LONG RANGE DEPENDENCE, HEAVY TAILS AND RARE EVENTS

GENNADY SAMORODNITSKY

Abstract. These notes were prepared for a Concentrated Advanced Course at the University of Copenhagen in the framework of the MaPhySto program. The notes present an attempt to propose a new approach to long range dependence, one that is not mainly based on correlations, or on (approximate) self-similarity. Rather, we advocate a point of view that regards the passage between short memory and long memory and a phase transition in the way certain rare events happen. We are especially interested in the heavy tailed case. These notes bring together ideas from large deviations, theory of heavy tailed processes, extreme value theory and relate those to the notion of memory. Main examples considered are those related to communication networks and risk theory.

Research partially supported by NSF grant DMS-0071073 at Cornell University.

Contents

1. Introduction

The phenomenon of long range dependence is widely believed to be both ubiquitous and important in data arising in a variety of different fields. Yet what is long range dependence? How does one measure it? Historically, long range dependence was viewed as a property of certain stochastic models with a finite variance, and then it was associated either with a particularly slow decay of correlation, or with a particular pole of the spectral density at the origin. In addition to obvious drawbacks of correlations that carry a limited amount of information away from the Gaussian case, this leaves one unable to define long memory for stochastic processes with infinite variance. Sometimes "correlation–like" notions have been used where correlations did not exist; see e.g. Astrauskas et al. (1991). Those, as expected, carry precious little information.

We propose to think about long range dependence in terms of the way rare events happen. This is particularly appropriate in the heavy tailed situations because most practitioners using heavy tailed models are interested precisely in certain rare events related to the tails. Specifically, in many cases one can split the parameter space of a stationary process into two parts, such that probabilities of certain rare events undergo a significant change at the boundary. Often in one part of the space the order of magnitude of the probabilities stays the same, and, once the boundary is crossed, the order of magnitude increases, and becomes dependent on the parameters. If one makes sure that this phenomenon is not related to a change in heaviness of the tails, then this *phase transition* may be called a passage between short and long memory. We provide many details and examples in the sequel.

These notes are organized as follows. We start with a discussion of heavy tails, which is by itself an often ambiguous notion. Section 3 introduces various points of view on long range dependence. Self-similar processes, an important example where the notion is most easily visible, are discussed in subsection 3.1. We also discuss spectral analysis of long range dependence and periodogram (subsection 3.2), as well as the classical topic of Hermite polynomials and their applications to studying the effect of pointwise transformations of Gaussian processes on the rate of correlations (subsection 3.3).

The most powerful approach to studying rare events is that of large deviations; it is introduced in section 4. Subsection 4.1 applies the ideas of large deviations to certain problems in communication networks.

All the ideas developed in the above sections are brought together in section 5, where we start developing the connections between memory and the way rare events happen. Two main classes of stochastic processes are considered: that of moving averages (subsection 5.1) and that of infinitely divisible processes (subsection 5.4). An important example of the latter is that of stationary stable processes in subsection 5.5. Two main test classes of rare events and associated functionals are considered: that of long strange intervals (subsection 5.2) and that of ruin probabilities in subsection 5.3.

2. Heavy Tails

In many ways the class of distributions with heavy tailed tails is that of *subexponential distributions.*

Definition 1. *A distribution F on $[0,\infty)$ is called subexponential if $\overline{F}(x) > 0$ for all $x \geq 0$ and*

$$\lim_{x\to\infty} \frac{\overline{F*F}(x)}{\overline{F}(x)} = 2.$$

Here for a distribution F we denote by $\overline{F} = 1 - F$ its tail. If X is a random variable with a subexponential distribution F, then we say that X is a subexponential random variable. Note that subexponentiality of X means that

$$\lim_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X > x)} = 2.$$

Here X_1 and X_2 are independent copies of X.

Notice also that

$$P(\max(X_1, X_2) > x)$$
$$= P(X_1 > x) + P(X_2 > x) - P(X_1 > x)P(X_2 > x)$$
$$= 2P(X > x) - (P(X > x))^2 \sim 2P(X > x)$$

as $x \to \infty$. That is, for a subexponential random variable

$$P(X_1 + X_2 > x) \sim P(\max(X_1, X_2) > x)$$

as $x \to \infty$. Since X is nonnegative, we always have $X_1 + X_2 \geq \max(X_1, X_2)$.

Therefore, for a subexponential random variable X the sum $X_1 + X_2$ is larger than a large value x when either X_1 or X_2 are larger than x. It is much less likely that both X_1 and X_2 are less than x, but they are still large enough so that their sum exceeds x. For example, for a subexponential random variable

$$P(x/2 < X_1 \leq x, x/2 < X_2 \leq x) \tag{2.1}$$
$$= (P(x/2 < X \leq x))^2 = o(P(X > x)).$$

Example 2. Let X be an exponential random variable with mean 1. Then

$$P(x/2 < X \leq x)^2 = (e^{-x/2} - e^{-x})^2$$
$$\sim e^{-x} = P(X > x)$$

as $x \to \infty$, and so (2.1) fails. That is, an exponential random variable is not subexponential.

This last fact also follows from second statement in the following proposition, that describe some basic properties of subexponential random variables. The proof may be found in, say, Embrechts et al. (1979).

Proposition 3. *Let X be a subexponential random variable, and $X_1, X_2, \ldots$ are iid copies of X. Then*

(i) $\lim_{x\to\infty} \frac{P(X>x+y)}{P(X>x)} = 1$ *uniformly in y over compact sets;*

(ii) $\lim_{x\to\infty} e^{\epsilon x} P(X > x) = \infty$, *for each $\epsilon > 0$;*

(iii) For every $n \geq 1$

$$\lim_{x\to\infty} \frac{P(X_1 + \ldots + X_n > x)}{P(X > x)} = n.$$

(iv) If N is a Poisson random variable with mean λ that is independent of the sequence $X_1, X_2, \ldots$ then

$$\lim_{x\to\infty} \frac{P(X_1 + \ldots + X_N > x)}{P(X > x)} = \lambda.$$

(v) If $\lim_{x\to\infty} P(Y > x)/P(X > x) = c \in (0, \infty)$, *then Y is subexponential as well.*

Even though we have defined subexponentiality for nonnegative random variables, the notion immediately extends to the class of general real valued random variables. We will call a real random variable X subexponential if its positive part, X_+, is subexponential. The properties of subexponentiality discussed above extend immediately to this more general case. For example, if X is subexponential, and X_1 and X_2 are independent copies of X, then

$$\limsup_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X > x)} \leq \limsup_{x\to\infty} \frac{P((X_1)_+ + (X_2)_+ > x)}{P(X_+ > x)} = 2,$$

and for every $M > 0$

$$P(X_1 + X_2 > x) \geq P(X_1 > x + M, X_2 > -M \text{or } X_1 > -M, X_2 > x + M)$$

$$= 2P(X_1 > x + M, X_2 > -M) - (P(X_1 > x + M))^2,$$

hence

$$\liminf_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X > x)} \geq \liminf_{x\to\infty} 2\frac{P(X_1 > x + M, X_2 > -M)}{P(X > x)}$$

$$= 2P(X > -M)$$

by part (i) of Proposition 3 above. Letting $M \to \infty$ we conclude that

$$\lim_{x\to\infty} \frac{P(X_1 + X_2 > x)}{P(X > x)} = 2.$$

Similarly, the parts (iii) and (i) of Proposition 3 extend to the general, not necessarily nonnegative, case as well.

A nice tool for checking whether a given distribution on $[0, \infty)$ is subexponential is the following result due to Pitman (1980). For a distribution F on $[0, \infty)$ let

$$g_F(x) = -\log \overline{F}(x), \ x \geq 0.$$

Theorem 4. *Suppose g_F has an eventually decreasing to 0 derivative g_F'. The a necessary and sufficient condition for F to be subexponential is*

$$\lim_{x\to\infty}\int_0^x \exp\{yg_F'(x)-g_F(y)\}g_F'(y)dy=1, \tag{2.2}$$

and a sufficient condition is

$$\exp\{yg_F'(y)-g_F(y)\}g_F'(y) \tag{2.3}$$

integrable over $[0,\infty)$.

Example 5. Let Y be the standard log-normal random variable. Then

$$P(Y>x)\sim\frac{1}{\sqrt{2\pi}}(\log x)^{-1}\exp\{-\frac{(\log x)^2}{2}\}$$

as $x\to\infty$. Choose a random variable X with a distribution F such that

$$P(Y>x)=\frac{1}{\sqrt{2\pi}}(\log x)^{-1}\exp\{-\frac{(\log x)^2}{2}\}$$

for $x\geq x_0>1$, and such that g_F has a bounded derivative on $[0,x_0]$. Notice that on $[x_0,\infty)$

$$g_F(x)=c+\log\log x+\frac{1}{2}(\log x)^2,$$

some constant c that may change as we go along, hence

$$g_F'(x)=\frac{1}{x\log x}+\frac{\log x}{x},$$

which is an eventually decreasing to zero function.

Hence

$$\begin{aligned}&\exp\{xg_F'(x)-g_F(x)\}g_F'(x)\\&=\exp\{\log x+(\log x)^{-1}-\log\log x-\frac{1}{2}(\log x)^2-c\}g_F'(x)\\&\leq\exp\{\log x+(\log x)^{-1}-\log\log x-\frac{1}{2}(\log x)^2-c\}\\&\leq\exp\{-c(\log x)^2\},\end{aligned}$$

some $c>0$, which is an integrable function. Hence by Theorem 4 X is subexponential and, hence, so is the log-normal random variable Y.

Distribution functions with the following tails can be similarly shown to be subexponential.

(1) $\overline{F}(x)\sim\exp\{-cx^a\}$ as $x\to\infty$, $0<a<1$ and $c>0$.
(2) $\overline{F}(x)\sim\exp\{-cx(\log x)^{-a}\}$ as $x\to\infty$, $a>0$ and $c>0$.

Arguably, the single most important class of subexponential distributions is given in the following example.

Example 6. A function $g : I\!R \to I\!R$ is called regularly varying at infinity with exponent $a \in I\!R$ if for every $c > 0$

$$\lim_{x\to\infty} \frac{g(cx)}{g(x)} = c^a. \tag{2.4}$$

If (2.4) holds with $a = 0$, then the function g is called slowly varying (at infinity). Any function g that is regularly varying with exponent a can, obviously, be written in the form

$$g(x) = x^a L(x), \; x > 0, \tag{2.5}$$

where L is a slowly varying function.

Definition 7. *A distribution function F is said to have regularly varying right tail with tail exponent $\alpha > 0$ if $\overline{F}$ is regularly varying at infinity with exponent $-\alpha$. If $\overline{F}$ is slowly varying at infinity, then we say that F has a slowly varying right tail.*

It is easy to see directly that a distribution with a regularly or slowly varying right tail is subexponential. For instance, if X is a nonnegative random variable with such a distribution, then for any $0 < \epsilon < 1$ we have

$$\begin{aligned} P(X_1 + X_2 > x) &\le P(X_1 > (1-\epsilon)x) + P(X_2 > (1-\epsilon)x) + P(X_1 > \epsilon x, X_2 > \epsilon x) \\ &= 2P(X > (1-\epsilon)x) + (P(X > \epsilon x))^2, \end{aligned}$$

and, hence,

$$\begin{aligned} \limsup_{x\to\infty} \frac{P(X_1+X_2>x)}{P(X>x)} &\le 2\limsup_{x\to\infty} \frac{P(X>(1-\epsilon)x)}{P(X>x)} + \limsup_{x\to\infty} \frac{(P(X>\epsilon x))^2}{(P(X>x))^2} P(X>x) \\ &= 2(1-\epsilon)^{-\alpha} + \epsilon^{-2\alpha}\cdot 0 = 2(1-\epsilon)^{-\alpha}. \end{aligned}$$

Letting $\epsilon \to 0$ we obtain

$$\limsup_{x\to\infty} \frac{P(X_1+X_2>x)}{P(X>x)} \le 2,$$

which means that X is subexponential.

There are several basic facts about regularly varying functions we will be using. Let g be a regularly varying at infinity function with exponent $a \in I\!R$.

(1) The convergence in

$$\lim_{x\to\infty} \frac{g(cx)}{g(x)} = c^a$$

takes place uniformly in c over compact intervals in $(0,\infty)$. If $a < 0$ then this convergence is also uniform in c over half-lines (b,∞) with $b > 0$.

(2) Let $\epsilon \in (0,1)$ and $b > 1$. There is a $x_0 > 0$ such that for all $c \ge b$ and $x \ge x_0$

$$(1-\epsilon)c^{a-\epsilon} \le \frac{g(cx)}{g(x)} \le (1+\epsilon)c^{a+\epsilon} \tag{2.6}$$

(the Potter bounds).

See e.g. Resnick (1987).

Distributions with regularly varying tails are by far the most popular in applications among all subexponential distributions. At the very least, many arguments involving distributions with regularly varying tails are much easier than the corresponding arguments involving subexponential distributions, and sometimes the parallel statements in the general subexponential case are simply false.

For example, it follows directly from the definition that if X and Y are independent, and have distributions with regularly varying right tails with exponents α and β accordingly, then $X+Y$ has also a distribution with a regularly varying right tail with exponent $\min(\alpha,\beta)$.

On the other hand, *it is not true* that if X and Y are independent and subexponential distributions, then $X+Y$ also has a subexponential distribution. A counterexample is given by Leslie (1989).

The situation is even more delicate with products instead of sums.

Let X be a random variable with a regularly varying right tail with exponent α, and Y is a positive random variable independent of X such that for some $\epsilon>0$ $EY^{\alpha+\epsilon}<\infty$. Let $Z=XY$. We claim that

$$\lim_{x\to\infty}\frac{P(Z>x)}{P(X>x)}=EY^{\alpha}. \tag{2.7}$$

In particular, Z has also a regularly varying right tail with exponent α.

To see that, let G be the distribution of Y. By the Potter bounds, there is a $C>0$ such that for all x large enough

$$\frac{P(X>x/y)}{P(X>x)}\le Cy^{\alpha+\epsilon}$$

for all $y>0$, where $\epsilon>0$ is such that $EY^{\alpha+\epsilon}<\infty$. This allows us to use the dominated convergence theorem in

$$\lim_{x\to\infty}\frac{P(Z>x)}{P(X>x)}=\lim_{x\to\infty}\int_0^\infty\frac{P(X>x/y)}{P(X>x)}G(dy)$$

$$=\int_0^\infty\lim_{x\to\infty}\frac{P(X>x/y)}{P(X>x)}G(dy)=\int_0^\infty y^{\alpha}G(dy)=EY^{\alpha}.$$

In fact, if X be a random variable with a regularly varying right tail with exponent α, and Y is a positive random variable independent of X, then it is also true that $Z=XY$ has also a regularly varying right tail with exponent α under the following assumptions:

(1) $EY^{\alpha}<\infty$.

(2) Y has a regularly varying right tail with exponent α.

However, the relation (2.7)

$$\lim_{x\to\infty}\frac{P(Z>x)}{P(X>x)}=EY^{\alpha}$$

may fail. See Embrechts and Goldie (1980).

Results of this type are much harder to obtain in the case when X is a general subexponential random variable. For example, it is not known, in general, whether or not it is true that if X and Y are independent subexponential random variables, then $Z = XY$ is also a subexponential random variable (probably not). However, it is still, roughly speaking, the case that if Y is a positive random variable independent of X, and the right tail of Y is sufficiently light in comparison with with the right tail of X.

The following result is from Cline and Samorodnitsky (1994). Let F and G be the distribution functions of X and Y correspondingly.

Theorem 8. *Assume that X is a subexponential random variable. If there is a function $a : (0, \infty) \to (0, \infty)$ satisfying*

- $a(t) \uparrow \infty \qquad$ *as* $t \to \infty$,
- $t/a(t) \uparrow \infty \qquad$ *as* $t \to \infty$,
- $\lim_{t\to\infty} \overline{F}(t - a(t))/\overline{F}(t) = 1$,
- $\overline{G}(a(bt)) = o(\overline{F}(t))$ *for some* $b > 0$

then $Z = XY$ is also a subexponential random variable.

In particular, if X is a subexponential random variable, and Y is a bounded positive random variable independent of X, then $Z = XY$ is also a subexponential random variable.

There are various statistical approaches to detecting heavy tails and measuring heaviness of the tails. Instead of looking at formal statistical tests, let us mention at the moment an "eyeball" approach to detecting heavy tails.

By plotting the data and seeing that most of the data is dominated by a few largest observations (see Figure 9) one concludes that heavy tails are present.

A typical plot without heavy tails is that of Figure 10. Here there are no obvious dominating observations.

We finish this introduction into heavy tails by observing that when random variables with heavy tails are mentioned in literature, the authors often mean different things. Some common possibilities:

- random variables with subexponential tails
- random variables with regularly varying right tails
- random variables with regularly varying right tails with exponent $\alpha < 1$
- random variables with infinite second moment

It is, therefore, important, to ascertain in what sense the notion of heavy tails is used in any given instance.

3. Long range dependence

As we mentioned above, there is no consensus on the notion of heavy tails. There is even less consensus on the notion of 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.

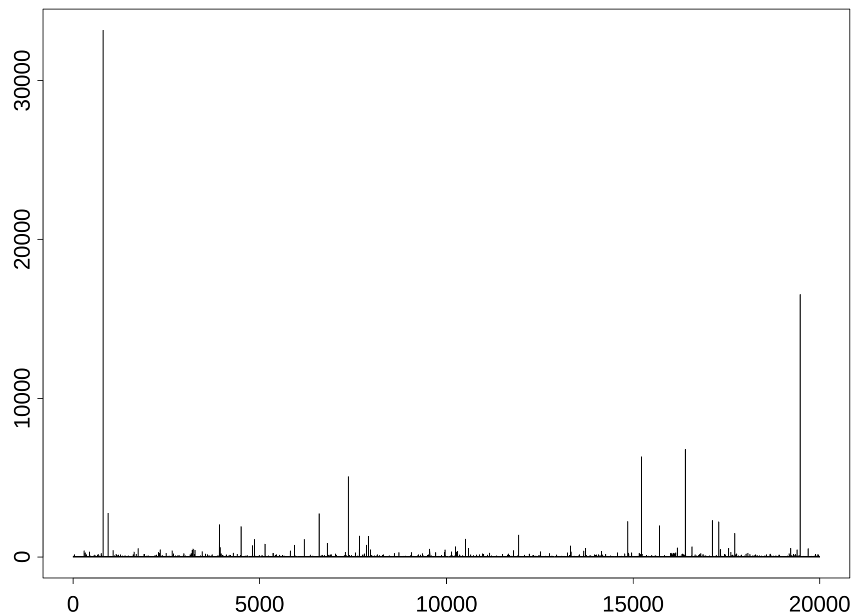

Figure 9. iid Pareto random variables with $\alpha = 1$.

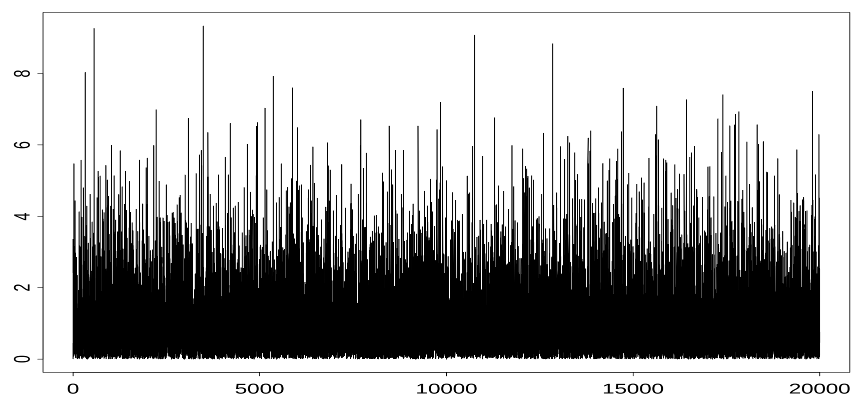

Figure 10. iid exponential random variables with mean 1.

Let us start with the discrete time case. Let X_n, $n = 0, 1, 2, \ldots$ be a stationary stochastic process with mean $\mu = EX_0$ and $0 < \sigma^2 = \mathrm{Var}X_0 < \infty$. Let $\rho_n = \mathrm{Corr}(X_0, X_n)$, $n = 0, 1, \ldots$ 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 implies, in particular, that

$$\sum_{n=0}^{\infty} |\rho_n| < \infty. \tag{3.1}$$

Summability of correlations means, in turn, many other things. For example, consider the partial sum process

$$S_n = X_1 + \ldots + X_n,\ n \geq 1,\ S_0 = 0,$$

and let us calculate its variance. We have

$$\text{Var}S_n = \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}(X_i, X_j)$$

$$= \sigma^2 \sum_{i=1}^{n}\sum_{j=1}^{n} \rho_{|i-j|} = \sigma^2 (n + 2\sum_{i=1}^{n-1}(n-i)\rho_i).$$

Under the assumption (3.1) we see by the dominated convergence theorem that

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

That is, if the correlations are summable, then the variance of the partial sum process increases linearly fast. Often it is the case that S_n itself has the order of magnitude $n^{1/2}$.

Moreover, under certain additional regularity assumptions the partial sum process (S_n) satisfies *Functional Central Limit Theorem*:

$$\frac{1}{\sqrt{n}} S^{(n)} \Rightarrow \sigma_* B \text{ in } D[0,1], \tag{3.2}$$

where

$$S^{(n)}(t) = S_{[nt]} - [nt]\mu,\ 0 \leq t \leq 1, \tag{3.3}$$

$$\sigma_*^2 = \sigma^2 (1 + 2\sum_{i=1}^{\infty} \rho_i) \geq 0$$

and B is the standard Brownian motion on $[0,1]$.

In 1951 a British hydrologist, H. Hurst, published a study of flow of water in the Nile river. The plot on Figure 11 shows the annual minima of the water level in the Nile river for the years 622-1281, measured at the Roda gauge near Cairo.

This plot looks interesting in many different ways, but H. Hurst was interested in a specific statistics, defined as follows. Let

$$\frac{R}{S}(X_1, \ldots, X_n) = \frac{\max_{0\leq i\leq n}(S_i - \frac{i}{n}S_n) - \min_{0\leq i\leq n}(S_i - \frac{i}{n}S_n)}{(\frac{1}{n}\sum_{i=1}^{n}(X_i - \frac{1}{n}S_n)^2)^{1/2}}. \tag{3.4}$$

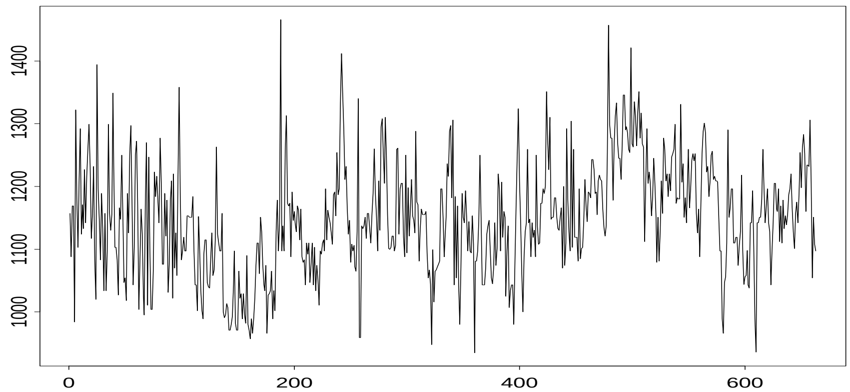

Figure 11. The Nile river annual minima

The numerator in this statistic is referred to as the range of the data, and the whole statistic (after scaling by the sample standard deviation) is referred to as the rescaled range, or *the R/S statistic.*

If one computes the R/S statistic on the increasing subsets $\{X_1, \ldots, X_n\}$ of the Nile river data set $\{X_1, \ldots, X_{660}\}$ then it turns out that, as a function of the sample size n, this statistic grows approximately as $n^{.75}$.

This has been observed to be strange. Too see why it is strange, suppose that our observations come from a stationary model with a finite variance and summable correlations, as in (3.1). In fact, assume that the Functional Central Limit Theorem (3.2) holds.

Notice that the range of the data, which is the numerator of the R/S statistic, can be expressed in the form $f(S^{(n)})$, where $S^{(n)}$ is the partial sum process in (3.3), and $f : D[0,1] \to I\!R$ is given by

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

$\mathbf{x} = (x(t),\ 0 \le t \le 1) \in D[0,1]$. It is easy to see that this is a continuous function on $D[0,1]$. Therefore, by the continuous mapping theorem,

$$\frac{1}{\sqrt{n}}(\textit{the range of the data}) = f\left(\frac{1}{\sqrt{n}} S^{(n)}\right) \Rightarrow f(\sigma_* B) \tag{3.5}$$

$$= \sigma_* \Big[\sup_{0 \le t \le 1} (B(t) - tB(1)) - \inf_{0 \le t \le 1} (B(t) - tB(1)) \Big] = \sigma_* \big[\sup_{0 \le t \le 1} B_0(t) - \inf_{0 \le t \le 1} B_0(t) \big],$$

where B_0 is the Brownian bridge on $[0,1]$.

Furthermore, if the observations come from an ergodic stationary process then the sample standard deviation is a consistent estimator of the population standard deviation;

$$(\frac{1}{n}\sum_{i=1}^{n}(X_i - \frac{1}{n}S_n)^2)^{1/2} \to \sigma. \tag{3.6}$$

Hence, if the observations come from a stationary ergodic model with a finite variance and summable correlations, and if the Functional Central Limit Theorem (3.2) holds, then by (3.5) and (3.6) we conclude that

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

That is, *the R/S statistic grows, with the sample size, as $n^{.5}$*. This was observed first by Feller (1952) (in the iid case).

