. A plot of $(\log R_n)/\log n$ for a linear process with $\alpha = 1.5$ and $\varphi_j = (1+j)^{-.8}$, $j \geq 0$.

Once again, even though a simulation study is not a conclusive evidence of a phase transition at the boundary between the set Ξ_1 and its complement, its results are consistent with such a phase transition.

References

A. Araujo and E. Giné (1980): *The Central Limit Theorem for Real and Banach Valued Random Variables*. Wiley, New York.

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

J. Beran (1992): Statistical Methods for Data with Long-Range Dependence. *Statistical Science* 7:404–416. With discussions and rejoinder, pages 404-427.

J. Beran (1994): *Statistics for Long–Memory Processes*. Chapman and Hall, New York.

P. Brockwell and R. Davis (1991): *Time Series: Theory and Methods*. Springer-Verlag, New York, 2nd edition.

V. Chistyakov (1964): A theorem on sums of independent random variables and its applications to branching random processes. *Theory of Probability and its Applications* 9:640–648.

J. Chover, P. Ney and S. Wainger (1973): Functions of probability measures. *J. Analyse Math.* 26:255–302.

D. Cline (1983): *Estimation and linear prediction for regression, autoregression and ARMA with infinite variance data*. Ph.D. thesis, Colorado State University.

D. Cline (1985): Linear prediction of ARMA processes with infinite variance. *Stochastic Processes and their Applications* 19:281–296.

D. Cox (1984): Long-range dependence: a review. In *Statistics: An Appraisal*, H. David and H. David, editors. Iowa State University Press, pp. 55–74.

A. Dembo and O. Zeitouni (1993): *Large Deviations Techniques and Applications*. Jones and Bartlett Publishers, Boston.

J.-D. Deuschel and D. Stroock (1989): *Large Deviations*. Academic Press, Boston.

P. Embrechts, C. Klüppelberg and T. Mikosch (1997): *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin.

C. Goldie and C. Klüppelberg (1998): Subexponential distributions. In *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed Distributions*. Birkhäuser, Boston, pp. 435–460.

C. Gomez, B. Selman and N. Crato (1997): Heavy–tailed probability distributions in combinatorial search. In *Principles and Practice of Constraint Programming*, G. Smolka, editor. Springer, Berlin, pp. 121–135.

A. Gross (1994): Some mixing conditions for stationary symmetric stable stochastic processes. *Stoch. Proc. Appl.* 51:277–295.

C. Heyde (1968): On large deviation probabilities in the case of attraction to a non-normal stable law. *Sankyā Ser. A* 30:253–258.

H. Hurst (1951): Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* 116:770–808.

H. Hurst (1955): Methods of using long-term storage in reservoirs. *Proceedings of the Institution of Civil Engineers, Part I* :519–577.

W. Linde (1986): *Probability in Banach Spaces-Stable and Infinitely Divisible Distributions.* Wiley, Chichester.

B. Mandelbrot (1965): Une classe de processus stochastiques homothetiques a soi; application a loi climatologique de H.E. Hurst. *Comptes Rendus Acad. Sci. Paris* 240:3274–3277.

B. Mandelbrot (1983): *The Fractal Geometry of Nature.* W.H. Freeman and Co., San Francisco.

B. Mandelbrot and J. Van Ness (1968): Fractional Brownian motions, fractional noises and applications. *SIAM Review* 10:422–437.

B. Mandelbrot and J. Wallis (1968): Noah, Joseph and operational hydrology. *Water Resources Research* 4:909–918.

B. Mandelbrot and J. Wallis (1969): Robustness of the rescaled range R/S in the measurement of noncyclic long-run statistical dependence. *Water Resour. Res.* 5:967–988.

P. Mansfield, S. Rachev and G. Samorodnitsky (1999): Long strange segments of a stochastic process and long range dependence. Preprint. Avalilable as rareseg.ps.gz at www.orie.cornell.edu/gennady/techreports.

G. Maruyama (1970): Infinitely divisible processes. *Theory Prob. Appl.* 15:1–22.

T. Mikosch and G. Samorodnitsky (2000a): Ruin probability with claims modeled by a stationary ergodic stable process. *Annals of Probability* 28:1814–1851.

T. Mikosch and G. Samorodnitsky (2000b): The supremum of a negative drift random walk with dependent heavy–tailed steps. *Annals of Applied Probability* 10:1025–1064.

U. Müller, M. Dacorogna and O. Pictet (1998): Hill, bootstrap and jacknife estimators for heavy tails. In *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy*

Tailed Distributions. Birkhäuser, Boston, pp. 283–310.

S. Nagaev (1979): Large deviations of sums of independent random variables. *Annals of Probability* 7:745–789.

K. Park and W. Willinger, editors (2000): *Self–similar Network Traffic and Performance Evaluation*, New York. John Wiley and Sons, Inc.

S. Rachev and G. Samorodnitsky (2001): Long strange segments in a long range dependent moving average. *Stochastic Processes and Their Applications* 93:119–148.

B. Rajput and J. Rosiński (1989): Spectral representations of infinitely divisible processes. *Probability Theory and Related Fields* 82:451–488.

S. Resnick (1987): *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York.

J. Rosiński (1989): On path properties of certain infinitely divisible processes. *Stoch. Proc. Appl.* 33:73–87.

J. Rosiński (1990): On series representation of infinitely divisible random vectors. *The Annals of Probability* 18:405–430.

J. Rosiński (1998): Structure of stationary stable processes. In *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed Distributions*, R. Adler, R. Feldman and M. Taqqu, editors. Birkhäuser, Boston, pp. 461–472.

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

Faculty of Economics and Business Administration, University of Sofia, and Bravo Risk Management Group, 1101-C Eugenia Place, Carpinteria, Santa Barbara CA 93013
E-mail address: borjana.racheva@bravo-group.com

School of Operations Research and Industrial Engineering, and Department of Statistical Science, Cornell University, Ithaca, NY 14853
E-mail address: gennady@orie.cornell.edu