For a long time people tried to explain what classes of stationary processes would lead to a faster rate of increase of the the R/S statistic than $n^{.5}$. This latter phenomenon has become known as *the Hurst phenomenon.*

Moran (1964) claimed that one can explain the Hurst phenomenon by assuming that the observations $X_1, X_2, \ldots$, while independent (and identically distributed), possess heavy tails in the sense of the infinite second moment. Specifically, he assumed that the observations have regularly varying tails with exponent $0 < \alpha < 2$, and even more specifically he assumed that these observations are in the domain of attraction of an α-stable distribution.

However, in Mandelbrot and Taqqu (1979) a heuristic argument was given that showed that even in that case one would expect the the R/S statistic to grow at the rate $n^{.5}$. It is not difficult to make this argument precise.

In fact, one can show that here (leaving aside slowly varying terms) that

$$\textit{the range of the data} \sim n^{1/\alpha}$$

and

$$\sum_{i=1}^{n}(X_i - \frac{1}{n}S_n)^2 \sim n^{2/\alpha}$$

(in distributional sense), so that the R/S statistic still has the order of $n^{.5}$.

An possible explanation of the Hurst phenomenon had to wait until the paper of Mandelbrot (1975), and it had to do with the length of memory as opposed to heavy tails.

3.1. **Self-similar processes.**

Definition 12. A stochastic process $(Y(t), t \ge 0)$ is called self-similar with exponent $H \in \mathbb{R}$ of self-similarity if for all $c > 0$

$$(Y(ct), t \ge 0) \stackrel{d}{=} c^H(Y(t), t \ge 0) \tag{3.7}$$

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

Obviously, if $H \neq 0$ then $Y(0) = 0$ a.s.

If the process $(Y(t), t \geq 0)$ also has stationary increments, then this process is often denoted SSSI (self-similar stationary increments).

A few facts about SSSI processes. Let $(Y(t), t \geq 0)$ be an SSSI process with exponent H of self-similarity.

(1) If $H < 0$ then $Y(t) = 0$ a.s. for every $t \geq 0$.
(2) If $H = 0$ and $(Y(t), t \geq 0)$ has a measurable modification (in particular, if it is continuous in probability), then for all $t \geq 0$, $P(X(t) = X(0)) = 1$.

Therefore, one assumes that $H > 0$ when studying SSSI processes.

Certain moment assumptions will further reduce the feasible range of the self-similarity exponent H. We will assume in the sequel that our process is not identically equal to zero at any given time t.

Lemma 13. *Let $(Y(t), t \geq 0)$ be an SSSI process with exponent H of self-similarity, and assume that $E|Y(1)| < \infty$. Then $H \leq 1$.*

Proof. We have by the self-similarity

$$n^{H-1}Y(1) \stackrel{\mathrm{d}}{=} \frac{Y(n)}{n} =$$
$$\frac{Y(1) + (Y(2) - Y(1)) + \ldots + (Y(n) - Y(n-1))}{n}$$

and by the stationarity of the increments and ergodic theorem the expression in the right hand side above converges a.s. to the conditional expectation of $Y(1)$ given the corresponding invariant σ-field. In particular, the family of the laws of the random variables in the right side above, indexed by $n \geq 1$, is tight. Hence $H \leq 1$. □

Two more facts are presented without a proof. Let $(Y(t), t \geq 0)$ be an SSSI process with exponent H of self-similarity.

(1) Assume $E|Y(1)| < \infty$. If $H = 1$ then for every $t \geq 0$ we have $X(t) = tX(1)$ with probability 1.
(2) Assume $E|Y(1)| < \infty$. If $0 < H < 1$ then $EY(1) = 0$.
(3) Assume that for some $0 < \gamma < 1$, $E|Y(1)|^\gamma < \infty$. Then $0 < H < 1/\gamma$.

Suppose that $(Y(t), t \geq 0)$ is an SSSI process with exponent H of self-similarity, and a finite non-zero variance at time 1. By the above we know that, if we want to avoid degenerate situations, we have to assume that $0 < H < 1$, and then the process has to have zero mean. For every $0 \leq s < t$ we must have then

$$E(Y(t) - Y(s))^2 = EY(t-s)^2 = (t-s)^{2H}EY(1)^2,$$

and so

$$\operatorname{Cov}(Y(s), Y(t)) = \frac{1}{2}\Big[EY(t)^2 + EY(s)^2 - E(Y(t) - Y(s))^2\Big] \tag{3.8}$$
$$= \frac{EY(1)^2}{2}\Big[t^{2H} + s^{2H} - (t-s)^{2H}\Big].$$

Therefore, if $(Y(t), t \geq 0)$ is a second order SSSI process with $0 < H < 1$, then it must have the covariance structure specified by (3.8).

Of course, we have not checked that the right hand side of (3.8) is a non-negative definite function. This, however, turns out to be the case. Therefore, second order SSSI processes do exist, and they all share the same covariance function given by the right hand side of (3.8).

In particular, **there is a unique up to a scale zero mean SSSI Gaussian process with** $0 < H < 1$. This process is called *Fractional Brownian motion* (FBM).

Since for $H = 1/2$ the covariance function given by the right hand side of (3.8) reduces to $\mathrm{Cov}(Y(s), Y(t)) = EY(1)^2 s$, one sees that a FBM with $H = 1/2$ is just a Brownian motion.

Going back to the Hurst phenomenon, let $(Y(t), t \geq 0)$ be a FBM, and let $X_i = Y(i) - Y(i-1)$, $i = 1, 2, \ldots$. By the stationarity of the increments of FBM, this is a stationary sequence, commonly referred to as the *Fractional Gaussian Noise* (FGN). Let us understand the behavior of the R/S statistic on FGN.

Notice that by the self-similarity

$$\max_{0 \leq i \leq n} (S_i - \frac{i}{n} S_n) = \max_{0 \leq i \leq n} (Y(i) - \frac{i}{n} Y(n)) \stackrel{\mathrm{d}}{=} n^H \max_{0 \leq i \leq n} (Y(\frac{i}{n}) - \frac{i}{n} Y(1)).$$

Since a FBM has continuous sample paths, we see that

$$\max_{0 \leq i \leq n} (Y(\frac{i}{n}) - \frac{i}{n} Y(1)) \to \sup_{0 \leq t \leq 1} (Y(t) - tY(1))$$

with probability 1.

The same argument applies to the second part of the range and, hence,

$$n^{-H} \Big[\max_{0 \leq i \leq n} (S_i - \frac{i}{n} S_n) - \min_{0 \leq i \leq n} (S_i - \frac{i}{n} S_n) \Big]$$
$$\Rightarrow \sup_{0 \leq t \leq 1} (Y(t) - tY(1)) - \inf_{0 \leq t \leq 1} (Y(t) - tY(1)).$$

Since the relation (3.6) still holds (the sample standard deviation converges a.s. to the population standard deviation) we conclude that

$$n^{-H} \frac{R}{S}(X_1, \ldots, X_n) \Rightarrow \frac{1}{\sigma} \Big[\sup_{0 \leq t \leq 1} (Y(t) - tY(1)) - \inf_{0 \leq t \leq 1} (Y(t) - tY(1)) \Big].$$

Therefore, a choice of an appropriate $H > 1/2$ will provide an explanation of the Hurst effect.

Let us do some covariance computations for a FGN. For $1 \leq i < j$ we have by (3.8)

$$\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}\Big[(Y(i) - Y(i-1))(Y(j) - Y(j-1)) \Big]$$
$$= \frac{EY(1)^2}{2} \Big[(j-i+1)^{2H} + (j-i-1)^{2H} - 2(j-i)^{2H} \Big]$$

and

$$\mathrm{Var}(X_i) = \mathrm{Var}(Y(i) - Y(i-1)) = EY(1)^2.$$

Hence, for $n \geq 1$

$$\rho_n = \mathrm{Corr}(X_0, X_n) = \frac{(n+1)^{2H} + (n-1)^{2H} - 2n^{2H}}{2}. \tag{3.9}$$

In particular,

$$\rho_n \sim H(2H-1)n^{-2(1-H)} \tag{3.10}$$

as $n \to \infty$. Of course, if $H = 1/2$, then $\rho_n = 0$ for all $n \geq 1$ (a Brownian motion has independent increments).

We conclude that the summability of correlations (3.1) holds if $0 < H \leq 1/2$ and it does not hold if $1/2 < H < 1$.

Therefore, a FGN with $H > 1/2$ has become commonly accepted as having long range dependence, and lack of summability of correlations as a popular definition of long range dependence.

Nonetheless, even those who view long range dependence through slow decay of correlations do not agree on what the right definition is. Here are several definitions that have been used.

(1) **Lack of summability of correlations**

$$\sum_{n=0}^{\infty} |\rho_n| = \infty. \tag{3.11}$$

(2) **Correlations are regularly varying at infinity with exponent** $-1 < d \leq 0$. Of course, this assumption implies lack of summability of correlations (3.11). FGN with $H > 1/2$ has this property.

(3) **Correlations are regularly varying at infinity with exponent** $d \leq 0$. This assumption does not imply (3.11), and it is designed, rather, for contrast with the case of exponentially decaying correlations.

Another possible angle of viewing long range dependence that is still closely related to correlations is through the spectral domain. Let X_n, $n = 0, 1, 2, \ldots$ be a stationary stochastic process with a finite variance σ^2. If its correlations are summable (i.e. (3.1) holds) then the process has a spectral density f satisfying

$$\sigma^2 \rho_n = \int_0^{\pi} \cos(nx) f(x) dx, \tag{3.12}$$

$n = 0, 1, 2, \ldots$. Moreover, in this case the spectral density is continuous on $[0, \pi]$.

On the other hand, it has been observed that a particular slow decay of correlations (condition 2 above), namely that the correlations are regularly varying at infinity with exponent $0 \leq d < 1$ often goes together with the spectral density "blowing up" at the origin. In fact, it has been observed that often the spectral density is then regularly varying at the origin, with exponent $-(1-d)$.

Example 14. It is easy to check that Fractional Gaussian Noise (FGN) $(Y_1, Y_2, \ldots)$ has a spectral density given by

$$f(x) = EY_1^2\, C(H)(1-\cos x) \sum_{j=-\infty}^{\infty} |2\pi j + x|^{-(1+2H)} \sim \frac{1}{2} EY_1^2 C(H) x^{-(2H-1)} \tag{3.13}$$

as $x \to 0$. Here

$$C(H) = \frac{2H(1-2H)}{\Gamma(2-2H)} \frac{1}{\cos \pi H}$$

($= 2/\pi$ if $H = 1/2$.) In particular, the spectral density is continuous at the origin if $0 < H \leq 1/2$ and "blows up" at the origin at the appropriate rate if $1/2 < H < 1$.

In fact, the equivalence between the regular variation of the correlations at infinity and the regular variation of the spectral density at the origin have become taken for granted, and often stated as a theorem (without proof).

To the best of our knowledge this equivalence is false in general, without extra regularity assumptions. Below is a rigorous result. Let $R_n = EY_1^2 \rho_n$, $n = 0, 1, 2, \ldots$ be the covariance function of a (weakly) stationary second order process.

Theorem 15. *(i) Assume that*

$$R_n = n^{-d} L(n),\ n = 0, 1, 2, \ldots, \tag{3.14}$$

where $0 < d < 1$ *and* L *is slowly varying at infinity, satisfying the following assumption:*

$$\text{for every } \delta > 0 \text{ both functions } g_1(x) = x^{\delta} L(x) \text{ and } g_2(x) = x^{-\delta} L(x) \text{ are eventually monotone.} \tag{3.15}$$

Then the process has a spectral density, say, f*, satisfying*

$$f(x) \sim x^{-(1-d)} L(x^{-1}) \frac{2}{\pi} \Gamma(1-d) \sin \frac{1}{2} \pi d \tag{3.16}$$

as $x \to 0$.

(ii) Conversely, assume that the process has a spectral density f *satisfying*

$$f(x) = x^{-d} L(x^{-1}),\ 0 < x < \pi, \tag{3.17}$$

where $0 < d < 1$*, and* L *is slowly varying at infinity, satisfying assumption (3.15) above. Suppose, further, that* f *is of bounded variation on the interval* (ϵ, π) *for any* $0 < \epsilon < \pi$*. Then the covariances of the process satisfy*

$$R_n \sim n^{-(1-d)} L(n) \Gamma(1-d) \sin \frac{1}{2} \pi d \tag{3.18}$$

as $n \to \infty$.

This theorem explains why is that that in practically all cases people looked at both appropriately slow regular decay of correlations at infinity and the appropriate "explosion" of the spectral density at the origin occur at the same time. It is easy to see, for example, that a slowly varying function

$$L(x) = (\log x)^\theta, \text{ any } \theta \in \mathbb{R}$$

satisfies the regularity assumption (3.15). This observation is a particular case of the following more general and easily verifiable statement.

Proposition 16. *Any eventually absolutely continuous, monotone, nonnegative function L such that L' is regularly varying at infinity with exponent -1 satisfies the regularity assumption (3.15).*

Nonetheless, **there are examples of slowly varying functions that do not satisfy the assumption** (3.15).

Whether or not there is equivalence between the rate of decay of correlations and the rate of "exposition" of the spectral density at the origin, two additional alternative definitions of long range dependence appeared.

(1) Spectral density is regularly varying at the origin with exponent $0 < d \le 1$.
(2) Spectral density has an infinite limit at the origin.

This last point of view on long range dependence has become one of the most commonly used ways by which people try to detect presence of long range dependence in the data.

3.2. **Periodogram.** For a sequence of observations $X_1, \ldots, X_n$ define

$$I_n(\lambda) = \frac{1}{\pi n} \left| \sum_{j=1}^{n} (X_j - \overline{X}) e^{ij\lambda} \right|^2, \tag{3.19}$$

to be the periodogram of the sequence computed at the frequency $\lambda \in (0, \pi)$. Here $\overline{X} = \sum_{j=1}^{n} X_j/n$ is the sample mean.

If $X_1, X_2, \ldots$ are observations from a second order stationary process with a spectral density f and either summable correlations, or certain kind of non-summable correlations, then it is not difficult to check that

$$EI_n(\lambda) \to f(\lambda) \text{ as } n \to \infty. \tag{3.20}$$

Moreover, under certain assumptions, one also knows the asymptotic distribution of the periodogram.

Hence, one can calculate the periodogram, take it as the estimate of the spectral density, and check if the estimated spectral density seems to have a pole at the origin. Furthermore, one may even try to estimate exactly how fast the estimated spectral density "blows up" at the origin.

Figure 17 contains the plot of the periodogram for the Nile river data. Note the obvious pole at the origin and lack of any structure away from the origin.

We have talked in details about points of view on long range dependence in the case of a finite variance. One way or the other these definitions of long range dependence

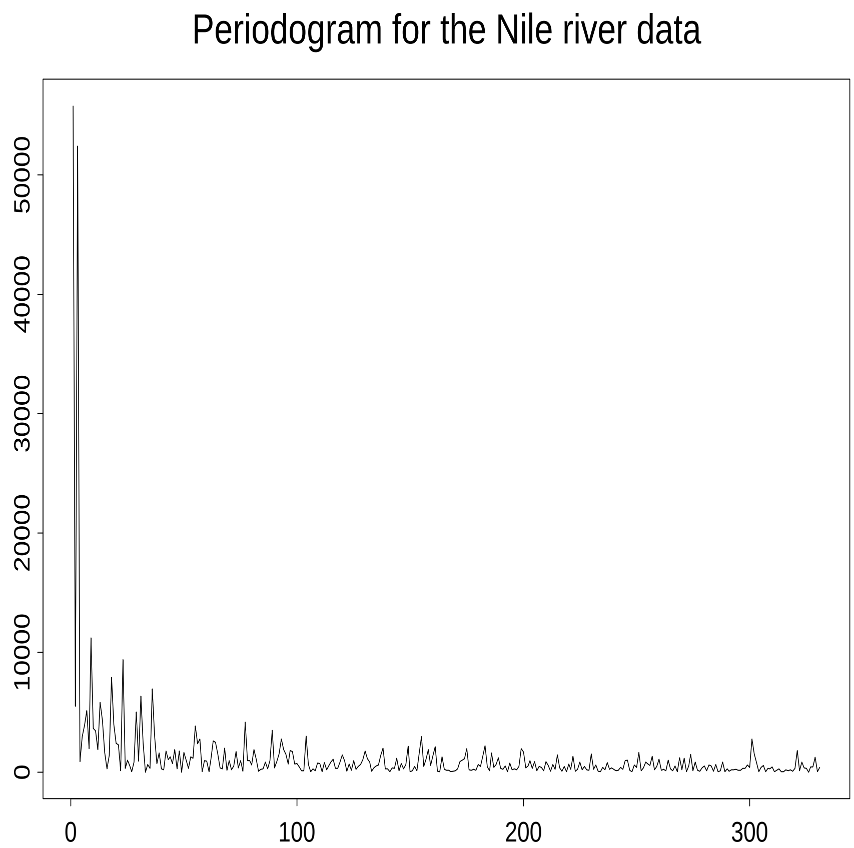

Figure 17.

concentrate on the correlations of the process. However, such approaches have drawbacks, that we will presently discuss.

Suppose that $X_1, X_2, \ldots$ is a zero mean stationary Gaussian process with covariance function satisfying

$$R_n \sim a\, n^{-d} \text{ for some } 0 < d < 1 \text{ as } n \to \infty. \tag{3.21}$$

Let $G : \mathbb{R} \to \mathbb{R}$ be a measurable function such that $EG(X_1) = 0$ and $EG(X_1)^2 < \infty$. We consider the stochastic process $Y_n = G(X_n)$, $n \geq 1$, an instantaneous transformation of the Gaussian process. How fast do the correlations of this process decay?

3.3. **Hermite polynomials.** For a fixed $x \in \mathbb{R}$ consider the function

$$h_x(a) = e^{-(a+x)^2/2},\ a \in \mathbb{R}.$$

This function has a Taylor expansion

$$h_x(a) = \sum_{k=0}^{\infty} \frac{a^k}{k!} \frac{d^k h}{da^k}(0) = \sum_{k=0}^{\infty} \frac{a^k}{k!} \frac{d^k(e^{-x^2/2})}{dx^k}.$$

Replacing a with $-a$ we get

$$e^{-x^2/2}e^{ax-a^2/2} = e^{-(x-a)^2/2} = \sum_{k=0}^{\infty}(-1)^k \frac{a^k}{k!}\frac{d^k(e^{-x^2/2})}{dx^k}.$$

Therefore

$$e^{ax-a^2/2} = \sum_{k=0}^{\infty}\frac{a^k}{k!}\left((-1)^k e^{x^2/2}\frac{d^k(e^{-x^2/2})}{dx^k}\right).$$

One denotes

$$H_k(x) = (-1)^k e^{x^2/2}\frac{d^k(e^{-x^2/2})}{dx^k},\ x \in \mathbb{R}, \tag{3.22}$$

for $k = 0, 1, 2, \ldots$, and we call H_k the kth Hermite polynomial. The way we introduced Hermite polynomials we see that

$$e^{ax-a^2/2} = \sum_{k=0}^{\infty} H_k(x)\frac{a^k}{k!},\ a \in \mathbb{R}, \tag{3.23}$$

and so, for a fixed $x \in \mathbb{R}$, the sequence $(H_k(x),\ k = 0, 1, 2, \ldots)$ is the sequence of the coefficients in the expansion of the function $g_x(a) = e^{ax-a^2/2}$ in a power series in a.

Some properties of Hermite polynomials.

1. H_k is a polynomial of degree k.

$$H_0(x) = 1,\ H_1(x) = x,$$

$$H_2(x) = x^2 - 1,\ H_3(x) = x^3 - 3x,$$

etc.

2. If X and Y are two jointly normal zero mean random variables with variances equal to one and correlation ρ, then

$$EH_k(X) = 0 \text{ for all } k \neq 0$$

and

$$E\left[H_k(X)\, H_m(Y)\right] = \begin{cases} 0 & \text{if } m \neq k \\ k!\rho^k & \text{if } m = k \end{cases}.$$

3. Hermite polynomials $H_0, H_1, \ldots$ form an orthogonal basis in the space

$$L^2 = L^2\left(\mathbb{R}, \mathcal{B}, \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx\right).$$

The last two properties imply that, if X is a standard normal random variable, then for any measurable function G such that $EG(X)^2 < \infty$ one can write

$$G(X) = \sum_{k=0}^{\infty}\frac{a_k}{k!}\, H_k(X), \tag{3.24}$$

with

$$a_k = E\left[H_k(X)G(X)\right],\ k = 0, 1, 2, \ldots, \tag{3.25}$$

in the sense that the series in the right hand side of (3.24) converges to $G(X)$ in L^2.

If, in addition, $EG(X) = 0$, then $a_0 = 0$, and so

$$G(X) = \sum_{k=1}^{\infty} \frac{a_k}{k!} H_k(X).$$

In general, for a given function G as above let

$$k_G = \inf\{k = 1, 2, \ldots :\ a_k \neq 0\}. \tag{3.26}$$

The number k_G is called *the Hermite rank* of the function G. We have

$$G(X) = \sum_{k=k_G}^{\infty} \frac{a_k}{k!} H_k(X),$$

and it turns out that the rate of decay of correlations of the stationary stochastic process $Y_n = G(X_n)$, $n = 1, 2, \ldots$, where $X_1, X_2, \ldots$ zero mean variance one stationary Gaussian process with covariance function satisfying (3.21) depends significantly on the the Hermite rank k_G of the function G.

Here is a theorem due to Taqqu (1975, 1979) and Dobrushin and Major (1979). Once again, we assume that the relation (3.21) holds:

$$R_n \sim a\, n^{-d} \text{ for some } 0 < d < 1 \text{ as } n \to \infty.$$

Theorem 18. *The covariance function R_n^Y of the process $Y_1, Y_2, \ldots$ satisfies*

$$R_n^Y \sim a \left(\frac{a_{k_G}}{k_G!}\right)^2 n^{-dk_G} \text{ as } n \to \infty. \tag{3.27}$$

Furthermore:

(i) Suppose that

$$k_G > \frac{1}{d}. \tag{3.28}$$

Then (by (3.27)) the correlations of the process $Y_1, Y_2, \ldots$ are summable,

$$\operatorname{Var}\left(\sum_{j=1}^{n} Y_j\right) \sim n\sigma_*^2 \text{ as } n \to \infty,$$

with

$$\sigma_*^2 = \sum_{k=k_G}^{\infty} \frac{a_k^2}{k!} \sigma_k^2 \in (0, \infty),$$

and

$$\sigma_k^2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} R_{|i-j|}^k < \infty \text{ for } k \geq k_G.$$

In addition, Functional Central Limit Theorem (3.2) holds:

$$\frac{1}{\sqrt{n}} S^{(n)} \Rightarrow \sigma_* B \ \textit{in} \ D[0,1],$$

where $S_n = Y_1 + \ldots Y_n$, $n = 0, 1, 2, \ldots$, $S^{(n)}(t) = S_{[nt]}, 0 \leq t \leq 1$, *and* B *is the standard Brownian motion on* $[0,1]$.

(ii) Suppose now that

$$k_G < \frac{1}{d}. \tag{3.29}$$

Then a Functional Non-Central Limit Theorem holds:

$$\frac{1}{n^{1-dk_G/2}} S^{(n)} \Rightarrow Z^{(d,k_G)} \ \textit{in} \ D[0,1], \tag{3.30}$$

where for $k < 1/d$

$$Z^{(d,k)}(t) = C(d,k) \int_{I\!R} \ldots \int_{I\!R} \frac{e^{it(x_1+\ldots+x_k)} - 1}{i(x_1 + \ldots + x_k)} |x_1|^{(d-1)/2} \ldots |x_k|^{(d-1)/2} W(dx_1) \ldots W(dx_k),$$

$0 \leq t \leq 1$, *where*

$$C(d,k) = 2\Gamma(d) \cos \frac{\pi d}{2} \frac{a_k}{k!},$$

and W *is a (complex-valued) Gaussian random measure with Lebesgue control measure.*

The process $(Z^{(d,k)}(t),\, t \geq 0)$ in part (ii) of Theorem 18) is a SSSI process with $H = 1 - dk/2$. This process is non-Gaussian if $k > 1$, and it is an FBM (with $H = 1 - d/2$) if $k = 1$.

One can argue that, if the Hermite rank k_G of the function G satisfies $k_G > 1/d$ (part (i) of Theorem 18), then the resulting process $Y_n = G(X_n)$, $n = 1, 2, \ldots$ has short memory: its correlation function is summable, and it satisfies Functional Central Limit Theorem.

On the other hand, if the Hermite rank k_G satisfies $k_G < 1/d$ (part (ii) of Theorem 18), then the process $Y_n = G(X_n)$, $n = 1, 2, \ldots$ has long memory: its correlations are regularly varying at infinity with exponent $-1 < d < 0$, and it satisfies Functional Non-Central Limit Theorem.

It turns out that the case $k_G = 1/d$ also puts us in the framework of Functional Central Limit Theorem, but the normalization is no longer $1/\sqrt{n}$ (the correlations are not summable).

Note that the situation described by Theorem 18 is worrisome.

There are two different one-to-one functions G_1 and G_2 of different Hermite ranks. Then, starting with a long memory (in the sense of the rate of decay of correlations) Gaussian process $(X_1, X_2, \ldots)$ we may end up with a process $Y_n^{(1)} = G_1(X_n)$, $n = 1, 2, \ldots$ that has a short memory in the sense of fast decaying correlations, and a process $Y_n^{(2)} = G_2(X_n)$, $n = 1, 2, \ldots$ that has a long memory in the sense of slowly decaying correlations.

However,

$$Y_n^{(2)} = G_2\left(G_1^{-1}(Y_n^{(1)})\right),\ n = 1, 2, \ldots.$$

and, since the function $G_2 \circ G_1^{-1}$ is one-to-one, the process $Y_n^{(2)} = G_2(X_n)$, $n = 1, 2, \ldots$ should "remember" exactly as much as the process $Y_n^{(1)} = G_1(X_n)$, $n = 1, 2, \ldots$ does!

All the points of view on long range dependence we have discussed so far rely one way or the other on correlations. Concentrating too much on the correlations has, however, a number of drawbacks.

- **Correlations provide only very limited information about the process if the process is "not very close" to being Gaussian**.

  Nobody has argued that Fractional Gaussian noise with $H > 1/2$ is not long range dependent, and in this case, indeed, correlations tell the entire story. However, for processes like ARCH or GARCH processes, or fractionally differenced processes of this kind, correlations are zero in spite of a very rich dependence structure in the process. Finally, it is often difficult to relate correlations to functional of the process that are of real interest.
- **Rate of decay of correlations may change** significantly after instantaneous one-to-one transformations of the process.
- **What to do if the variance is infinite**?

  Whatever the drawbacks of using correlations to measure length of memory in the L^2 case, the whole approach breaks down when the variance is infinite. Some of the proposed ways out in specific situations included computing "correlation-like" numbers, or using instead characteristic functions by studying the rate of convergence to zero of the difference

  $$\varphi_{X_1,X_{n+1}}(\theta_1, \theta_2) - \varphi_{X_1}(\theta_1)\varphi_{X_{n+1}}(\theta_2)$$

  $$= Ee^{i(\theta_1 X_1 + \theta_2 X_{n+1})} - Ee^{i\theta_1 X_1} Ee^{i\theta_2 X_{n+1}}$$

  for some θ_1, θ_2 not equal to zero. This approaches have met only with limited success.

We advocate a different approach to the problem of long range dependence, and we start, for now, with a generic example.

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

Suppose that R is a functional on $I\!R^\infty$. We view $R(X_1, X_2, \ldots)$ as a functional of the stochastic process; assume that it is a functional of interest. Of course, its behavior is different, in general, under different laws $\mathcal{P}_\theta$, and we are looking at how this behavior changes as $\theta \in \Theta$ changes.

Suppose that there is a partition of the parameter space Θ into two parts, Θ_0 and Θ_1, such that the behavior of the functional changes dramatically as one crosses the boundary between Θ_0 and Θ_1. Then it may make sense to talk about that boundary as the boundary between short range dependence and long range dependence. This approach makes long range dependence appear in a sort of a *phase transition*.

In many cases the functional R of the process is really a sequence of functionals, $R = (R_1, R_2, \ldots)$, where R_n is a functional on $\mathbb{R}^n$, $n = 1, 2, \ldots$. An example considered implicitly before is that of

$$R_n(x_1, \ldots, x_n) = x_1 + \ldots + x_n, \tag{3.31}$$

$n = 1, 2, \ldots$, and then the phase transition we can be looking for is that of the change in the rate of growth of the partial sums of the process.

Of course, not every phase transition indicates a change from short memory to long memory. For example, certain changes in parameters may mean changing heaviness of the tail, that is also likely to induce changes in the behavior of important functionals of the process. For example, for the functional described by (3.31) the transition between finite and infinite second moments can change the rate of growth of the partial sums of the process even without any change in the dependence structure of the process (e.g. in the iid situation).

In many cases we will be interested in phase transitions related to certain *rare events* and functionals related to such rare events.

4. Rare events and large deviations

By definition, rare events are those events that do not happen very often. The ordinary usage of the language is to associate rare events with certain limiting procedures. Let us start with some examples.

Let $(X_1, X_2, \ldots)$ be a stationary stochastic process.

Example 19. For large $\lambda > 0$ the event $\{X_1 > \lambda\}$ is a rare event.

The probability of this event is the tail probability for the process. On the other hand, this event is so elementary that it does not tell us anything about the memory in the process.

Example 20. For $k \geq 1$ and large $\lambda_1, \ldots, \lambda_k$ the event $\{X_1 > \lambda_1, \ldots, X_k > \lambda_k\}$ is also rare event.

The probabilities of such events have, obviously, a lot to do with the tails of the process. However, they can carry very important information about the dependence in the process.

Example 21. For large $n \geq 1$ and a positive sequence $(\lambda_j)_{j\geq 0}$ that does not converge to zero the event $\{X_j > \lambda_j,\ j = 1, \ldots, n\}$ is a rare event.

Even though it is less obvious here, probabilities of such events may have a lot to do with the tails of the process. The connection to the memory of the process is obvious here. The case $\lambda_j = \lambda > 0$ for all $j \geq 0$ is often interesting and appealing.

The term "large deviations" is a vague one, and is used in different ways by different people. An example of a fairly restrictive meaning of the term is "asymptotic computation of small probabilities on an exponential scale" (Dembo and Zeitouni (1993)). We will use this term in a much wider sense. For us "large deviations" are synonymous with "rare events", or with "things happening in other than expected way".

Large Deviations Approach: **unlikely things happen in the most likely way**.

By itself, this statement does not provide anything new. However, it guides one towards understanding of how rare events happen, and towards classifying rare events.

The first benefit of the large deviations approach is to realize that certain rare events tend to happen in different ways in the heavy tailed cases and in the light tailed cases.

Heavy tailed case: **rare events are caused by the smallest possible number of individual factors**

Light tailed case: **rare events are caused by "conspiracy" among all or most of individual factors**

Let us demonstrate what this means by considering a few examples.

Example 22. Let X_1 and X_2 be independent random variables with the same distribution. Consider the rare event

$$A = \{X_1 + X_2 > \lambda\}, \ \lambda \text{ large.}$$

Recall that, if the random variables X_1 and X_2 are subexponential, then

$$\begin{aligned} P(A) = P(X_1 + X_2 > \lambda) &\sim P(X_1 > \lambda) + P(X_2 > \lambda) \\ &\sim P\left(\{X_1 > \lambda\} \cup \{X_2 > \lambda\}\right). \end{aligned}$$

That is, in this case the rare event $A = \{X_1 + X_2 > \lambda\}$, is most likely to be caused by one of the two individual factors: X_1 is appropriately large (and X_2 is not outrageously small), or X_2 is appropriately large (and X_1 is not outrageously small). This is typical of the heavy tailed case.

On the other hand, let X_1 and X_2 be independent standard normal random variables, and consider the same event A as above. Obviously,

$$P(A) = P(X_1 + X_2 > \lambda) \sim \frac{1}{\sqrt{\pi}} \lambda^{-1} e^{-\lambda^2/4}.$$

On the other hand, for every $1/2 < \tau \leq 1$,

$$P(X_1 + X_2 > \lambda, X_1 > \tau\lambda) = P(X_1 + X_2 > \lambda, X_1 > \lambda) + P(X_1 + X_2 > \lambda, \tau\lambda < X_1 \leq \lambda).$$

Now,

$$P(X_1 + X_2 > \lambda, X_1 > \lambda) \leq P(X_1 > \lambda) \sim \frac{1}{\sqrt{2\pi}} \lambda^{-1} e^{-\lambda^2/2} = o(P(A)),$$

while

$$\begin{aligned} P(X_1 + X_2 > \lambda, \tau\lambda < X_1 \leq \lambda) &= \frac{1}{2\pi} \int_{\tau\lambda}^{\lambda} e^{-x^2/2}\, dx \int_{\lambda - x}^{\infty} e^{-y^2/2}\, dy \\ \leq \frac{1}{2\sqrt{2\pi}} \int_{\tau\lambda}^{\lambda} e^{-x^2/2} e^{-(\lambda - x)^2/2}\, dx &= \frac{1}{2\sqrt{2\pi}} e^{-\lambda^2/4} \int_{\tau\lambda}^{\lambda} e^{-(x - \lambda/2)^2}\, dx \end{aligned}$$

$$\leq \frac{1}{2\sqrt{2\pi}} e^{-\lambda^2/4} \int_{(\tau - 1/2)\lambda}^{\infty} e^{-x^2}\, dx \sim \frac{1}{2\sqrt{2\pi}} e^{-\lambda^2/4} ((2\tau - 1)\lambda)^{-1} e^{-((2\tau-1)\lambda)^2} = o(P(A)).$$

Therefore, every $1/2 < \tau \leq 1$,

$$P(X_1 + X_2 > \lambda, X_1 > \tau\lambda) = o(P(X_1 + X_2 > \lambda)).$$

In other words:

the rare event $P(X_1+X_2>\lambda)$ is most likely to occur because both of the terms in the sum, X_1 and X_2 are at the same time at about the level $\frac{1}{2}\lambda$.

This is an example of "conspiracy" (of X_1 and X_2).

Example 23. Let $X_1, X_2, \dots$ be iid random variables with a finite mean μ. For an $\epsilon>0$ consider the rare event

$$A=\left\{\frac{X_1+X_2+\ldots+X_n}{n}>\mu+\epsilon\right\},\ n \text{ large.}$$

Let consider first the heavy tailed case, which in the present case will mean that the random variables $X_1, X_2, \dots$ have a regularly varying right tail with exponent $\alpha>1$.

Observe that for any $\epsilon'>\epsilon$

$$P(A)=P\left(\frac{X_1+X_2+\ldots+X_n}{n}>\mu+\epsilon\right)$$

$$\geq P\left(\bigcup_{i=1}^n\left\{X_i>\epsilon' n,\ \sum_{j=1,\ldots,n,j\neq i} X_j>n(\mu+\epsilon-\epsilon')\right\}\right):=P(\cup_{i=1}^n B_i)$$

$$\geq \sum_{i=1}^n P(B_i)=nP(B_1)-\frac{n(n-1)}{2}P(B_1\cap B_2)$$

$$\geq nP(X_1>\epsilon' n)\,P(\sum_{j=2}^n X_j>n(\mu+\epsilon-\epsilon'))-\frac{n(n-1)}{2}P(X_1>\epsilon' n, X_2>\epsilon' n)$$

$$\sim nP(X_1>\epsilon n)(\frac{\epsilon}{\epsilon'})^\alpha-\frac{n(n-1)}{2}(P(X_1>\epsilon' n))^2,$$

because by the regular variation

$$\lim_{n\to\infty}\frac{P(X_1>\epsilon' n)}{P(X_1>\epsilon n)}=(\frac{\epsilon}{\epsilon'})^\alpha,$$

and by the law of large numbers,

$$P(\sum_{j=2}^n X_j>n(\mu+\epsilon-\epsilon'))\to 1.$$

Note that

$$nP(X_1>\epsilon n) \text{ is regularly varying with exponent } -(\alpha-1),$$

and

$$\frac{n(n-1)}{2}(P(X_1>\epsilon' n))^2 \text{ is regularly varying with exponent } -2(\alpha-1).$$

Therefore,

$$\liminf_{n\to\infty}\frac{P(A)}{nP(X_1>\epsilon n)}\geq(\frac{\epsilon}{\epsilon'})^\alpha,$$

and since this is true for any $\epsilon' > \epsilon$, we can let $\epsilon' \downarrow \epsilon$ to conclude that

$$\liminf_{n\to\infty} \frac{P(A)}{nP(X_1 > \epsilon n)} \geq 1. \tag{4.1}$$

To prove the corresponding asymptotic upper bound, let $0 < \tau < \epsilon$ be a small number. Observe that

$$P(A) = P\left(\frac{X_1 + X_2 + \ldots + X_n}{n} > \mu + \epsilon\right)$$

$$= P\left(A \cap \{\text{for some } i = 1, \ldots, n, \, X_i > \tau n\}\right) + P\left(A \cap \{\text{for all } i = 1, \ldots, n, \, X_i \leq \tau n\}\right).$$

We will show that, if $\tau > 0$ is small enough, then

$$\lim_{n\to\infty} \frac{P\left(A \cap \{\text{for all } i = 1, \ldots, n, \, X_i \leq \tau n\}\right)}{nP(X_1 > \epsilon n)} = 0. \tag{4.2}$$

In that case, for all τ small enough,

$$\limsup_{n\to\infty} \frac{P(A)}{nP(X_1 > \epsilon n)} \leq \limsup_{n\to\infty} \frac{P\left(A \cap \{\text{for some } i = 1, \ldots, n, \, X_i > \tau n\}\right)}{nP(X_1 > \epsilon n)}.$$

Take now any $\tau < \epsilon' < \epsilon$, and write

$$P\left(A \cap \{\text{for some } i = 1, \ldots, n, \, X_i > \tau n\}\right)$$

$$\leq P\left(\bigcup_{i=1}^{n} \{X_i > \epsilon' n\}\right) + P\left(\bigcup_{i=1}^{n} \left\{X_i > \tau n, \sum_{j=1,\ldots,n, j\neq i} X_j > n(\mu + \epsilon - \epsilon')\right\}\right)$$

$$\leq nP(X_1 > \epsilon' n) + nP\left(X_1 > \tau n, \sum_{j=2}^{n} X_j > n(\mu + \epsilon - \epsilon'))\right)$$

$$\sim nP(X_1 > \epsilon n)(\frac{\epsilon}{\epsilon'})^{\alpha} + nP(X_1 > \tau n)\, P(\sum_{j=2}^{n} X_j > n(\mu + \epsilon - \epsilon')).$$

Since by the law of large numbers,

$$P(\sum_{j=2}^{n} X_j > n(\mu + \epsilon - \epsilon')) \to 0,$$

we conclude that

$$\limsup_{n\to\infty} \frac{P\left(A \cap \{\text{for some } i = 1, \ldots, n, \, X_i > \tau n\}\right)}{nP(X_1 > \epsilon n)} \leq (\frac{\epsilon}{\epsilon'})^{\alpha},$$

and since this is true for any $\tau < \epsilon' < \epsilon$, we can let $\epsilon' \uparrow \epsilon$ to conclude that

$$\limsup_{n\to\infty} \frac{P(A)}{nP(X_1 > \epsilon n)} \leq 1. \tag{4.3}$$

Therefore, we will have established the relationship

$$\lim_{n\to\infty} \frac{P(A)}{nP(X_1 > \epsilon n)} = 1, \tag{4.4}$$

provided we can show that for all $\tau > 0$ small enough, the relationship (4.2) holds.

To this end, let $X_1^*, \ldots, X_n^*$ be iid random variables with the common distribution being that of X_1 conditioned on $|X_1| \leq \tau n$. Denote

$$\mu_* = EX_1^* \leq \mu + \frac{\epsilon}{2}$$

for large n. Note that

$$P\left(A \cap \{\text{for all } i = 1, \ldots, n,\ X_i \leq \tau n\}\right) \leq P\left(A \middle| X_i \leq \tau n,\ i = 1, \ldots, n\right)$$

$$= P\left(\frac{X_1 + \ldots + X_n}{n} > \mu + \epsilon \middle| X_i \leq \tau n,\ i = 1, \ldots, n\right)$$

$$\leq P\left(\frac{X_1 + \ldots + X_n}{n} > \mu + \epsilon \middle| |X_i| \leq \tau n,\ i = 1, \ldots, n\right)$$

$$= P\left(X_1^* + \ldots + X_n^* > n(\mu + \epsilon)\right) \leq P\left((X_1^* - \mu_*) + \ldots + (X_n^* - \mu_*) > n\frac{\epsilon}{2}\right).$$

The following lemma is very useful.

Lemma 24. *Let $Y_1, Y_2, \ldots$ be iid zero mean random variables, such that for some $c > 0$ we have $|Y_1| \leq c$ a.s. Let $S_n = Y_1 + \ldots + Y_n$, $n \geq 1$. Then*

$$P(S_n > \lambda) \leq \exp\left\{-\frac{\lambda}{2c} \operatorname{arsinh} \frac{c\lambda}{2 \operatorname{Var}(S_n)}\right\} \tag{4.5}$$

for every $\lambda > 0$.

See Prokhorov (1959), also Petrov (1995). Here

$$\operatorname{arsinh}(y) = \left(\frac{e^x - e^{-x}}{2}\right)^{-1}(y) \geq \log(2y) \tag{4.6}$$

for $y \geq 2$.

In our case, $Y_i = X_i^* - \mu_*$, $i = 1, \ldots, n$, $c = \tau n + |\mu_*| \leq 2\tau n$ for large n, and $\lambda = n\epsilon/2$. It is easy to check that there is $\beta < 2$ and $C > 0$ such that for every $n \geq 1$

$$\operatorname{Var}\left((X_1^* - \mu_*) + \ldots + (X_n^* - \mu_*)\right) \leq Cn^\beta.$$

Therefore,

$$\frac{c\lambda}{2 \operatorname{Var}(S_n)} \geq \frac{(\tau\epsilon/2)n^2}{2Cn^\beta} = C(\tau, \epsilon)n^{2-\beta},$$

and so by (4.5) and (4.6) we have

$$P\left((X_1^* - \mu_*) + \ldots + (X_n^* - \mu_*) > n\frac{\epsilon}{2}\right)$$

$$\leq \exp\left\{-\frac{n\epsilon/2}{2n\tau} \log(2C(\tau, \epsilon)n^{2-\beta})\right\} \sim \text{const}\, n^{-\epsilon(2-\beta)/(4\tau)}.$$

Therefore, if $\tau > 0$ is small enough for

$$\frac{\epsilon(2 - \beta)}{4\tau} > \alpha - 1,$$

then

$$\lim_{n\to\infty} \frac{P\left((X_1^* - \mu_*) + \ldots + (X_n^* - \mu_*) > n\frac{\epsilon}{2}\right)}{nP(X_1 > \epsilon n)} = 0,$$

which proves (4.2).

Summarizing, we have proved that in the heavy tailed case,

$$\lim_{n\to\infty} \frac{P\left(\frac{X_1+X_2+\ldots+X_n}{n} > \mu + \epsilon\right)}{nP(X_1 > \epsilon n)} = 1. \tag{4.7}$$

Recall that heavy tails here mean regularly varying right tails with exponent $\alpha > 1$.

Rewriting the statement (4.7) in the form

$$\lim_{n\to\infty} \frac{P\left(X_1 + X_2 + \ldots + X_n > n(\mu + \epsilon)\right)}{nP(X_1 > \epsilon n)} = 1$$

we see another demonstration of the large deviations approach in the heavy tailed case, saying that rare events are caused by the smallest possible number of individual factors.

Here the individual factors are the terms X_1 through X_n that behave in an unusual way, and to cause the particular rare event a given term X_i has to be greater than ϵn, because all the other terms in the sum, behaving *in the usual way*, add up to about $(n-1)\mu \sim n\mu$.

We emphasize that the large deviations approach is not a meta-theorem, and one has to make sure whether it provides the right intuition in any given situation. For example, the statement (4.7) does NOT hold for all subexponential distributions.

To see how the large deviations approach in the light tailed case works in this case assume that the random variables $X_1, X_2, \ldots$ have finite exponential moments:

$$Ee^{\theta X_1} < \infty \text{ for all } \theta > 0.$$

Note that the equation

$$\frac{E\left(X_1 e^{\theta X_1}\right)}{Ee^{\theta X_1}} = \mu + \epsilon$$

has a unique solution θ_* that belongs to $(0, \infty)$.

If F is the common law of $X_1, X_2, \ldots$ then F_* defined by

$$F_*(B) = \frac{1}{Ee^{\theta_* X_1}} \int_B e^{\theta_* x}\, F(dx),\ \ B \text{ a Borel set}$$

is another probability measure on $\mathbb{R}$, such that

$$\int_{\mathbb{R}} x\, F_*(dx) = \mu + \epsilon.$$

It turns out that the most likely way for the event

$$A = \left\{\frac{X_1 + X_2 + \ldots + X_n}{n} > \mu + \epsilon\right\}$$

to happen for large n is for $X_1, \ldots, X_n$ to take values as if their common law was not F but, rather, F_*.

That is, the terms $X_1, \ldots, X_n$ "conspire by changing their law", so that they will have "the right mean" $\mu + \epsilon$.

Then one shows that

$$\lim_{n\to\infty} \frac{1}{n} \left(-\log P\left(\frac{X_1 + X_2 + \ldots + X_n}{n} > \mu + \epsilon \right)\right) = \theta_*(\mu+\epsilon) - \log E e^{\theta_* X_1}.$$

Of course, the orders of magnitude of the probability of the rare event

$$A = \left\{ \frac{X_1 + X_2 + \ldots + X_n}{n} > \mu + \epsilon \right\}$$

are different in the light tailed and heavy tailed cases. The important thing for us is to notice that the most likely ways this rare events happen differ between the light tailed and heavy tailed cases.

4.1. **Applications to communication networks.** Consider a simple fluid queuing system in which work arrives according to a stochastic process $(A(t),\, t \geq 0)$ with non-decreasing right continuous sample paths. That is, the amount of work arriving in the system in the interval $[0,t]$ is equal to $A(t)$ for any $t \geq 0$.

The system has a single server of capacity $r > 0$ (i.e. the server can do r units of work per unit of time). All the work that has to wait is collected in a buffer. We will consider input processes $(A(t),\, t \geq 0)$ of two different types.

Superpositions of ON-OFF processes

Let F_{on} and F_{off} be two distributions on $[0,\infty)$ with finite means that we will denote by μ_{on} and μ_{off}. A (stationary) ON-OFF process with ON distribution F_{on} and OFF distribution F_{off} is defined as follows.

Let $(X_1, X_2, \ldots)$ and $(Y_0, Y_1, Y_2, \ldots)$ be two independent sequences of random variables; $X_1, X_2, \ldots$ are iid with a common law F_{on}, and $Y_0, Y_1, Y_2, \ldots$ are iid with a common law F_{off}.

Let, further,

$$F_{\text{on}}^{(0)}(t) = \frac{1}{\mu_{\text{on}}} \int_0^t \overline{F_{\text{on}}}(u)\, du$$

and

$$F_{\text{off}}^{(0)}(t) = \frac{1}{\mu_{\text{off}}} \int_0^t \overline{F_{\text{off}}}(u)\, du$$

be the residual life time distributions of F_{on} and F_{off} accordingly.

Let $X_0^{(0)}$ and $Y_0^{(0)}$ be independent random variables that are also independent of the sequences $(X_1, X_2, \ldots)$ and $(Y_0, Y_1, Y_2, \ldots)$, such that the law of $X_0^{(0)}$ is $F_{\text{on}}^{(0)}$ and the law of $Y_0^{(0)}$ is $F_{\text{off}}^{(0)}$.

Finally, let Z be a Bernoulli random variable with success probability $\mu_{\text{on}}/(\mu_{\text{on}} + \mu_{\text{off}})$, independent of all the rest of the random variables involved.

Define the process $(I(t),\, t \geq 0)$ as follows.

If $Z = 1$, set $U_0 = X_0^{(0)} + Y_0$, $D_n = U_{n-1} + X_n$ and $U_n = D_n + Y_n$ for $n = 1, 2, \ldots$, and

$$I(t) = 1 \text{ if } 0 \leq t < X_0^{(0)} \quad \text{or if} \quad U_{n-1} \leq t < U_{n-1} + X_n$$

for some $n \geq 1$. Otherwise set $I(t) = 0$.

If $Z = 0$, set $U_0 = Y_0^{(0)}$, and define, as before, $D_n = U_{n-1} + X_n$ and $U_n = D_n + Y_n$ for $n = 1, 2, \ldots$. Now set

$$I(t) = 1 \text{ if } U_{n-1} \leq t < U_{n-1} + X_n$$

for some $n \geq 1$. Otherwise set $I(t) = 0$.

This is a stationary alternating renewal process.

It turns out that there is a natural connection between one or both of the ON and OFF distributions F_{on} and F_{off} being heavy tailed, and long range dependence in the ON-OFF process $(I(t),\, t \geq 0)$. The intuitive reason is, simply, that a single long ON or OFF interval can cover time points far apart.

Here is a precise result.

Theorem 25. *Assume that the ON distribution F_{on} has a regularly varying right tail with exponent $1 < \alpha < 2$, and that $\overline{F_{\text{off}}}(t) = o(\overline{F_{\text{on}}}(t))$ as $t \to \infty$. Let $R(t) = \text{Cov}(I(s), I(s+t))$, $t \geq 0$. Then*

$$R(t) \sim \frac{\mu_{\text{off}}^2}{(\alpha - 1)(\mu_{\text{on}} + \mu_{\text{off}})^3} t \overline{F_{\text{on}}}(t) \text{ as } t \to \infty. \tag{4.8}$$

That is, under the assumption of the theorem we have that the correlation function of the ON-OFF process $(I(t),\, t \geq 0)$ is regularly varying at infinity with exponent $-(\alpha - 1)$ so the correlations decay slowly enough for the ON-OFF process to be called long range dependent.

The result is due to Heath et al. (1998).

Of course, one gets a similar result if the OFF distribution F_{off} has a regularly varying right tail with exponent $1 < \alpha < 2$, and that $\overline{F_{\text{on}}}(t) = o(\overline{F_{\text{off}}}(t))$ as $t \to \infty$.

We call the input processes $(A(t),\, t \geq 0)$ to a queuing system a superposition of ON-OFF processes if

$$A(t) = \sum_{j=1}^{k} a_j \int_0^t I_j(s)\, ds, \; t \geq 0, \tag{4.9}$$

where $I_1, \ldots, I_k$ are independent ON-OFF processes with ON distributions $F_{\text{on}}^{(j)}$ and OFF distributions $F_{\text{off}}^{(j)}$ accordingly, and $a_1, \ldots, a_k$ are positive numbers.

Note that the jth ON-OFF stream in (4.9) brings the work at the rate a_j during its ON period and no work during its OFF period, $j = 1, \ldots, k$. Observe also that the proportion of time an ON-OFF process is in the ON state is equal to $\mu_{\text{on}}/(\mu_{\text{on}}+\mu_{\text{off}})$. Hence the average input rate for the input process $(A(t),\, t \geq 0)$ given by (4.9) is

$$\sum_{j=1}^{k} a_j \frac{\mu_{\text{on}}^{(j)}}{\mu_{\text{on}}^{(j)}+\mu_{\text{off}}^{(j)}},$$

Here $\mu_{\text{on}}^{(j)}$ and $\mu_{\text{off}}^{(j)}$ are, correspondingly, the mean ON and OFF times for the jth ON-OFF process I_j, $k = 1, \ldots, k$.

In particular, the system is stable (reaches steady state) if and only iff the service rate is higher than the input rate:

$$r > \sum_{j=1}^{k} a_j \frac{\mu_{\text{on}}^{(j)}}{\mu_{\text{on}}^{(j)}+\mu_{\text{off}}^{(j)}}. \tag{4.10}$$

Superpositions of $M/G/\infty$ inputs

Let F be a distribution on $[0,\infty)$ with a finite mean μ. Let N be a homogeneous Poisson process on $(-\infty,\infty)$ with rate $\lambda > 0$, independent of a sequence $(X_i,\, i = 0, \pm 1, \pm 2, \ldots)$ of iid random variables with a common distribution F. Let $(\Gamma_i,\, i = 0, \pm 1, \pm 2, \ldots)$ be the atoms (arrival times) of the Poisson process N. Define

$$M(t) = \sum_{i=-\infty}^{\infty} \mathbf{1}\left(\Gamma_i \leq t < \Gamma_i + X_i\right),\ t \geq 0. \tag{4.11}$$

If one imagines that at time Γ_i a customer arrives in a system and stays there for X_i units of time, $i = 0, \pm 1, \pm 2, \ldots$, then $M(t)$ is the number of customers at the system at time $t \geq 0$.

By the definition $(M(t),\, t \geq 0)$ is a stationary stochastic process; it is the number of customers in the system in an $M/G/\infty$ queue with arrival rate λ and service time distribution F.

If $A(t) = \int_0^t M(s)\, ds$, $t \geq 0$, then the input A is called $M/G/\infty$ input. The computations dealing with $M/G/\infty$ inputs are often easier than those dealing with ON-OFF inputs because of the properties of Poisson random measures.

Indeed, the pairs (Γ_i, X_i), $i = 0, \pm 1, \pm 2, \ldots$ form the points of a Poisson random measure, say, N_*, on $\mathbb{R} \times (0,\infty)$ with mean measure $m_* = \lambda \cdot \text{Leb} \times F$.

For $t \geq 0$ let V_t be a wedge in $\mathbb{R} \times (0,\infty)$ defined by

$$V_t = \{(x,y) : x \leq t,\, x+y > t\}.$$

Then $M(t) = N_*(V_t)$, $t \geq 0$. In particular, $M(t)$ is a Poisson random variable with the mean

$$EM(t) = m_*(V_t) = \lambda \int_{-\infty}^{t} dx \int_{t-x}^{\infty} F(dy) = \lambda \int_0^{\infty} \overline{F}(u)\, du = \lambda\mu.$$

In a similar way we can compute the covariance function of the stationary process $(M(t),\, t \geq 0)$. For $t > 0$ write

$$M(0) = N_*(V_0) = N_*(A_t) + N_*(B_t),$$
$$M(t) = N_*(V_t) = N_*(B_t) + N_*(C_t),$$

where $A_t = V_0 \setminus V_t$, $B_t = V_0 \cap V_t$ and $C_t = V_t \setminus V_0$, and $N_*(A_t)$, $N_*(B_t)$ and $N_*(C_t)$ are independent Poisson random variables. Hence

$$R(t) = \mathrm{Cov}(M(0), M(t)) = \mathrm{Var}(N_*(B_t))$$
$$= m_*(B_t) = \lambda \int_{-\infty}^{0} dx \int_{t-x}^{\infty} F(dy) = \lambda \int_t^{\infty} \overline{F}(u)\, du.$$

One commonly refers to X_i as the length of the session starting at time Γ_i.

If the session length distribution F is regularly varying with exponent $\alpha > 1$, then

$$R(t) \sim \frac{\lambda}{\alpha - 1} t\overline{F_{\mathrm{on}}}(t) \text{ as } t \to \infty, \tag{4.12}$$

a similar behavior to the ON-OFF case.

We call the input processes $(A(t),\, t \geq 0)$ to a queuing system a superposition of $M/G/\infty$ input processes if

$$A(t) = \sum_{j=1}^{k} a_j \int_0^t M_j(s)\, ds,\ t \geq 0, \tag{4.13}$$

where $M_1, \ldots, M_k$ are independent $M/G/\infty$ input processes with Poisson rates $\lambda_1, \ldots, \lambda_k$ and session length distributions $F_1, \ldots, F_k$ accordingly, and $a_1, \ldots, a_k$ are positive numbers. Note that one can view the process

$$\sum_{j=1}^{k} a_j\ M_j(t),\ t \geq 0$$

in (4.13) as representing the instantaneous input rate in a system where sessions arrive according to a homogeneous Poisson process with a rate $\lambda = \lambda_1 + \ldots + \lambda_k$, and an arriving session will, with probability $p_i = \lambda_i/\lambda$, have its length distributed according to F_i, and will bring work at rate a_i, $i = 1, \ldots, k$.

Observe that the process $(M(t), t \geq 0)$ hits zero infinitely often, and then regenerates. Therefore,

$$\lim_{t\to\infty} \frac{\int_0^t M(s)\, ds}{t} = EM(0) = \lambda\mu.$$

Hence the average input rate for the input process $(A(t),\, t \geq 0)$ given by (4.13) is

$$\sum_{j=1}^{k} a_j \lambda_j \mu_j,$$

where μ_j is the mean session length for the jth $M/G/\infty$ process, $j = 1, \ldots, k$.

As before, the system is stable (reaches steady state) if and only iff the service rate is higher than the input rate:

$$r > \sum_{j=1}^{k} a_j \lambda_j \mu_j. \tag{4.14}$$

We will consider two problems related the fluid queuing model of a communication network considered above.

Problem 1. Suppose that the buffer in the system is of finite capacity H, and let T_H be the first time that the buffer overflows. What is the asymptotic behavior of ET_H as $H \to \infty$?

Problem 2 Suppose that the buffer is infinite, and let W represent the stationary buffer content. What is the tail behavior of the random variable W (what is the asymptotic behavior of $P(W > w)$ as $w \to \infty$)?

We will see what the large deviations approach tell us about these two problems.

Buffer overflow problem

Let us consider first the case of the input process $(A(t),\, t \geq 0)$ being a superposition of ON-OFF processes. We assume, as usually, the stability condition (4.10):

$$r > \sum_{j=1}^{k} a_j \frac{\mu_{\text{on}}^{(j)}}{\mu_{\text{on}}^{(j)} + \mu_{\text{off}}^{(j)}}.$$

Recall that $\mu_{\text{on}}^{(j)}$ and $\mu_{\text{off}}^{(j)}$ are, correspondingly, the mean ON and OFF times for the jth ON-OFF process, and let $\mu^{(j)} = \mu_{\text{on}}^{(j)} + \mu_{\text{off}}^{(j)}$, $j = 1, \ldots, k$. To develop our intuition, let us start with the simplest case.

Assume that there is a distribution F on $(0, \infty)$ with a regularly varying right tail with exponent $\alpha \geq 1$ and a subset J^* of $\{1, \ldots, k\}$ such that for every $j \in J^*$ the limit

$$p_j = \lim_{x \to \infty} \frac{\overline{F_{\text{on}}}^{(j)}(x)}{\overline{F}(x)} > 0$$

exists and is positive, while for every $j \notin J^*$

$$\lim_{x \to \infty} \frac{\overline{F_{\text{on}}}^{(j)}(x)}{\overline{F}(x)} = 0.$$

Assume, furthermore, that

$$\max_{j \in J^*} \left(a_j + \sum_{i=1,\ldots,k,\, i \neq j} a_i \frac{\mu_{\text{on}}^{(i)}}{\mu^{(i)}} \right) > r. \tag{4.15}$$

Let us try to figure out the asymptotic behavior of the buffer overflow time ET_H using the large deviations approach. Since we are in a heavy tailed case, the unlikely event that results in a buffer overflow is, most likely, caused by the smallest number of individual factors.

It is intuitively clear that such individual factors here are exceptionally long ON periods. Since the ON-OFF processes I_j with $j \in J^*$ have ON period distributions with the heaviest tails, these are the most likely to have an exceptionally long ON period.

Since the smallest possible number of individual factors cannot be less than 1, Let us check if a single exceptionally long ON period of an ON-OFF process I_j with $j \in J^*$ is sufficient to cause an overflow of a large buffer.

Suppose, therefore, that an ON-OFF process I_j with $j \in J^*$ has an ON period of a very large length Z. During that Z units of time the jth process I_j brings work at the rate a_j. If Z is large, the other $k-1$ ON-OFF processes I_i will bring work at their average rates

$$a_i \frac{\mu_{\text{on}}^{(i)}}{\mu_{\text{on}}^{(i)} + \mu_{\text{off}}^{(i)}}, \; i = 1, \ldots, k, \; i \neq j.$$

Therefore, the overall input rate during this exceptionally long ON period is

$$a_j + \sum_{i=1,\ldots,k,i\neq j} a_i \frac{\mu_{\text{on}}^{(i)}}{\mu^{(i)}}.$$

If

$$r > a_j + \sum_{i=1,\ldots,k,i\neq j} a_i \frac{\mu_{\text{on}}^{(i)}}{\mu^{(i)}},$$

then during this ON period buffer content tends to go down and, hence, this long ON period is not likely to be sufficient to cause buffer overflow.

On the other hand, if

$$r < a_j + \sum_{i=1,\ldots,k,i\neq j} a_i \frac{\mu_{\text{on}}^{(i)}}{\mu^{(i)}}, \tag{4.16}$$

then during this long ON period buffer content tends to go up at the rate

$$a_j + \sum_{i=1,\ldots,k,i\neq j} a_i \frac{\mu_{\text{on}}^{(i)}}{\mu^{(i)}} - r.$$

Assuming that buffer content is not exceptionally large before the beginning of the long ON period, the amount of work in the buffer at the end of this period will be about

$$Z\left(a_j + \sum_{i=1,\ldots,k,i\neq j} a_i \frac{\mu_{\text{on}}^{(i)}}{\mu^{(i)}} - r\right).$$

Therefore, an exceptionally long ON period of the ON-OFF process I_j, of length Z, is likely to cause overflow of a large buffer of size H if both (4.16) holds and

$$Z > \frac{H}{a_j + \sum_{i=1,\ldots,k,i\neq j} a_i \frac{\mu_{\text{on}}^{(i)}}{\mu^{(i)}} - r}. \tag{4.17}$$

Now we see the importance of the assumption (4.15): there are $j \in J^*$ for which (4.16) holds. Let us denote the set of such j by J^*_+, and we denote

$T_j =$ the first time jth ON-OFF process starts an ON period whose length satisfies (4.17), $j \in J^*_+$.

Let us mention, at this point, why we are assuming, in our informal calculations, that

(1) During an exceptionally long ON period of the jth ON-OFF process the other ON-OFF processes bring in work, roughly speaking, according to their average rate, and
(2) At the beginning on the exceptionally long ON period the buffer content is not very large.

Indeed, during an exceptionally long ON period of the jth ON-OFF process the law of large numbers kicks in for the other ON-OFF processes. If one of them violated the law of large numbers, this would be another exceptional event (factor) in addition to the already exceptionally long ON period, and the large deviations approach tells us that, if one factor is sufficient, then other factors will play no role.

For a similar reason the buffer is not likely to be very full at the beginning of the exceptionally long ON period. Since the overall service rate is higher than the overall input rate (condition 4.10) the is usually not very full, and this would have been a second exceptional event (factor), which should not be there according to the large deviations approach.

Let us go back to the above fact that for many purposes one can replace all the input streams except for the important ones by the non-random, averaged streams. Note that this results in effective reduction of the service rate from r to

$$r - \sum_{i=1,\ldots,k, i\neq j} a_i \frac{\mu^{(i)}_{\mathrm{on}}}{\mu^{(i)}}$$

(see (4.17)), while the process I_j runs its long ON period.

In general, input processes that are unlikely to produce exceptional factors that may cause a rare event of interest are replaced by the average inputs, reducing the effective service rates and leaving the "important" input processes as the remaining random input.

Results of this type are often called in communications research literature by the name *reduced load equivalence*. See, for example, Agrawal et al. (1999), Zwart et al. (2000) and Jelenković et al. (2002).

We are now ready to identify the asymptotic behavior of ET_H, the time until buffer overflow. Let

$$T^*_H = \min_{j\in J^*_+} T_j. \tag{4.18}$$

The large deviation approach tells us that buffer overflow occurs at approximately linear in H time after one of the ON-OFF processes I_j with $j \in J^*_+$ starts an ON period whose length satisfies (4.17). The time T^*_H is exactly the first time such an event happens. We will see shortly that ET_H^* is regularly varying with exponent $\alpha > 1$ as $H \to \infty$. Therefore, we conclude that

$$ET_H \sim ET_H^* \text{ as } H \to \infty. \tag{4.19}$$

Let us, therefore, figure out the asymptotic behavior of ET_H^* as $H \to \infty$.

We will consider, for simplicity, instead of the stationary input process the case where every ON-OFF process I_j starts, at time zero, at the beginning of an ON period. We will comment on what happens in the stationary case a bit later.

Denote

$$\beta_j = a_j + \sum_{i=1,\ldots,k,i\neq j} a_i \frac{\mu_{\text{on}}^{(i)}}{\mu^{(i)}} - r, \; j \in J_+^*. \tag{4.20}$$

Recall that at time T_j the process I_j starts an ON period of the length greater than H/β_j.

Observe that for every $j \in J_+^*$

$$T_j \stackrel{\text{d}}{=} \sum_{i=1}^{K_j} \left(X_i^{(H,-)} + Y_i \right).$$

Here $(X_1^{(H,-)}, X_2^{(H,-)}, \ldots)$ is a sequence of iid random variables with a common distribution given by

$$P(X_1^{(H,-)} \in A) = P\left(X_1 \in A \,\middle|\, X_1 \leq \frac{H}{\beta_j} \right) = \frac{F_{\text{on}}^{(j)}(A \cap [0, H/\beta_j])}{F_{\text{on}}^{(j)}([0, H/\beta_j])},$$

i.e. ON times conditioned on not exceeding H/β_j. Further, $(Y_1, Y_2, \ldots)$ is the usual sequence of iid OF times with their usual distribution $F_{\text{off}}^{(j)}$. All random variables involved are independent.

Note that, as $H \to \infty$,

$$\overline{F_{\text{on}}}^{(j)}(H/\beta_j) K_j \Rightarrow E \sim \exp(1),$$

while $X_1^{(H,-)}$ increases stochastically to X_1, a random variable with the law $F_{\text{on}}^{(j)}$. Therefore, by the law of large numbers,

$$\overline{F_{\text{on}}}^{(j)}(H/\beta_j) T_j \stackrel{\text{d}}{=} \left(\overline{F_{\text{on}}}^{(j)}(H/\beta_j) K_j \right) \frac{1}{K_j} \sum_{i=1}^{K_j} \left(X_i^{(H,-)} + Y_i \right) \Rightarrow \mu^{(j)} E. \tag{4.21}$$

We conclude that for any $t > 0$

$$P(T_H^* > t) = \prod_{j \in J_+^*} P(T_j > t) \sim \prod_{j \in J_+^*} P\left(E > t(\mu^{(j)})^{-1} \overline{F_{\text{on}}}^{(j)}(H/\beta_j) \right)$$

$$\sim \prod_{j \in J_+^*} \exp\{-t(\mu^{(j)})^{-1} \overline{F_{\text{on}}}^{(j)}(H/\beta_j)\} \sim \prod_{j \in J_+^*} \exp\{-t(\mu^{(j)})^{-1} p_j \overline{F_{\text{on}}}(H) \beta_j^{\alpha}\}$$

$$= \exp\left\{ -t \left(\sum_{j\in J_+^*} (\mu^{(j)})^{-1} p_j \beta_j^\alpha \right) \overline{F_{\text{on}}}(H) \right\}.$$

We obtain, therefore, a weak convergence result

$$\overline{F_{\text{on}}}(H) T_H^* \Rightarrow \left(\sum_{j\in J_+^*} (\mu^{(j)})^{-1} p_j \beta_j^\alpha \right) E \tag{4.22}$$

as $H \to \infty$. Since it is easy to check that $\overline{F_{\text{on}}}(H) T_H^*$ is uniformly integrable, we also obtain convergence of the means:

$$ET_H^* \sim \frac{1}{\overline{F_{\text{on}}}(H)} \frac{1}{\sum_{j\in J_+^*} (\mu^{(j)})^{-1} p_j \beta_j^\alpha}. \tag{4.23}$$

Using (4.19) we then obtain the solution to our problem: asymptotic behavior of the expected time until overflow of a large buffer:

$$ET_H \sim \frac{1}{\overline{F_{\text{on}}}(H)} \frac{1}{\sum_{j\in J_+^*} (\mu^{(j)})^{-1} p_j \beta_j^\alpha} \tag{4.24}$$

as $H \to \infty$, where $\beta_j, j \in J_+^*$ are given by (4.20) above.

Recall that we have derived the result (4.24) by replacing the assumption of stationary input by the assumption that every ON-OFF process I_j in the input to the system, starts, at time zero, at the beginning of an ON period.

What happens if the input is stationary? It is easy to see that the same asymptotic result remains true if one assumes, for example, that all OFF time distributions $F_{\text{off}}^{(j)}, j = 1, \ldots, k$ have a finite second moment.

To see why we are talking about a finite second moment, assume, for a moment, that $k = 1$ (the input stream consists of a single ON-OFF process) and that F_{off} has infinite second moment.

A stationary ON-OFF process starts, with a positive probability, with a special OFF period, whose distribution is the residual life time distribution $F_{\text{off}}^{(0)}$. The latter, of course, will have infinite mean in our situation. Since the buffer cannot overflow until this first OFF period is over, the expected time until overflow will be infinite in this case.

Assuming finite second moment for all the OFF time distributions eliminates problems of this kind. However, in the case of a superposition of multiple ON-OFF inputs this assumption is not necessary, and our conclusion (4.24) will hold under various rate assumptions even if the second moments of some of the OFF times are infinite.

The situation becomes quite different if the assumption (4.15)

$$\max_{j\in J^*} \left(a_j + \sum_{i=1,\ldots,k, i\neq j} a_i \frac{\mu_{\text{on}}^{(i)}}{\mu^{(i)}} \right) > r$$

does not hold.

To see what this means, let B be a subset of $\{1, \ldots, k\}$ and imagine that the ON-OFF processes $I_j, j \in B$ are running, *simultaneously*, very long ON periods.

Our previous discussion indicates that during this time the effective rate at which work enters the system is

$$\gamma_B = \sum_{j \in B} a_j + \sum_{j \notin B} a_j \frac{\mu_{\text{on}}^{(j)}}{\mu^{(j)}}. \tag{4.25}$$

The stability condition (4.10):

$$r > \sum_{j=1}^{k} a_j \frac{\mu_{\text{on}}^{(j)}}{\mu^{(j)}}$$

means that, on average, the server works at the rate higher than the input rate. That is, as long as the buffer is not empty, the buffer content has a negative drift: it tends to go down.

The right way to view the meaning of the assumption (4.15) is, then, as follows: there is a set $B \subset \{1, \ldots, k\}$ consisting of a single ON-OFF process whose ON distribution has the heaviest possible tail such that

$$\gamma_B > r. \tag{4.26}$$

That is, during a long ON period of that ON-OFF process *the direction of the drift in the system changes from negative to positive.*

If the assumption (4.15) does not hold, then either there is no B of cardinality 1 for which (4.26) holds, or, if such a B exists, the corresponding ON-OFF process does not have the heaviest possible tail of its ON distribution.

Recall that the large deviations approach tells us that our unlikely event of buffer overflow is likely to be caused by the smallest possible number of individual factors. Until now we have seen only example when this smallest possible number of factors is equal to one. When the assumption (4.15) fails, the smallest possible number of factors can be greater than one.

To get a better feeling about what is happening, let us consider a special case. Let $k = 2$, and assume that both ON-OFF sources are identical, in the sense that both have the same ON distribution F_{on}, the same OFF distribution F_{off}, same input rate a, such that

$$a\left(1 + \frac{\mu_{\text{on}}}{\mu}\right) \leq r < 2a.$$

In that case the singletons $B = \{1\}$ and $B = \{2\}$ for which $\gamma_B = a(1 + \mu_{\text{on}}/\mu)$ do not satisfy (4.26), but the set $B = \{1, 2\}$, for which $\gamma_B = 2a$ does.

When the two ON-OFF processes are simultaneously running long ON periods, the drift in the buffer content is equal to $2a - r > 0$. Therefore, the large deviations approach tells us that the overflow is likely to occur when both ON-OFF processes are running long ON periods whose common part is greater than $H/(2a - r)$.

In this case the smallest possible number of factors causing buffer overflow is 2: both ON-OFF sources have to run very long ON periods.

If we denote by T^*_H the first time such a long common ON time starts, we would expect that $ET_H \sim ET^*_H$ as $H \to \infty$. Let us try to figure out the order of magnitude of ET^*_H.

Let us view the situation as that of Bernoulli trials: every $(\overline{F_{\text{on}}}(H))^{-1}$ (order of magnitude) units of time we have one of the two ON-OFF processes start an ON period of order of magnitude H. During this ON period the other ON-OFF process has to start its ON period whose length is of order of magnitude H as well. The probability of this happening has order of magnitude $H\overline{F_{\text{on}}}(H)$, which we view as the success probability in our sequence of Bernoulli trials. Therefore, we expect to need about $(H\overline{F_{\text{on}}}(H))^{-1}$ trials before this happens. The conclusion is that

$$ET^*_H \approx H^{-1}(\overline{F_{\text{on}}}(H))^{-2} \text{ as } H \to \infty$$

in the sense of order of magnitude. Hence we expect that the same holds for the time until overflow:

$$ET_H \approx H^{-1}(\overline{F_{\text{on}}}(H))^{-2} \text{ as } H \to \infty \tag{4.27}$$

as well.

Note that we expect that ET_H is regularly varying at infinity with exponent $2\alpha - 1$, where α is the tail index of the ON distribution (common to the two ON-OFF processes).

Let us go back now to the general superposition of k ON-OFF processes. We assume that for any $j \in \{1, \ldots, k\}$ the ON distribution is regularly varying with exponent $\alpha_j \geq 1$. We are allowing $\alpha_j = \infty$ for some $j \in \{1, \ldots, k\}$, meaning distributions whose right tails are lighter than any regularly varying tail (exponential tails are allowed, for example).

Let

$$J_{\text{fin}} = \{j \in \{1, \ldots, k\} : \alpha_j < \infty\}. \tag{4.28}$$

We are assuming that $J_{\text{fin}} \neq \emptyset$.

Let apply the large deviations approach and the experience we gained while considering the simple case above to figure out the order of magnitude of the time until overflow.

Let $B \subset J_{\text{fin}}$. If the assumption (4.26) holds: $\gamma_B > r$, where we recall that γ_B is given by (4.25):

$$\gamma_B = \sum_{j\in B} a_j + \sum_{j\notin B} a_j \frac{\mu^{(j)}_{\text{on}}}{\mu^{(j)}},$$

then a long common ON period to the ON-OFF processes $(I_j,\, j \in B)$ of the length greater than $H/(\gamma_B - r)$ is likely to cause buffer overflow. Our previous discussion indicates that the expected time until this happens is regularly varying with exponent

$$1 + \sum_{j\in B}(\alpha_j - 1). \tag{4.29}$$

Therefore, we are looking for a subset $B \subset J_{\text{fin}}$ that will make the above event occur as soon as possible, which means that we are trying to minimize the expression in (4.29).

Consider, therefore, the optimization problem

$$(4.30) \qquad \min_{B\subset J_{\text{fin}}} \sum_{j\in B}(\alpha_j - 1)$$

$$\text{subject to} \quad \sum_{j\in B} a_j + \sum_{j\notin B} a_j \frac{\mu_{\text{on}}^{(j)}}{\mu^{(j)}} > r.$$

We assume that the set of the feasible solutions to the problem (4.30) is not empty.

This type of optimization problems was introduced by Zwart et al. (2000) in the context of **Problem 2** above (tail behaviour of the distribution of the stationary buffer content), who reformulate it as a *knapsack packing problem.*

Let κ_* be the optimal value of the cost function in the optimization problem above. Let also $B_1, \ldots, B_{k_*}$ be those subsets of J_{fin} on which this optimal value is achieved.

The large deviations approach suggests that the buffer overflow is likely to happen when all the ON-OFF processes in one of these sets run a long common ON period (whose length is of order H), and we have said that the expected time until this happens is regularly varying with exponent $1+\kappa_*$. Therefore,

$$(4.31) \qquad ET_H \ \text{ is regularly varying with exponent } 1+\kappa_* \text{ as } H\to\infty.$$

An interesting special case occurs when the input process consists of the superposition of k identical ON-OFF sources, again with the same ON distribution F_{on}, the same OFF distribution F_{off} and same input rate a. Assume that

$$(4.32) \qquad r < ka$$

(otherwise overflow cannot happen), and denote

$$(4.33) \qquad k_0 = \left\lceil \frac{r}{a}\frac{\mu}{\mu_{\text{off}}} - k\frac{\mu_{\text{on}}}{\mu_{\text{off}}} \right\rceil .$$

The assumption (4.32) guarantees that $k_0 \le k$.

In this case the constraint in the optimization problem (4.30) reduces to $\text{card}(B) \ge k_0$ and, hence, the optimal value of the cost function is achieved at the sets B with cardinality equal to k_0, and is equal to $k_0(\alpha-1)$. Therefore, (4.31) tells us that ET_H is regularly varying with exponent $1+k_0(\alpha-1)$ at infinity.

Note an interesting phase transition occurring here. Imagine that you are in charge of the system, and you can invest resources into getting a faster server, i.e. increasing the service rate r. You need to decide whether to do it or not.

Notice that k_0 above is a step function of the service rate r. Our discussion shows that if we increase the service rate in such a way that the parameter k_0 remains the same, there is no much gain, at least as far as the asymptotic behavior of the expected time until buffer overflow is concerned. If, however, the service rate r is increased in such a way that k_0 goes up, then there is a major gain in system performance.

One can do a similar analysis for the buffer overflow problem when the input a superposition of $M/G/\infty$ input processes, also using the large deviations approach. We will not go into the details of the argument again, leaving it instead to the interested readers to do it on their own.

We only mention a published result in the particular case of a superposition of k identical $M/G/\infty$ input processes. The k processes have a common Poisson arrival rate λ, a common session length distribution F with mean μ and regularly varying with exponent $\alpha \geq 1$ at infinity tail, and a common input rate $a > 0$.

Assume that

$$r > a\lambda\mu k$$

for stability of the system.

Here the crucial parameter is

$$k_0 = \left\lceil \frac{r}{a} - k\lambda\mu \right\rceil . \tag{4.34}$$

Then there is a finite constant $C \geq 1$ such that

$$C^{-1} a^{1+k_0(\alpha-1)} H \left(H\overline{F}(H) \right)^{-k_0} \leq ET_H \leq C a^{1+k_0(\alpha-1)} H \left(H\overline{F}(H) \right)^{-k_0} \tag{4.35}$$

for all H large enough; see Heath et al. (1999).

Note, once again, that ET_H is regularly varying with exponent $1 + k_0(\alpha - 1)$ at infinity, and that we have a phase transition similar to that in the case of the superposition of ON-OFF processes as the input.

Steady state buffer content problem

Here we consider in detail the case of the input process $(A(t),\, t \geq 0)$ being a superposition of k independent $M/G/\infty$ input processes. We assume the stability condition (4.14):

$$r > \sum_{j=1}^{k} a_j \lambda_j \mu_j .$$

We recall that here a_j, λ_j and μ_j are, correspondingly, the input rate, arrival rate and mean session time for the jth $M/G/\infty$ input process, $j = 1, \ldots, k$.

Furthermore, the $M/G/\infty$ input processes have session length distributions F_j, $j = 1, \ldots, k$, on which we will impose certain assumptions as we go along.

The key ingredient to working with the distribution of the stationary buffer content, which we will denote by $(W(t),\, t \geq 0)$ is the representation

$$W(t) = \sup_{u \leq t} \int_u^t \left(\sum_{j=1}^{k} a_j\, M_j(s) - r \right) ds,\; t \geq 0, \tag{4.36}$$

where we recall that for $j = 1, \ldots, k$, the process $(M_j(t),\, -\infty < t < \infty)$ is the stationary process describing the number of the open at time t sessions of the jth $M/G/\infty$ input process (alternatively, the number of customers in the system at time t in the jth $M/G/\infty$ queue). The processes $(M_j(t),\, -\infty < t < \infty)$, $j = 1, \ldots, k$ are independent.

Notice that if one is interested only in the one-dimensional marginal distribution of the buffer content, then one can use the representation

$$W(0) = \sup_{u\leq 0} \int_u^0 \left(\sum_{j=1}^k a_j\, M_j(s) - r \right) ds \tag{4.37}$$

$$= \sup_{-u\geq 0} \int_0^{-u} \left(\sum_{j=1}^k a_j\, M_j(-s) - r \right) ds \stackrel{\mathrm{d}}{=} \sup_{u\geq 0} \int_0^{u} \left(\sum_{j=1}^k a_j\, M_j(s) - r \right) ds$$

since the processes $(M_j(t),\, -\infty < t < \infty)$, $j = 1, \ldots, k$ are *reversible*:

$$(M_j(-t),\, -\infty < t < \infty) \stackrel{\mathrm{d}}{=} (M_j(t),\, -\infty < t < \infty)$$

$j = 1, \ldots, k$.

Representations of the type (4.36) and (4.37) have been widely used in a variety of queuing models; see for example Asmussen (1987), Prabhu (1998) or Whitt (1999).

Let us note the structure of the stochastic process $(M_j(t),\, -\infty < t < \infty)$ describing the number of customers in the system at time t in the jth $M/G/\infty$ queue. At any time t, $M_j(t)$ is a Poisson random variable with mean $\lambda_j\mu_j$ and, conditionally on $M_j(t) = k$, the remaining lengths of the k sessions present are independent and identically distributed with the common distribution

$$F_j^{(0)}(x) = \frac{1}{\mu_j} \int_0^x (1 - F_j(u))\, du,\ x \geq 0\,,$$

$j = 1, \ldots, k$.

Once again, with start with the same particular case as in the buffer overflow problem. Assume that there is a distribution F on $(0, \infty)$ with a regularly varying right tail with exponent $\alpha \geq 1$ and a subset J^* of $\{1, \ldots, k\}$ such that for every $j \in J^*$ the limit

$$p_j = \lim_{x\to\infty} \frac{\overline{F}_j(x)}{\overline{F}(x)} > 0$$

exists and is positive, while for every $j \notin J^*$

$$\lim_{x\to\infty} \frac{\overline{F}_j(x)}{\overline{F}(x)} = 0.$$

Assume, furthermore, that

$$\max_{j\in J^*} a_j > r - \sum_{i=1}^k a_i\lambda_i\mu_i. \tag{4.38}$$

As in the case of the buffer overflow problem we will use the logic of large deviations to figure out the asymptotic behavior of the tail probability

$$P\left(W(0) > w\right) = P\left(\sup_{u\geq 0} \int_0^u \left(\sum_{j=1}^k a_j\, M_j(s) - r \right) ds > w \right) \tag{4.39}$$

as $w \to \infty$. This is exactly the kind of situation one expects this approach to be particularly applicable. The stability condition (4.14) means that the process

$$\int_0^u \left(\sum_{j=1}^k a_j\, M_j(s) - r \right) ds, \;\; u \ge 0$$

has a negative drift. Hence, we are talking about a rare event.

We will look for a single factor that is most likely to cause this rare event to happen. It turns out that it is exactly the assumption (4.38) that causes the usual smallest number of factors that are most likely to cause the rare event, to be equal to 1.

Indeed, when one of the $M/G/\infty$ input processes with $j \in J^*$ runs a very long session, during that time the long session itself generates input at rate a_j, while we expect the "normal" additional input rate of $\sum_{i=1}^k a_i\lambda_i\mu_i$. If

$$a_j + \sum_{i=1}^k a_i\lambda_i\mu_i > r,$$

then during this time the process

$$\int_0^u \left(\sum_{j=1}^k a_j\, M_j(s) - r \right) ds, \;\; u \ge 0$$

has, actually, a positive drift of

$$\beta_j = a_j + \sum_{i=1}^k a_i\lambda_i\mu_i - r, \tag{4.40}$$

and, hence, can reach a high level w if the length of the long session is large enough.

Of course, the assumption (4.38) means that there is at least one of the $M/G/\infty$ input processes with $j \in J^*$ for which this scenario is feasible. Recall that the input processes for which $j \in J^*$ are the most likely to produce a very long session.

Now, for the jth $M/G/\infty$ input process, considered on $[0, \infty)$ we have to think both of the remainders of the sessions already present at time 0, and the new sessions that start after time 0. Note that the remainders of the sessions present at time 0 are distributed according to the life time distribution $F_j^{(0)}$, and this distribution has regularly varying tails with exponent $\alpha - 1$, while the new sessions that start after time 0 are distributed according to F_j, and this distribution has regularly varying tails with exponent α. Hence we need to pay special attention to the remainders of the sessions present at time 0.

To separate the effect of the remainders of the initial sessions and that of the newly arriving sessions, let us denote by $(M_j^0(t),\, 0 \le t < \infty)$, $j = 1, \ldots, k$ independent stochastic processes describing the number of customers in the system at time t in the jth $M/G/\infty$ queue, $j = 1, \ldots, k$, *that starts empty at time* 0.

Let us denote, as before, by J_+^* the set of all $j \in J^*$ for which (4.38) holds. Then appealing to the logic of heavy tailed large deviations, we see that

$$P\left(W(0) > w\right) = P\left(\sup_{u\geq 0}\int_0^u \left(\sum_{j=1}^k a_j\, M_j(s) - r\right)\, ds > w\right)$$

$$\sim P\Big(\text{for some } j \in J_+^* \text{ one of the customers present at time } 0 \tag{4.41}$$

$$\text{in the } j\text{th } M/G/\infty \text{ queue has more than } \tfrac{w}{\beta_j} \text{ units time left to be in the system}\Big)$$

$$+P\left(\sup_{u\geq 0}\int_0^u \left(\sum_{j=1}^k a_j\, M_j^0(s) - r\right)\, ds > w\right) := P_{\text{rem}}(w) + P_{\text{new}}(w).$$

Recalling that we do not expect more than one session present at time zero to have a very long remaining lifetime, we see that

$$P_{\text{rem}}(w) \sim \sum_{j\in J_+^*} P\big(\text{one of the customers present at time 0 in the } j\text{th } M/G/\infty \text{ queue}$$

$$\text{has more than } \tfrac{w}{\beta_j} \text{ units of time left to be in the system}\Big) := \sum_{j\in J_+^*} P_{\text{rem},j}(w). \tag{4.42}$$

For every $j \in J_+^*$ we have

$$P_{\text{rem},j}(w) = \sum_{i=1}^{\infty} P\left(M_j(0) = i\right) P\big(\text{one of the } i \text{ remaining sessions}$$

$$\text{has more than } \tfrac{w}{\beta_j} \text{ units of time left to be in the system}\Big). \tag{4.43}$$

Now, $M_j(0)$ has the Poisson distribution with mean $\lambda_j\mu_j$, and so by (4.43) we have

$$P_{\text{rem},j}(w) \sim \sum_{i=1}^{\infty} P\left(M_j(0) = i\right) iP\big(\text{ a given remaining sessions has more than}$$

$$\tfrac{w}{\beta_j} \text{ units of time left to be in the system}\Big)$$

$$= EM_j(0)\overline{F_j^{(0)}}\left(\frac{w}{\beta_j}\right) \sim \lambda_j\mu_j\,\frac{1}{\mu_j}\,\frac{1}{\alpha-1}\,\frac{w}{\beta_j}\,\overline{F}_j\left(\frac{w}{\beta_j}\right) \sim p_j\lambda_j\beta_j^{\alpha-1}\frac{1}{\alpha-1}w\overline{F}(w)$$

as $w \to \infty$. Substituting into (4.42) we obtain

$$P_{\text{rem}}(w) \sim \left(\sum_{j\in J_+^*} p_j\lambda_j\beta_j^{\alpha-1}\right)\frac{1}{\alpha-1}w\overline{F}(w) \tag{4.44}$$

as $w \to \infty$.

It remains to deal with the second term in (4.41), $P_{\text{new}}(w)$, which is the probability of the unlikely event

$$\left\{ \sup_{u\geq 0} \int_0^u \left(\sum_{j=1}^k a_j\, M_j^0(s) - r \right) ds > w \right\}. \tag{4.45}$$

We already know that the logic of large deviations tells us that this event, if it occurs at all, is likely to be caused by a single very long session during the duration of which the drift in the system is positive and it drifts upwards to the level w. Such very long session is most likely to occur in one $M/G/\infty$ input processes with $j \in J_+^*$.

Let D be the duration of such a session, and T the time this session arrives. Note that the duration D has to be very long, hence the arrival time T has to be large as well. As the result, by the time the very long session arrives, the process

$$\int_0^u \left(\sum_{j=1}^k a_j\, M_j^0(s) - r \right) ds, \ u \geq 0$$

is likely to be at the negative level of about

$$-T \left(r - \sum_{j=1}^k a_j \lambda_j \mu_j \right).$$

Suppose that the long session arriving at the time T belongs to the jth $M/G/\infty$ input process with $j \in J_+^*$. Then during that long session the system experiences a temporary positive drift of β_j. Hence, the duration D of the long session is sufficient to cause the event in (4.45) to occur if

$$D\beta_j > w + T \left(r - \sum_{i=1}^k a_i \lambda_i \mu_i \right).$$

Therefore, we expect that

$$P_{\text{new}}(w) \sim P \Bigg(\bigcup_{j\in J_+^*} \{ \ j\text{th } M/G/\infty \text{ input process has a session of length } D$$

$$\text{arriving at time } T \text{ such that } \ D\beta_j > w + T \left(r - \sum_{i=1}^k a_i \lambda_i \mu_i \right) \Big\} \Bigg) := P \left(\bigcup_{j\in J_+^*} A_j \right).$$

Since the events A_j above are unlikely, we have

$$P_{\text{new}}(w) \sim \sum_{j\in J_+^*} P(A_j). \tag{4.46}$$

Observe that for the jth $M/G/\infty$ input process the pairs (T, D) of the times the sessions are initiated and of their lengths form a Poisson random measure M_* on $\mathbb{R}_+^2$ with mean measure

$$m(dt, dx) = \lambda_j \, dt \, F_j(dx).$$

Therefore, for $j \in J_+^*$,

$$P(A_j) = P\left(M_*(B_j) > 0\right),$$

where

$$B_j = \left\{(t,x) \in \mathbb{R}_+^2 : x\beta_j > w + t\left(r - \sum_{i=1}^k a_i\lambda_i\mu_i\right)\right\}.$$

It is easy to compute the measure m of the set B_j. We conclude that

$$P(A_j) = 1 - e^{-m(B_j)} \sim m(B_j)$$

$$= \frac{\lambda_j\beta_j}{r - \sum_{i=1}^k a_i\lambda_i\mu_i}\int_{w/\beta_j}^{\infty} \overline{F}_j(x)\,dx \sim \frac{p_j\lambda_j\beta_j^{\alpha}}{r - \sum_{i=1}^k a_i\lambda_i\mu_i}\,\frac{1}{\alpha-1} w\overline{F}(w) \tag{4.47}$$

as $w \to \infty$. Substituting the result of (4.47) into (4.46) we conclude that

$$P_{\text{new}}(w) \sim \left(\sum_{j\in J_+^*} p_j\lambda_j\beta_j^{\alpha}\right)\frac{1}{r - \sum_{i=1}^k a_i\lambda_i\mu_i}\,\frac{1}{\alpha-1} w\overline{F}(w) \tag{4.48}$$

as $w \to \infty$.

Now that we know the asymptotic behavior of both $P_{\text{rem}}(w)$ and $P_{\text{new}}(w)$, we can substitute both expressions into (4.41) and obtain, after some easy algebra,

$$P\left(W(0) > w\right) \sim \left(\sum_{j\in J_+^*} p_j a_j\lambda_j\beta_j^{\alpha-1}\right)\frac{1}{r - \sum_{i=1}^k a_i\lambda_i\mu_i}\,\frac{1}{\alpha-1} w\overline{F}(w) \tag{4.49}$$

as $w \to \infty$. That is, the tail of the steady state buffer content is regularly varying with exponent $\alpha - 1$.

In the particular case of homogeneous input $k = 1$ (4.49) has been proved in Resnick and Samorodnitsky (2001).

Let us briefly review what happens in the case when the assumption (4.38)

$$\max_{j\in J^*} a_j > r - \sum_{i=1}^k a_i\lambda_i\mu_i$$

does not hold. We know, from our discussion of the buffer overflow problem, that, in this case, the most likely way for the unlikely event

$$A = \{W(0) > w\} \tag{4.50}$$

to occur may not be due to a single very long session but, rather, due to several long sessions running at the same time.

To see what is the most likely way for this to happen, suppose that, at some point, the k $M/G/\infty$ input processes start running, simultaneously, $n_1, n_2, \ldots, n_k$ long sessions, for some $n_j = 0, 1, 2\ldots, j = 1, \ldots, k$. Let us call the collection $(n_1, \ldots, n_k)$ a *configuration*.

For a given configuration $(n_1, \ldots, n_k)$ to be able to cause the unlikely event A in (4.50), the drift during the long common part of the sessions in that configuration should be positive. Our usual arguments tells us that this drift is equal to

$$\beta_{n_1,\ldots,n_k} = \sum_{j=1}^{k} a_j \left(n_j + \lambda_j \mu_j\right) - r. \tag{4.51}$$

We need, therefore, for the jth $M/G/\infty$ input process to run n_j sufficiently long sessions at the same time (the length of the session is most likely to have order of magnitude ow w) and early enough to make sure that the event A in (4.50) happens. This last requirement says that these should either be the remainders of the sessions present at time zero, or the sessions, or sessions arriving within about linear in w time interval after time zero.

Arguments similar to the one used above indicate that, the probability for this to happen has the same order of magnitude as

$$\left(w\overline{F}_j(w)\right)^{n_j}.$$

Therefore, the probability for a configuration $(n_1, \ldots, n_k)$ satisfying (4.51) to occur has the same order of magnitude as

$$\prod_{j=1}^{k} \left(w\overline{F}_j(w)\right)^{n_j}. \tag{4.52}$$

Observe that the expression in (4.52) is regularly varying at infinity with exponent

$$-\sum_{j=1}^{k} n_j(\alpha_j - 1).$$

Therefore, the configuration the most likely to cause the event A in (4.50) to happen is the one for which this exponent of the regular variation is the largest.

Therefore, we are led to the following optimization problem:

$$\min_{\substack{n_j=0,1,\ldots \\ j=1\ldots,k}} \sum_{j=1}^{k} n_j(\alpha_j - 1) \tag{4.53}$$

$$\text{subject to } \sum_{j=1}^{k} a_j \left(n_j + \lambda_j \mu_j\right) - r > 0.$$

As before, we are allowing $\alpha_j = \infty$ for some $j = 1, \ldots, k$, indicating that the corresponding distributions have right tails that are lighter than any regularly varying tails (e.g. exponentially fast decaying tails). Note that the set of feasible solutions to the problem (4.53) is always non-empty.

Let κ_* be the optimal value of the cost function in this optimization problem. Then we expect that

$$P\left(W(0) > w\right) \quad \text{is regularly varying with exponent } -\kappa_* \tag{4.54}$$

as $w \to \infty$.

In fact, this statement was proved in Borst and Zwart (2001), who introduced the above optimization problem, and called it a knapsack packing problem as well. The above paper also provides a certain asymptotic expression for the tail probability $P(W(0) > w)$.

Finally, in the case of the input being a superposition of ON-OFF processes, one can use similar arguments to arrive, once again,e to the optimization problem (4.30):

$$\min_{B \subset J_{\text{fin}}} \sum_{j \in B} (\alpha_j - 1)$$

$$\text{subject to} \quad \sum_{j \in B} a_j + \sum_{j \notin B} a_j \frac{\mu_{\text{on}}^{(j)}}{\mu^{(j)}} > r$$

(recall that $J_{\text{fin}} = \{j \in \{1, \ldots, k\} : \alpha_j < \infty\}$) and then we expect that

$$P\left(W(0) > w\right) \quad \text{is regularly varying with exponent } -\kappa_*$$

as $w \to \infty$, where we assume that the optimization problem has a feasible solution, and κ_* is the optimal value of the cost function. This result has been established by Zwart et al. (2000).

5. Rare events and long range dependence

We switch now to a discussion of long range dependence in the context of how rare events happen. This will be done using the large deviations approach: we would like to understand how memory in the process determines which configurations of possible factors may causes certain unlikely events to happen.

We will concentrate on two major classes of heavy tailed stochastic models: *moving average processes* and *infinitely divisible processes*. For these two classes of models we can identify the different factors affecting rare events.

5.1. **Moving average processes.** These are stochastic processes in discrete time defined by

$$X_n = \mu + \sum_{j=-\infty}^{\infty} \varphi_{n-j}\, \varepsilon_j\,, \quad n = 0, 1, \ldots \tag{5.1}$$

(note that we are considering two-sided moving averages). Here $(\varepsilon_n,\, n = \ldots, -1, 0, 1, 2, \ldots)$ (*the noise variables*) are iid random variables, and μ is a constant.

If the random variables (ε_n) have a finite mean, we will assume that the mean is equal to zero (it is simply incorporated in the constant μ).

Depending on the law of the noise variables we will, clearly, need to impose certain requirements on the coefficients (φ_n) in (5.1) for the process to be well defined.

We assume that the random variables (ε_n) have heavy tails, and our specific assumption is that of regular variation and tail balance: for $\varepsilon = \varepsilon_0$

$$P(|\varepsilon| > \lambda) \quad \text{is regularly varying with exponent } \alpha > 0, \tag{5.2}$$

and

$$\lim_{\lambda\to\infty} \frac{P(\varepsilon > \lambda)}{P(|\varepsilon| > \lambda)} = p, \quad \lim_{\lambda\to\infty} \frac{P(\varepsilon < -\lambda)}{P(|\varepsilon| > \lambda)} = q\,. \tag{5.3}$$

Here $p, q \geq 0$ and $p + q = 1$. We will see below that (under appropriate assumptions on the coefficients) this implies that the tails of the stochastic process (X_n) in (5.1) have the same order of magnitude.

We will assume that the coefficients (φ_n) in (5.1) satisfy the following assumptions, sufficient for the series in (5.1) to converge.

If $\alpha > 2$, we will assume that

$$\sum_{j=-\infty}^{\infty} \varphi_j^2 < \infty\,, \tag{5.4}$$

whereas if $0 < \alpha \leq 2$, we will assume that

$$\sum_{j=-\infty}^{\infty} |\varphi_j|^{\alpha-\epsilon} < \infty \tag{5.5}$$

for some $\epsilon > 0$.

In the case $\alpha > 2$ it is easy to check, using the three-series theorem, that condition (5.4) is also necessary for convergence of the series in (5.1). In the case $0 < \alpha \leq 2$ condition (5.4) is not necessary for convergence. However, absent information on the slowly varying function in the regular variation of the tail of $|\varepsilon|$, this condition is as good as one can hope for.

Under the above conditions on the coefficients (φ_n) the moving average stochastic process (X_n) defined in (5.1) is, obviously, a stationary process. Moreover, it is heavy tailed, and the tail of X_1 turns out to be the same, up to a multiplicative constant, as the tail of the noise variable $|\varepsilon|$:

$$\lim_{\lambda\to\infty} \frac{P(X_1 > \lambda)}{P(|\varepsilon| > \lambda)} = \sum_{j=-\infty}^{\infty} |\varphi_j|^{\alpha} \big[p\, \mathbf{1}_{\{\varphi_j > 0\}} + q\, \mathbf{1}_{\{\varphi_j < 0\}}\big]\,. \tag{5.6}$$

See Mikosch and Samorodnitsky (2000b). It is also a mixing, hence ergodic, process. See Rosenblatt (1962).

The most familiar classes of moving average processes are, of course, the classical ARMA processes. These are especially popular in the case of finite second moments. According to (5.6) this is guaranteed if $\alpha > 2$.

Assuming finite variances, it is easy to see that

$$\mathrm{Cov}\,(X_n, X_0) = \left(\sum_{j=-\infty}^{\infty} \varphi_j \varphi_{n+j}\right) \mathrm{Var}(\varepsilon)\,, \tag{5.7}$$

and for the traditional ARMA models covariances decay exponentially fast; see e.g. Brockwell and Davis (1991). Notice that the condition

$$\sum_{j=-\infty}^{\infty} |\varphi_j| < \infty \tag{5.8}$$

guarantees absolute summability of correlations:

$$\sum_{n=0}^{\infty} |\mathrm{Cov}\,(X_n, X_0)\,| < \infty.$$

Historically, those wishing to model long range dependence, while staying "not too far" from ARMA processes, used the so called *fractionally differenced ARMA models.*

These are parametric models, the crucial parameter of which is the parameter of fractional differencing $d \in (-1/2, 1/2)$. Except for the case $d = 0$ which corresponds to no fractional differencing, one has

$$\mathrm{Cov}\,(X_n, X_0) \sim \mathrm{const}\, n^{-(1-2d)} \text{ as } n \to \infty.$$

Hence, it is common to refer to fractionally differenced ARMA processes with $0 < d < 1/2$ as long range dependent. In this case correlations are not absolutely summable, and are regularly varying at infinity with exponent $2d - 1 > -1$.

Our approach to deciding which moving average processes (5.1) have long memory and which have short memory, is not based on correlations. Rather, we will be looking for *phase transitions* as one moves from one part of the parameter space to the other one.

In this case the parameter space Θ consists of the sequences $(\varphi_j\,, j = \ldots, -1, 0, 1, \ldots)$ of coefficients satisfying (5.4) or (5.5). We view the noise variables distribution or, rather, the tail parameter $\alpha > 0$ of the noise variables, as being fixed, rather than as a part of the parameter space. The reason is that changes in the tail parameter mostly affect the heaviness of the tail of the moving average process (X_n), while we would like to concentrate on the length of memory *with given tails.*

Similarly, we will also view the shift μ in (5.1) for the obvious reason that it does not affect the memory of the process.

Assume that $\alpha > 1$ and consider the subset Θ_0 of the parameter space consisting of absolutely summable coefficients:

$$\Theta_0 = \Big\{(\varphi_j) \in \Theta : \ \ \textstyle\sum_j |\varphi_j| < \infty.\Big\} \tag{5.9}$$

As we will see shortly, important things happen when one crosses the boundary between $\Theta_1 = \Theta \setminus \Theta_0$, i.e. when the coefficients in the moving average process stop being absolutely summable. The changes are so dramatic that they qualify as "a phase transition".

Hence, we will argue that the coefficients in the set Θ_1 in (5.9) correspond to long memory processes, while the coefficients in $\Theta_0 = \Theta \setminus \Theta_1$ correspond to short memory processes.

We have seen above that, if the process has a finite variance, then the coefficients in Θ_0 also guarantee absolutely summable correlations. However, we will not concentrate on correlations and, moreover, our analysis will apply also in the cases when the variance does not exist.

5.2. **Long strange segments.** Here is our first test problem on which we will look for significant changes as the memory of the process changes.

Let $(X_1, X_2, \ldots)$ be a stationary and ergodic stochastic process with a finite mean μ (which may or may not a moving average process). For a $\theta > \mu$ we define

$$R_n(\theta) = \sup \left\{ j - i : \, 0 \le i < j \le n, \, \frac{X_{i+1} + \ldots + X_j}{j - i} > \theta \right\} , \tag{5.10}$$

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

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

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

should be about the mean μ. Since there are many different intervals of this type between 1 and n for large n, one would expect that over some of them this average may exceed a given $\theta > \mu$. The statistic $R_n(\theta)$ gives the length of the longest interval over which this happens.

Long intervals whose length statistic $R_n(\theta)$ is intended to measure are those on which the law of large numbers appears to break down. Hence we refer to them as *long strange intervals.*

If X_n refers to, say, the amount of work arriving in a service station during nth time period, then during long strange intervals the system appears to be running under load higher than the nominal load μ. Therefore, the longer are the long strange intervals, the worst we expect system performance to be.

This description indicates that such time intervals can be of a crucial importance in manufacturing and insurance applications. Functionals like $R_n(\theta)$ are also important in finance, comparative analysis of DNA sequences and analysis of computer search algorithms.

In the case of light tailed processes (X_n) and short or no memory, long strange intervals have been considered before; see for example Dembo and Zeitouni (1993). Typically, $R_n(\theta)$ grows as $c \log n$ for some $c > 0$ (if the process has exponentially fast decaying tails).

The fact that heaviness of the tails of the process (X_n) affects the length of the long strange intervals is quite intuitive; heavy tailed values of X_n make it more likely that the average over an intervals is away from the theoretical mean. We will see this quantified below. We will also see the effect of memory on the length of long strange intervals.

We will study the long strange intervals in the case when the process (X_n) is a moving average process (5.1). We are assuming now that the noise variables (ε_j) have a finite mean, which is equal to zero. Specifically, we will assume that $\alpha \ge 1$. Then $EX_1 = \mu$.

The very name *long strange intervals* implies that we are dealing with rare events. Hence we expect that the large deviations approach may be useful here.

We start with noting that it follows directly from the definition of $R_n(\theta)$ that for any $1 \leq m \leq n$

$$R_n(\theta) \geq m \quad \text{if and only if } X_{i+1} + \ldots + X_{i+k} > k\theta \tag{5.11}$$

$$\text{for some } k = m, m+1, \ldots, n, \text{ and some } \; i = 0, \ldots, n-k\,.$$

By the definition (5.1) of the moving average process we know that

$$X_{i+1} + \ldots + X_{i+k} = k\mu + \sum_{j=-\infty}^{\infty} \left(\sum_{d=i+1-j}^{i+k-j} \varphi_d \right) \varepsilon_j\,. \tag{5.12}$$

The logic of large deviations tells us that, for a large k, the unlikely event

$$X_{i+1} + \ldots + X_{i+k} > k\theta$$

is likely to be caused by a single large positive or negative value of a noise variable. The coefficients in (5.12) above will determine how large this value has to be.

That is, we expect that

$$P(R_n(\theta) \geq m) \sim P\left(\left(\sum_{d=i+1-j}^{i+k-j} \varphi_d \right) \varepsilon_j > k(\theta - \mu) \;\text{ for some } \; j = \ldots, -1, 0, 1, \ldots, \right. \tag{5.13}$$

$$\left. \text{some } \; k = m, m+1, \ldots, n \text{and some } \; i = 0, \ldots, n-k \right).$$

To obtain a meaningful result out of (5.13) we will let the number m increase to infinity as a function of the sample size n. For the "right" choice of $m = m(n)$ we obtain a non-degenerate limit in (5.13), and then we can say that the length $R_n(\theta)$ of the long strange intervals grows of $m(n)$ with n.

Suppose, first of all, that the parameters of our process are in the set Θ_0 of absolutely summable coefficients.

We will see what is "right" choice of $m = m(n)$ as we go along, but whatever it may be, we expect that

$$m(n) \to \infty, \quad \frac{m(n)}{n} \to 0 \quad \text{as } n \to \infty.$$

It follows from (5.13) that

$$P(R_n(\theta) < m(n)) \tag{5.14}$$

$$\sim \prod_{j=-\infty}^{\infty} P\left(\left(\sum_{d=i+1-j}^{i+k-j} \varphi_d \right) \varepsilon_j \leq k(\theta - \mu) \;\text{ for all } k = m(n), \ldots, n \text{ and } i = 0, \ldots, n-k. \right)$$

Recall that we are talking about "strange" intervals, and so for a given initial point of an interval, it is easier for the average over the interval to be above $\theta > \mu$ if the interval is shorter. Hence, we expect that $k = m(n)$ will make this event the most likely. Hence, using $k = m(n)$ in (5.14) we see that

$$P(R_n(\theta) < m(n)) \tag{5.15}$$

$$\sim \prod_{j=-\infty}^{\infty} P\left(\left(\sum_{d=i+1-j}^{i+m(n)-j} \varphi_d\right) \varepsilon_j \leq m(n)(\theta-\mu) \text{ for all } i=0,\ldots,n-m(n)\text{-}\right)$$

as $n \to \infty$. Remembering that the coefficients may be either positive or negative, we see that the event in the probability above imposes restrictions on ε_j both from above and from below: we need

$$\varepsilon_j \leq (\theta-\mu)\left(\frac{1}{m(n)} \sup_{i=0,\ldots,n-m(n)} \left(\sum_{d=i+1-j}^{i+m(n)-j} \varphi_d\right)_+\right)^{-1} := b_+(j,n) \tag{5.16}$$

and

$$\varepsilon_j \geq -(\theta-\mu)\left(\frac{1}{m(n)} \sup_{i=0,\ldots,n-m(n)} \left(\sum_{d=i+1-j}^{i+m(n)-j} \varphi_d\right)_-\right)^{-1} := -b_-(j,n)\,. \tag{5.17}$$

Therefore,

$$P(R_n(\theta) < m(n)) \sim \prod_{j=-\infty}^{\infty} P\left(-b_-(j,n) \leq \varepsilon_j \leq b_+(j,n)\right)\,. \tag{5.18}$$

Since $m(n)$ grows to infinity, we expect that the individual terms in the infinite product in (5.18) converge to 1 as $n \to \infty$. In general, a little thought about the structure of $b_+(j,n)$ and $b_-(j,n)$ above shows that only a particular range of j is likely to affect significantly the product in (5.18).

Denote

$$M_{+,1}(\varphi) = \sup_{-\infty<k<\infty} \left(\sum_{j=-\infty}^{k} \varphi_j\right)_+, \quad M_{+,2}(\varphi) = \sup_{-\infty<k<\infty} \left(\sum_{j=k}^{\infty} \varphi_j\right)_+,$$

$$M_{-,1}(\varphi) = \sup_{-\infty<k<\infty} \left(\sum_{j=-\infty}^{k} \varphi_j\right)_-, \quad M_{-,2}(\varphi) = \sup_{-\infty<k<\infty} \left(\sum_{j=k}^{\infty} \varphi_j\right)_-.$$

Then for j in the range $m \leq j \leq m+n$, some integer m,

$$b_+(j,n) \sim (\theta-\mu)m(n)\left(\max\left(M_{+,1}(\varphi), M_{+,2}(\varphi)\right)\right)^{-1}$$

and

$$b_-(j,n) \sim (\theta-\mu)m(n)\left(\max\left(M_{-,1}(\varphi), M_{-,2}(\varphi)\right)\right)^{-1}.$$

This turns out to be the only important range of j.

Let

$$M_+(\varphi) = \max\left(M_{+,1}(\varphi), M_{+,2}(\varphi)\right)$$

and

$$M_-(\varphi) = \max\left(M_{-,1}(\varphi), M_{-,2}(\varphi)\right)\,.$$

Then we expect that

$$P(R_n(\theta) < m(n)) \sim \prod_{j=m}^{m+n} P\Big(-(\theta-\mu)m(n)M_-(\varphi)^{-1} \le \varepsilon_j \le (\theta-\mu)m(n)M_+(\varphi)^{-1}\Big)$$

$$= \Big(P\left(-(\theta-\mu)m(n)M_-(\varphi)^{-1} \le \varepsilon_1 \le (\theta-\mu)m(n)M_+(\varphi)^{-1}\right) \Big)^{n+1} . \tag{5.19}$$

Now we can figure out what is the right choice for $m = m(n)$ above. Writing in (5.19)

$$P(R_n(\theta) < m(n)) \tag{5.20}$$

$$\sim \Big(1 - P\left(\varepsilon_1 < -(\theta-\mu)m(n)M_-(\varphi)^{-1}\right) - P\left(\varepsilon_1 > (\theta-\mu)m(n)M_+(\varphi)^{-1}\right) \Big)^{n+1}$$

we see that one should define $m(n)$ in such a way that

$$P\left(\varepsilon_1 < -m(n)\right) \approx P\left(\varepsilon_1 > m(n)\right) \approx \frac{1}{n}. \tag{5.21}$$

Let F be the distribution function of $|\varepsilon|$. For $n \ge 1$ define

$$a_n = \left(\frac{1}{\overline{F}}\right)^{\leftarrow}(n)\,. \tag{5.22}$$

Here, for a nondecreasing function U, we use the notation $U^{\leftarrow}$ to denote the left continuous inverse of U

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

Since $1/\overline{F}$ is regularly varying at infinity with exponent α, we immediately conclude that (a_n) is regularly varying at infinity with exponent $1/\alpha$ and

$$\lim_{n\to\infty} n\overline{F}(a_n) = 1\,. \tag{5.23}$$

For $x > 0$ let $m(n) = xa_n$, $n \ge 1$. Because of the regular variation and tail balance conditions (5.2) and (5.2) we see that

$$P\Big(\varepsilon_1 < -(\theta-\mu)\,m(n)\,M_-(\varphi)^{-1}\Big) = P\Big(\varepsilon_1 < -(\theta-\mu)\,xa_n\,M_-(\varphi)^{-1}\Big)$$

$$\sim \left((\theta-\mu)\,x\,M_-(\varphi)^{-1}\right)^{-\alpha} P\left(\varepsilon_1 < -a_n\right) \sim \left(q\,(\theta-\mu)^{-\alpha}\,M_-(\varphi)^{\alpha}\,x^{-\alpha}\right)\frac{1}{n}$$

and

$$P\Big(\varepsilon_1 > (\theta-\mu)\,m(n)\,M_-(\varphi)^{-1}\Big)$$

$$\sim \left((\theta-\mu)\,x\,M_+(\varphi)^{-1}\right)^{-\alpha} P\left(\varepsilon_1 > a_n\right) \sim \left(p\,(\theta-\mu)^{-\alpha}\,M_+(\varphi)^{\alpha}\,x^{-\alpha}\right)\frac{1}{n}$$

as $n \to \infty$. Substituting the above expressions into (5.20) we conclude that for any $x > 0$

$$P(R_n(\theta) < xa_n)) \to \exp\Big\{ -(\theta-\mu)^{-\alpha}\left(pM_+(\varphi)^{\alpha} + qM_-(\varphi)^{\alpha}\right)x^{-\alpha}\Big\} \tag{5.24}$$

as $n \to \infty$.

Therefore, if the parameters of our process are in the set Θ_0 of absolutely summable coefficients,

$$\frac{R_n(\theta)}{a_n} \Rightarrow (\theta-\mu)^{-1}\left(pM_+(\varphi)^\alpha + qM_-(\varphi)^\alpha\right)^{1/\alpha} Z_\alpha \tag{5.25}$$

for every $\theta > \mu$. Here Z_α has the standard extreme value distribution of Φ_α type:

$$P(Z_\alpha \le z) = \exp\left\{-z^{-\alpha}\right\}, \ z > 0. \tag{5.26}$$

The result (5.25) is due to Mansfield et al. (1999).

The conclusion is that, if the parameters of our process are in the set Θ_0 of absolutely summable coefficients, then the length of the long strange intervals grows about as a_n (a regularly varying sequence with exponent $1/\alpha$).

Notice that the order of magnitude of $R_n(\theta)$ does not change as the parameters vary within Θ_0 (the multiplicative constant in (5.25) does change).

We discuss next what happens when the parameters of the moving average process (5.1) cross the boundary between Θ_0 and Θ_1, the set where the coefficients are not absolutely summable.

Qualitatively, we expect that the behavior of the long strange intervals will be different in that case. For example, if the coefficients are not summable, then the quantities $M_{+,1}(\varphi)$, $M_{+,2}(\varphi)$, $M_{-,1}(\varphi)$ and $M_{-,2}(\varphi)$ may not be finite (this would be the case if, for example, the coefficients are of a constant sign), and so the multiplicative constant in (5.25) may become infinite.

To quantify the order of magnitude of $R_n(\theta)$ when the parameters are in the set Θ_1, we will need to look at a more concrete structure of the coefficients (φ) of the moving average process.

Specifically, we will assume that the coefficients (φ_j) are regularly varying and balanced. That is, there is a function $\varphi: [0,\infty) \to [0,\infty)$ such that

$$\varphi(t) = L_2(t)\, t^{-h}, \tag{5.27}$$

$0 < h < 1$, as $t \to \infty$ and such that

$$\lim_{j\to\infty} \frac{\varphi_j}{\varphi(j)} = \xi_+, \quad \lim_{j\to\infty} \frac{\varphi_{-j}}{\varphi(j)} = \xi_-, \tag{5.28}$$

for some $\xi_+, \xi_- \ge 0$, at least one of which is positive.

Here

$$h > \max\left\{\frac{1}{\alpha}, \frac{1}{2}\right\} \tag{5.29}$$

and L_2 is a slowly varying function.

Note that the conditions (5.4) - (5.5) for convergence of the moving average process still hold and, in particular, the tail behavior of the process remains the same: the distribution of X_1 has regularly varying right tail with exponent α; see (5.6). The length of the memory of the process has changed, however.

Now, the relation (5.18)

$$P(R_n(\theta) < m(n)) \sim \prod_{j=-\infty}^{\infty} P\left(-b_-(j,n) \le \varepsilon_j \le b_+(j,n)\right)$$

still holds, but the behavior of the numbers $(b_+(j,n))$ and $(b_-(j,n))$ is different now.

The regular variation of the coefficients means that, for all j large enough, all φ_j are of the same sign (at least if $\xi_+ > 0$), and similarly with the negative j.

As before, we do not expect that individual terms in the product above have a significant effect. We may assume, therefore, that all the coefficients are of the same sign, and we assume, for example, that the coefficients are all nonnegative.

In that case $b_-(j,n) = \infty$ for all j and n and the expression (5.18) becomes

$$P(R_n(\theta) < m(n)) \sim \prod_{j=-\infty}^{\infty} P\left(\varepsilon_j \le b_+(j,n)\right) . \tag{5.30}$$

Denote

$$\rho = \frac{\xi_-^{1/h}}{\xi_+^{1/h} + \xi_-^{1/h}} . \tag{5.31}$$

Arguments similar to the ones we used in the case when the parameters of the process belonged to the set Θ_0 give us that the important range of j in the infinite product in (5.31) is that between $\rho m(n)$ and $n - (1-\rho)m(n)$. Therefore (not worrying about integer values), we expect that

$$P(R_n(\theta) < m(n)) \sim \prod_{j=\rho m(n)}^{n-(1-\rho)m(n)} P\left(\varepsilon_j \le b_+(j,n)\right) . \tag{5.32}$$

The next step is to understand the structure of the numbers $(b_+(j,n))$ here. Since the coefficients (φ_m) converge to zero as $m \to \infty$ and $m \to -\infty$, we expect that the average in $b_+(j,n)$,

$$\sup_{i=0,\dots,n-m(n)} \frac{1}{m(n)} \left(\sum_{d=i+1-j}^{i+m(n)-j} \varphi_d \right)$$

is the biggest when the sum is "hugging zero". Hence, we expect that the maximum is achieved when $i - j \approx -a\, m(n)$, some $0 \le a \le 1$. In that case,

$$\begin{aligned}
\sum_{d=i+1-j}^{i+m(n)-j} \varphi_d &\sim \sum_{d=1-am(n)}^{(1-a)m(n)} \varphi_d = \sum_{d=0}^{(1-a)m(n)} \varphi_d + \sum_{d=1}^{am(n)-1} \varphi_{-d} \\
&\sim \xi_+ \int_0^{(1-a)m(n)} \varphi(t)\,dt + \xi_- \int_0^{am(n)} \varphi(t)\,dt \\
&\sim \xi_+(1-h)^{-1}(1-a)m(n)\,\varphi\left((1-a)m(n)\right) + \xi_-(1-h)^{-1}\,a\,m(n)\,\varphi\left(a\,m(n)\right) \\
&\sim (1-h)^{-1}\left(\xi_+(1-a)^{1-h} + \xi_- a^{1-h}\right) m(n)\varphi\left(m(n)\right) .
\end{aligned}$$

It is trivial to check that the expression

$$\xi_+(1-a)^{1-h} + \xi_- a^{1-h}$$

is maximized when $a = \rho$. Hence we expect that in our range of j

$$b_+(j,n) \approx (\theta-\mu)\left(\frac{1}{m(n)}\sum_{d=0}^{(1-\rho)m(n)} \varphi_d + \sum_{d=1}^{\rho m(n)-1} \varphi_{-d}\right)$$

$$\sim (\theta-\mu)\frac{1}{m(n)}(1-h)^{-1}\left(\xi_+(1-\rho)^{1-h} + \xi_-\rho^{1-h}\right) m(n)\varphi\left(m(n)\right)$$

$$= (\theta-\mu)(1-h)^{-1}\left(\xi_+^{1/h} + \xi_-^{1/h}\right)^h \varphi\left(m(n)\right).$$

Substituting this expression into (5.32), we see that

$$P(R_n(\theta) < m(n)) \sim \prod_{j=\rho m(n)}^{n-(1-\rho)m(n)} P\left(\varepsilon_j \le (\theta-\mu)(1-h)\left(\xi_+^{1/h} + \xi_-^{1/h}\right)^{-h}\frac{1}{\varphi\left(m(n)\right)}\right)$$

$$= \left(P\left(\varepsilon \le (\theta-\mu)(1-h)\left(\xi_+^{1/h} + \xi_-^{1/h}\right)^{-h}\frac{1}{\varphi\left(m(n)\right)}\right)\right)^n \tag{5.33}$$

$$= \left(1 - P\left(\varepsilon > (\theta-\mu)(1-h)\left(\xi_+^{1/h} + \xi_-^{1/h}\right)^{-h}\frac{1}{\varphi\left(m(n)\right)}\right)\right)^n.$$

Once again, we are now in a position to identify the right choice for $m = m(n)$ above. We need to define $m(n)$ in such a way that

$$P\left(\varepsilon > \frac{1}{\varphi\left(m(n)\right)}\right) \approx \frac{1}{n}. \tag{5.34}$$

By the above, we know that we should have

$$\frac{1}{\varphi\left(m(n)\right)} \approx a_n = \left(\frac{1}{\overline{F}}\right)^{\leftarrow}(n),$$

and so we should use

$$m(n) \approx \left(\frac{1}{\varphi}\right)^{\leftarrow}(a_n) := b_n. \tag{5.35}$$

Since a_n is regularly varying at infinity with exponent $1/\alpha$, we conclude that b_n is regularly varying at infinity with exponent $1/(\alpha h)$.

As before, for $x > 0$ we let $m(n) = xb_n$, $n \ge 1$. By the regular variation we have

$$\frac{1}{\varphi\left(x\, b_n\right)} \sim x^h \frac{1}{\varphi\left(b_n\right)} \sim x^h a_n.$$

Therefore,

$$P\left(\varepsilon > (\theta-\mu)(1-h)\left(\xi_+^{1/h} + \xi_-^{1/h}\right)^{-h}\frac{1}{\varphi\left(m(n)\right)}\right)$$

$$= P\left(\varepsilon > (\theta-\mu)(1-h)\left(\xi_+^{1/h}+\xi_-^{1/h}\right)^{-h}\frac{1}{\varphi\left(x\, b_n\right)}\right)$$
$$\sim P\left(\varepsilon > (\theta-\mu)(1-h)\left(\xi_+^{1/h}+\xi_-^{1/h}\right)^{-h} x^h a_n\right)$$
$$\sim p\left((\theta-\mu)(1-h)\left(\xi_+^{1/h}+\xi_-^{1/h}\right)^{-h} x^h\right)^{-\alpha}\frac{1}{n}.$$

Substituting this expression into (5.33) we conclude that for every $x > 0$

$$P(R_n(\theta) < x\, b_n) \to \exp\left\{-p(\theta-\mu)^{-\alpha}(1-h)^{-\alpha}\left(\xi_+^{1/h}+\xi_-^{1/h}\right)^{\alpha h} x^{-\alpha h}\right\} \tag{5.36}$$

as $n \to \infty$.

That is, if the parameters of our process satisfy the regular variation and balance assumptions (5.28), and are non-negative, then

$$\frac{R_n(\theta)}{b_n} \Rightarrow p^{1/\alpha h}\Big(\,(\theta-\mu)(1-h)\Big)^{-1/h}\left(\xi_+^{1/h}+\xi_-^{1/h}\right) Z_{\alpha h} \tag{5.37}$$

for every $\theta > \mu$. Here, once again, $Z_{\alpha h}$ is an extreme value distribution given by (5.26).

Notice that under the above assumptions the length of the long strange intervals grows about as b_n (a regularly varying sequence with exponent $1/(\alpha h) > 1/\alpha$).

In particular, *the rate of growth of the long strange intervals depends on the coefficients* (through the exponent h).

The result (5.37) is due to Rachev and Samorodnitsky (2001).

Notice that:

- If the parameters of the process are within the set Θ_1 of not absolutely summable coefficients, then one can expect that the length of the long strange intervals grows faster than it is the case when the parameters of the process are within the set Θ_0 of absolutely summable coefficients.
- Within Θ_1 the rate of growth of the long strange intervals may change when the parameters change.

Therefore, there is a *phase transition* as one crosses the boundary between Θ_0 and Θ_1, and so it makes sense to say that absolutely summable coefficients correspond to a short memory process and not absolutely summable coefficients correspond to a long memory process.

5.3. **Ruin probabilities.** Here is our second test problem on which we will look for significant changes as the memory of the process changes.

Let $(X_1, X_2, \ldots)$ be a stationary and ergodic stochastic process with a finite mean μ (which may or may not a moving average process). One can view X_n as the amount of claims paid by an insurance company during the nth period, $n = 1, 2, \ldots$.

Let $c > \mu$ be a constant; think of c as the premium received by the insurance company in a single time period. The assumption $c > \mu$ is called *positive loading*. It assures that, on average, the company receives more in premium than it pays out in claims.

If λ is the initial capital of the company, then after n time periods company's capital is $\lambda + cn - (X_1 + \ldots + X_n)$. Hence, if

$$\sup_{n\geq 0}(X_1 + \ldots + X_n - cn) > \lambda$$

then at some point in time the company runs out of money, hence this event is called ruin occurrence, and its probability

$$P\left(\sup_{n\geq 0}(X_1 + \ldots + X_n - cn) > \lambda\right) \tag{5.38}$$

is called *ruin probability.*

Probabilities of the type (5.38) arise not only in insurance context, but also in solutions of stochastic recurrence equations. This includes study of the tails of ARCH and GARCH processes (see e.g. Embrechts et al. (1997)) and study of the stationary distributions in queuing theory through the so-called *Lindley equation* (see e.g. Baccelli and Brémaud (1994)).

We are, mostly, interested in understanding the asymptotic behaviour of the ruin probability as the initial capital λ increases to infinity.

Consider the "random walk"

$$S_0 = 0,\ S_n = X_1 + \ldots + X_n - cn,\ n \geq 1\,.$$

Since the random walk has a negative drift, the ruin occurrence is, for large values of the initial capital λ, an unlikely event. It is not surprising, therefore, that the large deviations approach is very useful in studying the ruin probability

$$P_{\text{ruin}}(\lambda) = P\left(\sup_{n\geq 0} S_n > \lambda\right). \tag{5.39}$$

Several factors affect the asymptotic behaviour of the ruin probability. It is evident that the right tail of the claim size distribution has a major effect on the ruin probability. The heavier is the right tail of the distribution of the claim size X_1, the more likely it is that ruin will occur. It is less evident that the memory in the claim size process affects the ruin probability as well, and it is this effect that we are after.

As always, we are interested in the case when the claim sizes have a heavy tailed distribution.

We start with reviewing what is known in the classical case, when the claim sizes $X_1, X_2, \ldots$ are iid.

Recall that the large deviations logic tells us that ruin is most likely caused by the smallest number of causes. If the claim sizes are iid, then the causes are individual claim sizes being very large. It also turns out that here this smallest number of individual causes is equal to 1. Knowing this, we can figure out what we expect the asymptotic behaviour of the ruin probability to be.

If the ruin is caused by the nth claim X_n, then the ruin time n is likely to be large itself, because very large claims do not come very often. Since the mean claim size is μ, by the time the very large claim arrives, the fortune of the company is about

$$\lambda + c(n-1) - (X_1 + \ldots + X_{n-1}) \sim \lambda + (c-\mu)n .$$

Letting F denote the clime size distribution, we would, therefore, expect that

$$(5.40) \qquad P_{\text{ruin}}(\lambda) \sim \sum_{n=1}^{\infty} P\left(X_n > \lambda + (c-\mu)n\right) = \sum_{n=1}^{\infty} \overline{F}(\lambda + (c-\mu)n)$$

$$\sim \int_0^{\infty} \overline{F}(\lambda + (c-\mu)x)\, dx = \frac{1}{c-\mu} \int_{\lambda}^{\infty} \overline{F}(x)\, dx$$

as $\lambda \to \infty$.

In fact, the asymptotic expression (5.40) of the ruin probability in the heavy tailed case holds if the integrated tails of the claim sizes are subexponential. This was proved by Embrechts and Veraverbeke (1982).

There is also evidence that this asymptotic equivalence remains valid for many heavy tailed stationary ergodic claim size processes (X_n) that are not iid; see e.g. Asmussen et al. (1999).

We will see shortly that, in fact, memory may cause the asymptotic equivalence (5.40) to break down, and in the case of long memory, even the order of magnitude of the ruin probability may differ from the one prescribed by (5.40).

Let us now go back and study the ruin probability when the claim size process (X_n) is a moving average process (5.1). We are assuming, once again, that the noise variables (ε_j) have a finite mean, which is equal to zero. Specifically, we assume that $\alpha \geq 1$. Then $EX_1 = \mu$.

Suppose, first of all, that the parameters of our process are in the set Θ_0 of absolutely summable coefficients. Similarly to our study of long strange intervals, two quantities will play a major role in our calculations:

$$(5.41) \qquad m_\varphi^+ = \sup_{-\infty < n < \infty} \sum_{k=-\infty}^{n} \varphi_k\,, \quad \text{and} \quad m_\varphi^- = \sup_{-\infty < n < \infty} \sum_{k=-\infty}^{n} (-\varphi_k)\,.$$

According to the logic of large deviations we should look for individual causes of the ruin probability. In the case of moving averages these individual factors are evident: those are the noise variables (ε_j). Let us figure out what is the most likely way a given noise variable ε_j can cause ruin.

Observe that for $n \geq 1$

$$(5.42) \qquad S_n = X_1 + \ldots + X_n - cn$$

$$= -(c-\mu)n + \sum_{k=1}^{n} \sum_{j=-\infty}^{\infty} \varphi_{k-j}\, \varepsilon_j = -(c-\mu)n + \sum_{j=-\infty}^{\infty} \varepsilon_j \sum_{k=1-j}^{n-j} \varphi_k\,.$$

Notice that the noise variables can take both positive and negative values. An extremely large value of a noise variable ε_j can, potentially, cause a very large value of a claim size if it is multiplied by a positive factor

$$\sum_{k=1-j}^{n-j} \varphi_k \tag{5.43}$$

in (5.42). Similarly, an extremely small negative value of a noise variable ε_j can cause a very large value of a claim size if it is multiplied by a negative factor in (5.43) in the expression (5.42).

Let us concentrate first on the large positive values of the noise. That is, we are looking at unusually large values of ε_j^+ for all possible j, and let us discuss the factor (5.43) that determines the contribution of ε_j^+ to S_n in (5.42).

Since we are assuming that the coefficients of the moving average process are absolutely summable, when j is a very small negative number, the factor in (5.43) is small uniformly in n. Furthermore, for each fixed j, we do not expect that ε_j^+ is likely to cause the ruin because the tail $P(\varepsilon_j^+ > \lambda)$ is of a smaller order than that predicted even by Embrechts and Veraverbeke theorem (5.40). Since neither noise variables ε_j^+ with very small negative j nor any fixed individual j are expected to play an important role in the ruin probability for large values of λ, the really "important" noise variables are those with high j.

For those noise variables ε_j^+ with a high j the multiplicative factor in (5.43) becomes about

$$\sum_{k=-\infty}^{n-j} \varphi_k \,,$$

and the largest this multiplicative factor ever becomes is exactly

$$m_\varphi^+ = \sup_{-\infty<n<\infty} \sum_{k=-\infty}^{n} \varphi_k \,.$$

It is important to realize that the values of S_n in which ε_j^+ gets multiplied by a factor close to m_φ^+ are those with n equal to about j, because $n-j$ has to lie in a particular region. Because of the ergodicity of the process, the random walk is, at that time, at about the level $-(c-\mu)j$.

Of course, a similar analysis applies to very small negative values of the noise variables ε_j, and the largest negative factor a very large value of ε_j^- can be multiplied with is equal to exactly

$$m_\varphi^- = \sup_{-\infty<n<\infty} \sum_{k=-\infty}^{n} (-\varphi_k) \,.$$

Therefore, the logic of large deviations leads us to expect that

$$P_{\text{ruin}}(\lambda) \sim \sum_{j=1}^{\infty} \left(P\left(m_\varphi^+ \varepsilon_j^+ > \lambda + j(c-\mu)\right) + P\left(m_\varphi^- \varepsilon_j^- > \lambda + j(c-\mu)\right)\right) \tag{5.44}$$

$$\sim \int_1^\infty P\left(m_\varphi^+ \varepsilon^+ > \lambda + y(c-\mu)\right)\, dy + \int_1^\infty P\left(m_\varphi^- \varepsilon^- > \lambda + y(c-\mu)\right)\, dy$$

$$= \int_1^\infty P\left(\varepsilon > \frac{\lambda}{m_\varphi^+} + y\frac{(c-\mu)}{m_\varphi^+}\right)\,dy + \int_1^\infty P\left(\varepsilon < -\frac{\lambda}{m_\varphi^+} - y\frac{(c-\mu)}{m_\varphi^+}\right)\,dy$$

$$= \frac{m_\varphi^+}{c-\mu}\int_{\lambda/m_\varphi^+}^\infty P(\varepsilon > y)\,dy + \frac{m_\varphi^-}{c-\mu}\int_{\lambda/m_\varphi^-}^\infty P(\varepsilon < -y)\,dy$$

as $\lambda \to \infty$.

Using the regular variation assumption (5.2) we conclude that

$$P_{\text{ruin}}(\lambda) \sim \frac{m_\varphi^+}{c-\mu}\,\frac{\lambda}{m_\varphi^+}\,P\left(\varepsilon > \frac{\lambda}{m_\varphi^+}\right)\frac{1}{\alpha-1} + \frac{m_\varphi^-}{c-\mu}\,\frac{\lambda}{m_\varphi^-}\,P\left(\varepsilon < -\frac{\lambda}{m_\varphi^-}\right)\frac{1}{\alpha-1} \tag{5.45}$$

$$\sim \frac{\left[p\,(m_\varphi^+)^\alpha + q\,(m_\varphi^-)^\alpha\right]}{\alpha-1}\frac{1}{c-\mu}\,\lambda\,P(|\varepsilon| > \lambda)$$

as $\lambda \to \infty$.

In fact, under the assumptions slightly stronger than the summability of the coefficients, the asymptotic expression (5.45) was established in Mikosch and Samorodnitsky (2000b).

We do not know if the previous expression, (5.44), holds under the assumption of subexponentiality of the noise variables distribution.

Let us compare the asymptotic result (5.45) with the Embrechts and Veraverbeke theorem (5.40) under the assumption of regular variation. In that case, the Embrechts and Veraverbeke result (in the iid case) says that

$$P_{\text{ruin}}(\lambda) \sim \frac{1}{c-\mu}\,\frac{1}{\alpha-1}\lambda P(X_1 > \lambda) \tag{5.46}$$

$$\sim \frac{\sum_{j=-\infty}^{\infty}|\varphi_j|^\alpha\left[p\,\mathbf{1}_{\{\varphi_j>0\}} + q\,\mathbf{1}_{\{\varphi_j<0\}}\right]}{\alpha-1}\frac{1}{c-\mu}\,\lambda\,P(|\varepsilon| > \lambda)$$

as $\lambda \to \infty$, where we have used (5.6).

In principle, the asymptotic behavior of the ruin probability described by (5.45) can be of a smaller order than that predicted by the Embrechts and Veraverbeke result result (5.46). This would be the case if, say, $p = 1$ and $m_\varphi^+ = 0$ (just take $\varphi_{-1} = -1$ and $\varphi_0 = 1$, with $\varphi_j = 0$ for the rest of the js).

This is the case of strong negative dependence. Apart from such cases, we see that, when the parameters of our process are in the set Θ_0 of absolutely summable coefficients, the ruin probability has the same order of magnitude of decay as in the iid case. While the multiplicative constant in the asymptotic form of the ruin probability is a function of the parameters, the order of magnitude is not, as long as the parameters belong to the set Θ_0.

The ruin probability problem for linear processes when the parameters are in the set Θ_1 of not absolutely summable coefficients, has not, to our knowledge, been considered.

Nonetheless, if one makes the assumption (5.28) of regularly varying and balanced coefficients, then an application of logic of large deviations makes one expect that the ruin probability, $P_{\text{ruin}}(\lambda)$, is, in this case, regularly varying at infinity with exponent $-(\alpha h - 1)$. We will not go into the details.

If this conjecture holds, then one sees that the conclusion here is similar to what one gets by considering long strange intervals: there is a *phase transition* as one crosses the boundary between Θ_0 and Θ_1, and so it makes sense to say that absolutely summable coefficients correspond to a short memory process and not absolutely summable coefficients correspond to a long memory process.

5.4. **Infinitely divisible processes.** A stochastic process $(X(t),\, t \in T)$ is infinitely divisible if all of its finite dimensional distributions are infinitely divisible. Equivalently, for every $k \geq 1$ there is a stochastic process $(Y^{k)}(t),\, t \in T)$ such that

$$(X(t),\, t \in T) \stackrel{\mathrm{d}}{=} \left(\sum_{i=1}^{k} Y_i^{(k)}(t),\, t \in T \right)$$

in terms of equality of finite dimensional distributions.

An infinitely divisible process has two independent components, a Poisson component and a Gaussian component. Since we are considering heavy tails, the Gaussian component with its light tails is usually negligible, and so we will consider here only infinitely divisible processes without a Gaussian component. Such processes have characteristic functions of the form

$$E \exp\left\{ i \sum_{t \in T} \theta(t) X(t) \right\} \tag{5.47}$$

$$= \exp\left\{ \int_{\mathbb{R}^T} \left(\exp\left\{ \sum_{t\in T} i\theta(t)x(t) \right\} - 1 - i \sum_{t \in T} \theta(t)x(t)\mathbf{1}_{[0,1]}(|x(t)|) \right) \nu(d\mathbf{x}) + i \sum_{t \in T} \theta(t) b(t) \right\}$$

for every $(\theta(t),\, t \in T)$ at most finitely many of whose values are different from zero. Here ν is a measure on R^T equipped with the cylindrical σ-field (the *Lévy measure of the process*) and $(b(t),\, t \in T)$ a function on T.

An infinitely divisible process is α-stable, $0 < \alpha < 2$, if its Lévy measure scales:

$$\nu(aA) = a^{-\alpha} \nu(A) \tag{5.48}$$

for every measurable A and $a > 0$.

The Lévy measure ν of an infinitely divisible process is its most important characteristic. It gives us the most direct view of how the Poissonian jumps underlying an infinitely divisible process combine and how they affect the properties of the process.

These Poissonian jumps also turn out to be the factors that can cause unlikely events. We will see the details a number of times in the sequel.

We start with a very general result related to rare events. For most applications one can assume that the parameter set T is countable (by reducing, if necessary, everything to a countable subset of T, e.g. to countable numbers as a subset of $\mathbb{R}$). In that case the Lévy measure ν in (5.47) is σ-finite.

Many questions of interest can be thought of in the following way. We are given a measurable functional $\phi : I\!R^T \to I\!R$, and we are interested in the right tail of the distribution of $\phi(\mathbf{X})$:

$$P\left(\phi(\mathbf{X}) > \lambda\right), \quad \lambda \to \infty . \tag{5.49}$$

Typical examples include

$$\phi_1(\mathbf{x}) = \sup_{t \in T} x(t) \quad \text{and} \quad \phi_2(\mathbf{x}) = \sup_{t \in T} |x(t)|. \tag{5.50}$$

Define

$$H(\lambda) = \nu\left\{\mathbf{x} \in I\!R^T : \phi(\mathbf{x}) > \lambda\right\}, \tag{5.51}$$

where ν is the Lévy measure of an infinitely divisible process $\mathbf{X}$.

If one compares (5.49) with (5.51) then one sees that the two expressions compute the measures of the same event $\{\phi(\mathbf{x}) > \lambda\}$. In the former case the measure is the probability law of the process, whereas in the latter case the measure is the Lévy measure of the process. In particular, $H(\lambda)$ may be infinite for some or all λ.

It turns out that, *in heavy tailed cases*, the two expressions are often asymptotically equivalent. Specifically, assume that the functional ϕ is *subadditive*:

$$\phi(\mathbf{x}_1 + \mathbf{x}_2) \leq \phi(\mathbf{x}_1) + \phi(\mathbf{x}_2) \tag{5.52}$$

for all $\mathbf{x}_i \in I\!R^T$, $i = 1, 2$. Note that the functionals given in (5.50) are subadditive.

Let $q : I\!R^T \to [0, \infty]$ be a measurable subadditive function such that $q(\mathbf{0}) = 0$ and $q(c\mathbf{x}) \leq q(\mathbf{x})$ for all $\mathbf{x} \in I\!R^T$ and $c \in [-, 1, 1]$. Such q is called a measurable pseudonorm. Assume that

$$|\phi(\mathbf{x})| \leq q(\mathbf{x}) \quad \text{for all } \mathbf{x} \in I\!R^T, \tag{5.53}$$

for a lower-semicontinuous pseudonorm q, such that

$$P\left(q(\mathbf{X}) < \infty\right) = 1 . \tag{5.54}$$

Under the above assumptions $H(\lambda) \to 0$ as $\lambda \to \infty$. If

$$1 - \min\left(H(\lambda), 1\right) \quad \text{is a subexponential distribution function}, \tag{5.55}$$

then

$$\lim_{\lambda \to \infty} \frac{P\left(\phi(\mathbf{X}) > \lambda\right)}{H(\lambda)} = 1 . \tag{5.56}$$

The above result is due to Rosiński and Samorodnitsky (1993).

Example 26. Let $(X(t),\, 0 \leq t \leq 1)$ be a Lévy process (i.e. a process with stationary independent increments). This is an infinitely divisible process whose Lévy measure is given by

$$\nu = (\text{Leb} \times \rho) \circ J^{-1}, \tag{5.57}$$

where $J : [0, 1] \times I\!R \to I\!R^{[0,1]}$ is a measurable map given by

$$J(s, y)(t) = y\mathbf{1}(t \leq s),\ 0 \leq t \leq 1$$

for $s \in [0,1]$ and $y \in \mathbb{R}$. Further, ρ is a one-dimensional Lévy measure. That is, ρ is a σ-finite measure on $\mathbb{R}$ such that $\int (1 \wedge x^2)\rho(dx) < \infty$.

Take $\phi_1(\mathbf{x}) = \sup_{0\leq t\leq 1} x(t)$ in (5.50). In this case for $\lambda > 0$

$$H(\lambda) = (\text{Leb} \times \rho)\left\{(s,y) : \sup_{0\leq t\leq 1} y\mathbf{1}(t \leq s) > \lambda\right\}$$

$$= \rho\left((\lambda, \infty)\right) .$$

Assume that the tail of ρ is subexponential. That is,

$$1 - \min\left(\rho\left((\lambda,\infty)\right), 1\right)$$

is a subexponential distribution function. Then (5.56) applies (use $q(\mathbf{x}) = \sup_{0\leq t\leq 1} |x(t)|$), and so we conclude that

$$\lim_{\lambda\to\infty} \frac{P\left(\sup_{0\leq t\leq 1} X(t) > \lambda\right)}{\rho\left((\lambda,\infty)\right)} = 1 . \tag{5.58}$$

Notice that the requirement of heavy tails needed for the result (5.56) to hold is that of subexponentiality.

To see how (5.56) is related to the large deviations idea, let us assume for simplicity that the Lévy measure ν of the infinitely divisible process X in (5.47) is, actually, finite. Let $\|\nu\| = \nu(\mathbb{R}^T)$. Let N be a Poisson random variable with the mean $\|\nu\|$, and let $(\mathbf{Y}_j,\ j \geq 1)$ be a sequence of iid $\mathbb{R}^T$ valued random variables with common law $\nu/\|\nu\|$, independent of N.

It is elementary to check by comparing the characteristic functions that the process

$$\sum_{j=1}^{N} Y_j(t) + a(t) , \tag{5.59}$$

with

$$a(t) = b(t) - \int_{\mathbb{R}^T} x(t)\mathbf{1}(|x(t)| \leq 1)\,\nu(dx),\ \ t \in T$$

is a version of $(X(t),\ t \in T)$. Therefore,

$$\phi(\mathbf{X}) \stackrel{\text{d}}{=} \phi\left(\sum_{j=1}^{N} \mathbf{Y}_j + \mathbf{a}\right) . \tag{5.60}$$

Notice that the collection $(\mathbf{Y}_j,\ j = 1, \ldots, N)$ forms the points of a Poisson random measure on $\mathbb{R}^T$ with mean measure ν. The large deviations approach tells us that it is one of these Poisson points that is responsible for the rare event $\{\phi(\mathbf{X}) > \lambda\}$ (the non-random function $\mathbf{a}$ does not contribute much). That is, we expect that

$$P\left(\phi(\mathbf{X}) > \lambda\right) \sim P\left(\phi(\mathbf{Y}_j) > \lambda \ \text{ for some } j = 1, \ldots, N\right) .$$

Conditioning on N and using the inclusion-exclusion formula we see that

$$(5.61) \qquad P\left(\phi(\mathbf{X}) > \lambda\right) \sim E\left[\sum_{j=1}^{N} P\left(\phi(\mathbf{Y}_j) > \lambda\right)\right] = EN\, P\left(\phi(\mathbf{Y}_1) > \lambda\right) .$$

Notice that

$$P\left(\phi(\mathbf{Y}_1) > \lambda\right) = \frac{1}{\|\nu\|}\nu\left\{\mathbf{x} \in I\!R^T : \phi(\mathbf{x}) > \lambda\right\} = \frac{1}{\|\nu\|}H(\lambda).$$

Substituting this into (5.61) gives us

$$P\left(\phi(\mathbf{X}) > \lambda\right) = \|\nu\| \left(\frac{1}{\|\nu\|}H(\lambda)\right) = H(\lambda)$$

as $\lambda \to \infty$.

One can see the Poissonian structure of an infinitely divisible process, resulting in an even better understanding of the way large deviations work here, in the case when an infinitely divisible process is given as a stochastic integral.

Let $(S, \mathcal{A})$ be a measurable space and M an infinitely divisible random measure on this space, with Lévy measure G and shift measure γ_0. Here G is a σ-finite measure on $(S \times I\!R, \mathcal{A} \times \mathcal{B})$ and γ_0 is a σ-finite signed measure on $(S, \mathcal{A})$. Let

$$(5.62) \qquad \mathcal{A}_0 = \left\{ A \in \mathcal{A} :\ \gamma(A) := |\gamma_0|(A) + \int_A \int_{I\!R} \min(1, x^2) G(ds, dx) < \infty \right\}.$$

The set function γ extends to a σ-finite measure on $(S, \mathcal{A})$, called a *control measure* of the random measure M.

One can view the random measure M as a stochastic process of the type $(M(A), A \in \mathcal{A}_0)$, such that

- M is independently scattered. That is, for any disjoint $\mathcal{A}_0$ sets $A_1, \ldots, A_n$, the random variables $M(A_1), \ldots, M(A_n)$ are independent
- M is σ-additive. That is, for any disjoint $\mathcal{A}_0$ sets $A_1, A_2, \ldots$ such that $\cup_{i=1}^{\infty} A_i \in \mathcal{A}_0$ we have $M(\cup_{i=1}^{\infty} A_i) = \Sigma_{i=1}^{\infty} M(A_i)$ a.s.
- For every $A \in \mathcal{A}_0$, $M(A)$ is an infinitely divisible random variable with

$$(5.63) \qquad E\exp(i\theta M(A)) = \exp\left\{ i\theta\gamma_0(A) + \int_A \int_{I\!R} (e^{i\theta x} - 1 - i\theta x \mathbf{1}_{[0,1]}(|x|)) G(ds, dx) \right\}.$$

Consider a stochastic process given in the form

$$(5.64) \qquad X(t) = \int_S f(t, s) M(ds),\ t \in T,$$

where $f : T \times S \to I\!R$ is a deterministic function. Such process is always infinitely divisible. See Rajput and Rosiński (1989) for more details on infinitely divisible random measures and on conditions on the kernel $f(t, s)$ in (5.64) ensuring that the stochastic integral is well defined.

For an infinitely divisible process given by in (5.64) its Lévy measure ν is given in the form

$$\nu = G \circ J^{-1}\,, \tag{5.65}$$

where $J:\ S\times I\!R \to I\!R^T$ is given by

$$J(s,y)(t) = yf(t,s),\ t\in T \tag{5.66}$$

for $s\in S$ and $y\in I\!R$.

Expression (5.65) often allows one a straightforward way to compute the function H in (5.51) above.

Since the signed measure γ_0 in (5.63) plays no role in Lévy measure of the process X (other than affecting the conditions on the kernel f under which the process is well defined) and, hence, does not affect the function H, we will simply assume that $\gamma_0 = 0$.

Below we will explore some important special cases.

The computations become especially convenient if the Lévy measure G of the random measure M is given in the form

$$G\,(A\times B) = \int_A \rho(s,B)\,\pi(ds),\ A\in\mathcal{A},\ B\in\mathcal{B}\,, \tag{5.67}$$

where π is a probability measure on S (which is, necessarily, equivalent to the extended measure γ in (5.62); π is also called a control measure measure of M), and $\rho(s,\cdot),\ s\in S$ is a family of Lévy measures on $I\!R$.

Suppose also that the functional ϕ is *homogeneous*:

$$\phi(c\mathbf{x}) = c\phi(\mathbf{x}),\ c>0,\ \mathbf{x}\in I\!R^T. \tag{5.68}$$

Many functionals of interest are homogeneous: various *sup* functionals, L^p-norm functionals, etc.

Under the above assumptions we have

$$\begin{aligned} H(\lambda) &= \int_S \pi(ds)\int_{I\!R} \mathbf{1}\left(\phi\left(y\,f(\cdot,s)\right) > \lambda\right)\,\rho(s,dy) \\ &= \int_S \left[\rho\left(s,\left(\frac{\lambda}{\phi\left(f(\cdot,s)\right)_+},\infty\right)\right) + \rho\left(s,\left(-\infty,-\frac{\lambda}{\phi\left(-f(\cdot,s)\right)_+}\right)\right)\right]\,\pi(ds)\,. \end{aligned} \tag{5.69}$$

One obtains an α-stable process with $0<\alpha<2$ if

$$\rho(s,dy) = w(s)\left(\frac{1+\beta(s)}{2}y^{-(\alpha+1)}\mathbf{1}(y>0) + \frac{1-\beta(s)}{2}|y|^{-(\alpha+1)}\mathbf{1}(y<0)\right)\,dy\,. \tag{5.70}$$

Here $w:\ S\to(0,\infty)$ and $\beta:\ S\to[-1,1]$ are measurable functions.

If the functional ϕ is homogeneous, we obtain by (5.69) and (5.70) that

$$H(\lambda) = \lambda^{-\alpha}\int_S \left[\frac{1+\beta(s)}{2}\left(\phi\left(f(\cdot,s)\right)_+\right)^\alpha + \frac{1-\beta(s)}{2}\left(\phi\left(-f(\cdot,s)\right)_+\right)^\alpha\right]\frac{1}{\alpha}w(s)\,\pi(ds)\,. \tag{5.71}$$

In particular, the function H is regularly varying at infinity with exponent $-\alpha$ and, hence, the subexponentiality assumption (5.55) holds.

Example 27. One of interesting ways to measure the length of dependence is by looking at joint tails. Suppose $(X_1, X_2, \ldots)$ is a stationary process. For some (large) $\lambda > 0$ consider

$$P\left(X_1 > \lambda, \ldots, X_n > \lambda\right)$$

as a function of n. Obviously, if the sequence $(X_1, X_2, \ldots)$ is iid, then the above expression is an exponentially decaying function of n.

If $(X_1, X_2, \ldots)$ is an α-stable process, then the limit

$$K_n = \lim_{\lambda\to\infty} \lambda^\alpha P\left(X_1 > \lambda, \ldots, X_n > \lambda\right) \tag{5.72}$$

always exists, and one can look at the rate of decay of K_n as an indication of the length of dependence. As above, $K_n = 0$ for $n \geq 2$ if $(X_1, X_2, \ldots)$ is a sequence of iid α-stable random variables.

Suppose that the α-stable process $(X_1, X_2, \ldots)$ is given by an integral representation (5.64). We see immediately by (5.71) and (5.56) that

$$\lim_{\lambda\to\infty} \lambda^\alpha P\left(\max_{i=1,\ldots,n} X_i > \lambda\right) \tag{5.73}$$

$$= \int_S \left[\frac{1+\beta(s)}{2} \max_{i=1,\ldots,n} \left(f(i,s)_+\right)^\alpha + \frac{1-\beta(s)}{2} \max_{i=1,\ldots,n} \left(f(i,s)_-\right)^\alpha\right] \frac{1}{\alpha} w(s)\, \pi(ds)\,.$$

An application of the inclusion - exclusion formula to (5.73) shows that

$$K_n = \lim_{\lambda\to\infty} \lambda^\alpha P\left(\min_{i=1,\ldots,n} X_i > \lambda\right) \tag{5.74}$$

$$= \int_S \left[\frac{1+\beta(s)}{2} \min_{i=1,\ldots,n} \left(f(i,s)_+\right)^\alpha + \frac{1-\beta(s)}{2} \min_{i=1,\ldots,n} \left(f(i,s)_-\right)^\alpha\right] \frac{1}{\alpha} w(s)\, \pi(ds)\,.$$

This shows that the rate of decay of K_n is related to the rate of decay of $(f(n,s)_+,\, n \geq 1)$ and $(f(n,s)_-,\, n \geq 1)$ as n increases, for "most" of $s \in S$.

As a concrete example let us consider the so-called (one-sided) *linear fractional α-stable noises* given by

$$X_n = \int_{I\!R} g(s-n)\, M(ds), \ \ n = 1, 2, \ldots\,, \tag{5.75}$$

where

$$g(s) = (-s)_+^{H-1/\alpha} - (-s-1)_+^{H-1/\alpha}, \ \ s \in I\!R\,, \tag{5.76}$$

$0 < H < 1$, and we interpret $0^a = 0$ for any real a. The α-stable random measure M has the Lebesgue control measure π, weight function $w(s) \equiv 1$ and for simplicity we will choose the α-stable random measure M to be symmetric ($\beta(s) = 0$ for all $s \in I\!R$). In that case the process $(X_1, X_2, \ldots)$ is itself a symmetric α-stable process.

Notice that, if $1 < \alpha < 2$, and $H = 1/\alpha$, then the function g in (5.76) becomes

$$g(s) = \mathbf{1}\left(-1 < s \leq 0\right)\,,$$

so that the resulting process in (5.75)

$$X_n = \int_{\mathbb{R}} \mathbf{1}(-1 < s - n \le 0)\ M(ds) = M\left((n-1, n]\right),\ n = 1, 2, \ldots,$$

is just a sequence of iid α-stable random variables.

The process $(X_1, X_2, \ldots)$ in (5.75) is the increment process of one-sided linear fractional α-stable motions given by

$$Y(t) = \int_{\mathbb{R}} \left((t-s)_+^{H-1/\alpha} - (-s)_+^{H-1/\alpha}\right) M(ds), \tag{5.77}$$

$t \in \mathbb{R}$, which are SSSI (self-similar stationary increment) α-stable Specifically,

$$X_n = Y(n) - Y(n-1),\ n \ge 1,$$

In particular, linear fractional α-stable noises are stationary processes (as are, in general, any processes of the form (5.75), with any kernel f, which are referred to as *continuous time moving averages*).

Suppose first that $H < 1/\alpha$. In that case it is straightforward to check that for $n = 1, 2, \ldots$

$$\min_{i=1,\ldots,n} f(i,s)_+ = \min_{i=1,\ldots,n} g(s-i)_+ = 0,\ s \in \mathbb{R}$$

and

$$\min_{i=1,\ldots,n} f(i,s)_- = \min_{i=1,\ldots,n} g(s-i)_- = \begin{cases} 0 & \text{if } s \ge 0 \\ (n-1-s)^{H-1/\alpha} - (n-s)^{H-1/\alpha} & \text{if } s < 0 \end{cases}.$$

Therefore, by (5.74)

$$K_n = \frac{1}{2\alpha}\int_{-\infty}^{0} \left((n-1-s)^{H-1/\alpha} - (n-s)^{H-1/\alpha}\right)^\alpha ds$$

$$= \frac{1}{2\alpha}\int_{n}^{\infty} \left((x-1)^{H-1/\alpha} - (x)^{H-1/\alpha}\right)^\alpha ds \sim \frac{\left(\frac{1}{\alpha} - H\right)^\alpha}{2\alpha^2(1-H)}\, n^{-\alpha(1-H)}$$

as $n \to \infty$.

Using a similar argument for $1/\alpha < H < 1$ we conclude that

$$K_n \sim \frac{\left|\frac{1}{\alpha} - H\right|^\alpha}{2\alpha^2(1-H)}\, n^{-\alpha(1-H)} \tag{5.78}$$

as $n \to \infty$.

Notice that the rate of decay of K_n is slower when H is large, and in this sense the memory becomes longer as H increases.

Let us go back to the general infinitely divisible stochastic process given in the form of a stochastic integral (5.64). Let the Lévy measure of the random measure M be given in the form (5.67). We will assume, for simplicity, that the random measure M is symmetric, which means that

$$\rho(s, -B) = \rho(s, B),\ B \in \mathcal{B}$$

(on a set of π measure 1). For $s \in S$ and $u > 0$ let

$$R(u,s) = \inf\{y > 0 : \ \rho(s,(y,\infty)) \le u\} \tag{5.79}$$

be the right-continuous inverse of the tail of $\rho(s,\cdot)$.

Let (ε_j), (Γ_j) and (ξ_j) be three independent sequences of random variables such that:

- $\varepsilon_1, \varepsilon_2, \ldots$ are iid Rademacher random variables. That is,
$$P(\varepsilon_1 = 1) = P(\varepsilon_1 = -1) = \frac{1}{2}.$$
- $\Gamma_1, \Gamma_2, \ldots$ are the arrival times of a unit rate Poisson process on $(0,\infty)$.
- $\xi_1, \xi_2, \ldots$ are iid S-valued random variables with a common law π.

Then the stochastic process in (5.64) can be represented in the form

$$(X(t),\, t \in T) \stackrel{\mathrm{d}}{=} \left(\sum_{j=1}^{\infty} \varepsilon_j\, R\left(\Gamma_j, \xi_j\right)\, f\left(t, \xi_j\right),\, t \in T \right) \tag{5.80}$$

in the sense of equality of finite-dimensional distributions.

The representation (5.80) is called a *series representation* of the infinitely divisible process (5.64).

Such representation is not unique (one can select a function R of a different form), and representations also exist in the non-symmetric case. Then one may need to center the series to make it converge. See Rosiński (1990) for a general treatment of series representations.

Series representations of infinitely divisible processes are very important. In particular, it emphasizes the role of Poisson jumps in the structure of an infinitely divisible process. In large deviations type of situations the Poisson points $\Gamma_1, \Gamma_2, \ldots$ are identifiable as individual factors.

Notice that, for every fixed $s \in S$, the function $(R(u,s),\, u > 0)$ is nonincreasing. Since $\Gamma_1 \le \Gamma_2 \le \ldots$, we see that $R(\Gamma_1, \xi_1)$ is stochastically larger than $R(\Gamma_j, \xi_j)$ for $j \ge 2$.

Therefore, if a functional ϕ has certain monotonicity properties with respect to the scale of its argument (in particular, if the functional ϕ has the homogeneity property (5.68)), then

we expect that the very first term in the series representation (5.80) is most likely to cause the rare event $\{\phi(\mathbf{X}) > \lambda\}$ for large λ.

To see how this works, notice that, with this logic,

$$P\left(\phi(\mathbf{X}) > \lambda\right) \sim P\left(\phi\left(\varepsilon_1\, R\left(\Gamma_1, \xi_1\right)\, f\left(\cdot, \xi_1\right)\right) > \lambda\right) \tag{5.81}$$
$$= \frac{1}{2} P\left(\phi\left(R\left(\Gamma_1, \xi_1\right)\, f\left(\cdot, \xi_1\right)\right) > \lambda\right) + \frac{1}{2} P\left(\phi\left(-R\left(\Gamma_1, \xi_1\right)\, f\left(\cdot, \xi_1\right)\right) > \lambda\right).$$

By the homogeneity property (5.68),

$$P\left(\phi\left(R\left(\Gamma_1, \xi_1\right)\, f\left(\cdot, \xi_1\right)\right) > \lambda\right) = \int_S P\left(R\left(\Gamma_1, s\right) > \frac{\lambda}{\phi\left(f\left(\cdot, s\right)\right)_+}\right)\, \pi(ds). \tag{5.82}$$

Observe that for every $x > 0$

$$P\Big(R(\Gamma_1, s) > x\Big) = P\Big(\rho(s, (x, \infty)) > \Gamma_1\Big) = 1 - \exp\Big\{ -\rho(s, (x, \infty))\Big\} \sim \rho(s, (x, \infty))$$

as $x \to \infty$. Assuming that using this asymptotic equivalence inside the integral in (5.82) is justified, we expect that

$$P\Big(\phi\Big(R(\Gamma_1, \xi_1)\, f(\cdot, \xi_1)\Big) > \lambda\Big) \sim \int_S \rho\left(s, \left(\frac{\lambda}{\phi\left(f(\cdot, s)\right)_+}, \infty\right)\right) \pi(ds) \tag{5.83}$$

as $\lambda \to \infty$.

Similarly, we expect that

$$P\Big(\phi\Big(- R(\Gamma_1, \xi_1)\, f(\cdot, \xi_1)\Big) > \lambda\Big) \sim \int_S \rho\left(s, \left(\frac{\lambda}{\phi\left(-f(\cdot, s)\right)_+}, \infty\right)\right) \pi(ds)\,. \tag{5.84}$$

Substituting (5.83) and (5.84) into (5.81), we conclude that

$$P(\phi(\mathbf{X}) > \lambda)$$
$$\sim \int_S \left[\rho\left(s, \left(\frac{\lambda}{\phi\left(f(\cdot, s)\right)_+}, \infty\right)\right) + \rho\left(s, \left(-\infty, -\frac{\lambda}{\phi\left(-f(\cdot, s)\right)_+}\right)\right)\right] \pi(ds) = H(\lambda)$$

by (5.69), as promised, for example, by (5.56).

Obviously, the above argument is valid only under certain conditions. Some of these situations are described above in (5.56). Below we will see additional situations where this general principle holds, and is useful.

Note, however, that whether or not it is the first term in the series representation (5.80) is most likely to cause the event $\{\phi(\mathbf{X}) > \lambda\}$ for large λ,

the tail measure $H(\lambda)$ is always the sum of the contributions to the tail probability $P(\phi(\mathbf{X}) > \lambda)$ from one of the Poisson points $\Gamma_1, \Gamma_2, \ldots$ in that series representation.

Therefore, the series representation (5.80) is exactly the right way to look at individual causes of rare events associated with infinitely divisible processes:

the individual causes are the points of the Poisson random measure on $I\!R^T$ given by

$$\left(\varepsilon_j\, R(\Gamma_j, \xi_j)\, f(t, \xi_j)\,,\, t \in T\right)\,,\ j = 1, 2, \ldots\,. \tag{5.85}$$

Indeed, as above,

$$\sum_{j=1}^{\infty} P\Big(\phi\Big(\varepsilon_j\, R(\Gamma_j, \xi_j)\, f(\cdot, \xi_j)\Big) > \lambda\Big)$$
$$= \sum_{j=1}^{\infty} \int_0^{\infty} \frac{z^{j-1}}{(j-1)!}\, e^{-z} P\Big(\phi\Big(\varepsilon_j\, R(z, \xi_j)\, f(\cdot, \xi_j)\Big) > \lambda\Big)\, dz \tag{5.86}$$

$$= \int_0^\infty P\left(\phi\left(\varepsilon_1 R(z,\xi_1)\, f(\cdot,\xi_1)\right) > \lambda\right) dz$$

$$= \frac{1}{2}\int_0^\infty P\left(\phi\left(R(z,\xi_1)\, f(\cdot,\xi_1)\right) > \lambda\right) dz + \frac{1}{2}\int_0^\infty P\left(\phi\left(-R(z,\xi_1)\, f(\cdot,\xi_1)\right) > \lambda\right) dz\,.$$

Proceeding as before, we see that

$$P\left(\phi\left(R(z,\xi_1)\, f(\cdot,\xi_1)\right) > \lambda\right) = \int_S \mathbf{1}\left(\rho\left(s, \left(\frac{\lambda}{\phi\left(f(\cdot,s)\right)_+}, \infty\right)\right) > z\right) \pi(ds)$$

and, hence, the first term in the right hand side in (5.86) is

$$\frac{1}{2}\int_0^\infty dz \int_S \mathbf{1}\left(\rho\left(s, \left(\frac{\lambda}{\phi\left(f(\cdot,s)\right)_+}, \infty\right)\right) > z\right) \pi(ds)$$

$$= \int_S \rho\left(s, \left(\frac{\lambda}{\phi\left(f(\cdot,s)\right)_+}, \infty\right)\right) \pi(ds)\,.$$

Treating in the same way the second term the right hand side in (5.86), we obtain

$$\sum_{j=1}^\infty P\left(\phi\left(\varepsilon_j R(\Gamma_j,\xi_j)\, f(\cdot,\xi_j)\right) > \lambda\right)$$

$$= \int_S \left[\rho\left(s, \left(\frac{\lambda}{\phi\left(f(\cdot,s)\right)_+}, \infty\right)\right) + \rho\left(s, \left(-\infty, -\frac{\lambda}{\phi\left(-f(\cdot,s)\right)_+}\right)\right)\right] \pi(ds) = H(\lambda),$$

as required.

In fact, what the series representation (5.80) does is represent an infinitely divisible process as an integral with respect to a Poisson random measure on $\mathbb{R}^T$.

For the remainder of our discussion we will concentrate on symmetric α-stable (SαS) processes. Recall that, in this case,

$$\rho(s,dy) = \frac{1}{2} w(s)\, y^{-(\alpha+1)}\, dy\,,\ s\in S\,,\ y>0\,,$$

and so the function R in (5.79) is given by

$$R(u,s) = \left(\frac{1}{2\alpha}\right)^{1/\alpha} w(s)^{1/\alpha}\, u^{-1/\alpha} := c_\alpha\, w(s)^{1/\alpha}\, u^{-1/\alpha} \tag{5.87}$$

for $s\in S$ and $u>0$.

Therefore, the most commonly used series representation of SαS processes has the form

$$(X(t),\, t\in T) \stackrel{\mathrm{d}}{=} \left(c_\alpha \sum_{j=1}^\infty \varepsilon_j\, \Gamma_j^{-1/\alpha}\, w(\xi_j)^{1/\alpha} f(t,\xi_j)\,,\, t\in T\right)\,. \tag{5.88}$$

There are several interesting facts about series representations for SαS processes and random variables.

If X is a SαS random variable, then it has a characteristic function of the form

$$Ee^{i\theta X} = e^{-\sigma^\alpha|\theta|^\alpha},\ \theta\in\mathbb{R} \tag{5.89}$$

for some $\sigma \geq 0$ which is called *the scale parameter* of X. Then X has a series representation of the form

$$X \stackrel{\mathrm{d}}{=} C_\alpha \sum_{j=1}^{\infty} \varepsilon_j \Gamma_j^{-1/\alpha} Y_j \,, \tag{5.90}$$

some $C_\alpha > 0$, where this time $Y_1, Y_2, \ldots$ are iid random variables independent of the sequence $\varepsilon_1, \varepsilon_2, \ldots$ and $\Gamma_1, \Gamma_2, \ldots$ as above.

The random variables (Y_j) must have a finite absolute α moment, and

$$E|Y_1|^\alpha = \sigma^\alpha \,. \tag{5.91}$$

Conversely, the distribution of the random series in the right hand side of (5.90) depends only on the absolute α moment of the random variables $Y_1, Y_2, \ldots$.

One application of the above fact is that, in the series representation (5.88) of a SαS process one may replace the Rademacher random variables $\varepsilon_1, \varepsilon_2, \ldots$ by iid zero mean normal random variables with the variance selected in such a way so as to make the absolute α moment equal to 1.

Therefore, an alternative series representation is

$$(X(t),\, t \in T) \stackrel{\mathrm{d}}{=} \left(c'_\alpha \sum_{j=1}^{\infty} G_j \Gamma_j^{-1/\alpha} w(\xi_j)^{1/\alpha} f(t, \xi_j) \,,\, t \in T \right) \,, \tag{5.92}$$

where c'_α is a finite positive cons-tat, and $G_1, G_2, \ldots$ are iid standard normal random variables. In particular, any SαS process is a mixture of zero mean Gaussian processes (but we will not use this fact in the sequel).

Many facts about α-stable processes can be found in Samorodnitsky and Taqqu (1994).

We will study now the question of short versus long large dependence for stationary SαS processes by concentrating on the *ruin problem*. As before, let (X_n) be an ergodic stationary SαS process, which is given in the form of a stochastic integral

$$X_n = \int_S f_n(s)\, M(ds) \,,\; n = 1, 2, \ldots, \tag{5.93}$$

where M is a SαS random measure with a corresponding probability measure π in (5.67) and a function w in (5.70). The σ-finite measure

$$m(A) = \int_A w(s)\, \pi(ds) \,,\; A \in \mathcal{A} \tag{5.94}$$

is also called a control measure of M, and it is the most common usage of this term in the stable context.

Finally, $f_1, f_2, \ldots$ are measurable functions on S such that $\int_S |f_n(s)|^\alpha m(ds) < \infty$ for all $n \geq 1$ (so that all the integrals are well defined).

Since we are studying ruin probabilities and need to have a finite mean, we will assume that $1 < \alpha < 2$, in which case $EX_1 = 0$.

Let $c > 0$ and consider, as before, the ruin probability in (5.39):

$$P_{\text{ruin}}(\lambda) = P\left(\sup_{n\geq 0} S_n > \lambda\right)$$

with

$$S_0 = 0,\ S_n = X_1 + \ldots + X_n - cn,\ n \geq 1\,.$$

Notice that we are in an apparently similar situation to the one we had before: namely, we are looking, as in (5.49), at the tail

$$P\left(\phi(\mathbf{X} - c\mathbf{I}) > \lambda\right),\quad \lambda \to \infty$$

of a measurable functional ϕ given by

$$\phi_{\text{ruin}}(\mathbf{x}) = \sup_{n\geq 0}\,(x_1 + \ldots + x_n) \tag{5.95}$$

for $\mathbf{x} = (x_1, x_2, \ldots)$.

Even though the functional ϕ_{ruin} is, clearly, subadditive, the result of Rosiński and Samorodnitsky (1993) mentioned above that justified the asymptotic relation (5.56), does not apply here because that theorem requires the functional to be bounded from above by a finite pseudonorm (requirement (5.53)). This, intuitively, means that the process

$$X_1 + \ldots + X_n - cn,\ n \geq 1$$

has to be bounded on *both sides*, whereas in our case it is only bounded from above because of the negative drift. Nonetheless, it turns out that under certain assumptions the asymptotic relation (5.56) still holds.

Specifically, let m_n be the scale parameter of the SαS random variable $X_1 + \ldots + X_n$, $n = 1, 2, \ldots$. One can show that the ergodicity of the process (X_n) implies that $m_n = o(n)$ as $n \to \infty$. Assume a stronger requirement: for some $\beta \in (0, 1)$

$$\lim_{n\to\infty} \frac{m_n}{n^\beta} = 0\,. \tag{5.96}$$

The assumption (5.96) turns out to hold for virtually all stationary ergodic SαS processes of interest. Under the assumption (5.96), the equivalence (5.56)

$$P\left(\phi(\mathbf{X}) > \lambda\right) \sim H(\lambda)$$

as $\lambda \to \infty$ still holds. This result is due to Mikosch and Samorodnitsky (2000a).

It is interesting to note that, in this case, it is *not* necessarily the case that it is only the first term in the series representation (5.88) (the largest Poisson jump) may cause the ruin. It is the sum of possible effects of all Poisson jumps that causes the ruin.

In a recent paper Braverman (2002) stated conditions on the series representation of the process (X_n) (as we know, it is not unique) under which it is still the case that only the first term in the series representation (5.88) is likely to cause ruin.

Note that, in this case, by (5.65) and (5.67),

$$H(\lambda) = \nu\left\{\mathbf{x} \in I\!R^T :\ \phi(\mathbf{x}) > \lambda\right\} = \int_S m(ds) \int_0^\infty x^{-(\alpha+1)}\,dx \tag{5.97}$$

$$\left[\frac{1}{2}\mathbf{1}\left(\sup_{n\geq 0}\left(x\sum_{k=1}^{n} f_k(s)-cn\right)>\lambda\right)+\frac{1}{2}\mathbf{1}\left(\sup_{n\geq 0}\left(-x\sum_{k=1}^{n} f_k(s)-cn\right)>\lambda\right)\right].$$

However,

$$\left\{\sup_{n\geq 0}\left(x\sum_{k=1}^{n} f_k(s)-cn\right)>\lambda\right\}=\left\{x\sum_{k=1}^{n} f_k(s)>cn+\lambda \text{ for some } n\geq 0\right\}$$
$$=\left\{x>\inf_{n\geq 0}\frac{\lambda+cn}{\left(\sum_{k=1}^{n} f_k(s)\right)_+}\right\}.$$

Substituting this expression into (5.97) above we obtain that, under the assumption (5.96)

(5.98)
$$P_{\text{ruin}}(\lambda)\sim H(\lambda)=\frac{1}{2\alpha}\int_S \sup_{n\geq 0}\frac{\left(\sum_{k=1}^{n} f_k(s)\right)_+^\alpha}{(\lambda+cn)^\alpha}\,m(ds)+\frac{1}{2\alpha}\int_S \sup_{n\geq 0}\frac{\left(-\sum_{k=1}^{n} f_k(s)\right)_+^\alpha}{(\lambda+cn)^\alpha}\,m(ds)$$

as $\lambda\to\infty$.

We will use the above asymptotic equivalence to study the length of memory in stationary ergodic SαS processes via the rate of decay of the ruin probability.

5.5. **Structure of stationary SαS processes.** Even though the integral representation of α-stable processes is not unique, stationary SαS processes have integral representations of a special form, that allows one to get quite a bit of insight into the structure of these processes. Theory of such representations is due to Rosiński (1995).

Let $(S,\mathcal{A},m)$ be a σ-finite measure space, and let $\varphi: S\to S$ be a one-to-one map, such that both φ and φ^{-1} are measurable. We assume that φ is a non-singular map (that is, the measure $m\circ\varphi^{-1}$ is equivalent to the measure m). The family $(\varphi^n,\ n=\ldots,-1,0,1,2,\ldots)$ is called *a non-singular flow* on $(S,\mathcal{A},m)$.

According to Rosiński (1995), every stationary SαS process has a representation of the form

$$X_n=\int_S a_n(s)\left(\frac{d\,(m\circ\varphi^{-n})}{dm}(s)\right)^{1/\alpha} f\circ\varphi^n(s)\,M(ds) \tag{5.99}$$

for $n=\ldots,-1,0,1,2,\ldots$. Here M is a SαS random measure on $(S,\mathcal{A})$ with control measure m, $f\in L^\alpha(m)$, the functions (a_n) is a family of ±1-valued functions satisfying

$$a_{n+m}(s)=a_m(s)\,a_n\circ\varphi^m(s) \text{ for all integer } m,n\,, \tag{5.100}$$

$s\in S$. The functions (a_n) form a so-called *cocycle* for the flow (φ^n).

That is, apart from a cocycle and the Radon-Nykodim derivative, the kernel in the integral representation (5.99) consists of a single function f being "shifted" by the flow (φ^n). Of course, $f\circ\varphi^n(s)$ is just $f(\varphi^n(s))$.

Recall that a non-singular flow (φ^n) is called *conservative* if there is no wandering set of a positive measure m. That is, there no set $A\in\mathcal{A}$ such that that the sets $\varphi^{-n}A$, $n=0,1,\ldots$ are disjoint.

Given a non-singular flow (φ^n), there is a unique (up to a set of m-measure zero) partition of S into φ-invariant measurable sets C and D ($\varphi(C) = C$ and $\varphi(D) = D$), such that

- the restriction of (φ^n) to C is conservative
- D is either empty or $D = \cup_{n=-\infty}^{\infty} \varphi^n(A)$ for some wandering set A.

The decomposition $S = C \cup D$ as above is called the *Hopf decomposition*. See e.g. Krengel (1985).

A flow is conservative if $D = \emptyset$. If $C = \emptyset$ then the flow is called *dissipative*. Intuitively, conservative flows tend to come back to the initial point, while dissipative flows do not.

Given a stationary SαS process with an integral representation (5.99), if the flow (φ^n) is conservative, we say that the stationary process (X_n) is *generated by a conservative flow*. If the flow (φ^n) is dissipative, we say that the stationary process (X_n) is *generated by a dissipative flow*. It is known that if a stationary SαS process is generated by a conservative flow, then in any other integral representation of the form (5.99) it will be be represented by a conservative flow as well. Similarly with stationary SαS processes generated by a dissipative flow (Rosiński (1995)).

Since SαS random measures assign independent values to disjoint sets, it is, intuitively, clear that SαS processes generated by a conservative flow should have a longer memory than those generated by a dissipative flow. Indeed, if a process is generated by a conservative flow, then the same values of the random measure tend to contribute to the values of, say, X_0 and X_n for large n.

Suppose now that a general stationary SαS process is given by an integral representation (5.99). Let $S = C \cup D$ be the Hopf decomposition of S with respect to the flow (φ^n). Denote

$$X_n^{(1)} = \int_C a_n(s) \left(\frac{d\,(m \circ \varphi^{-n})}{d\,m}(s) \right)^{1/\alpha} f \circ \varphi^n(s)\, M(ds) \tag{5.101}$$

and

$$X_n^{(2)} = \int_D a_n(s) \left(\frac{d\,(m \circ \varphi^{-n})}{d\,m}(s) \right)^{1/\alpha} f \circ \varphi^n(s)\, M(ds)\,, \tag{5.102}$$

$n = \ldots, -1, 0, 1, 2, \ldots$. Notice that $(X_n^{(1)})$ and $(X_n^{(2)})$ are independent stationary SαS processes, generated, correspondingly, by a conservative flow and a dissipative flow, and that

$$X_n = X_n^{(1)} + X_n^{(2)}, \quad n = \ldots, -1, 0, 1, 2, \ldots. \tag{5.103}$$

Decomposition (5.103) represents a stationary SαS process as a sum of two independent stationary SαS processes, generated by a conservative flow and a dissipative flow, correspondingly. Such a decomposition is unique in distribution (Rosiński (1995)).

Therefore, to understand the length of memory in a stationary SαS process, one should start with trying to understand the length of memory in a stationary SαS processes generated by conservative flows, and those generated by dissipative flows.

It is known that the processes generated by dissipative flows are the so-called *mixed moving average processes.* They are always ergodic (even mixing) and have the form

$$X_n = \int_W \int_{I\!R} g(w, s-n)\, M(dw, ds), \tag{5.104}$$

for $n = \ldots, -1, 0, 1, 2, \ldots$, where $(W, \mathcal{W}, \nu)$ is a σ-finite measure space, M a SαS random measure on $W \times I\!R$ with control measure $\nu \times$ Leb, and $g \in L^\alpha(\nu \times \text{Leb})$. See Surgailis et al. (1993) and Rosiński (1995).

There is no known concise description of a similar kind of general stationary SαS processes generated by conservative flows, and these may or may not be ergodic. Intuitively, for a stationary SαS process generated by conservative flow to be ergodic the flow should not "come back too often". It also turns out that if such a process is ergodic, then the longer it takes for the flow "to come back", the shorter is the memory of the process. We will see these phenomena in the sequel.

Let us go back to studying the ruin probability

$$P_{\text{ruin}}(\lambda) = P\left(\sup_{n\geq 0} S_n > \lambda\right)$$

with

$$S_0 = 0,\ S_n = X_1 + \ldots + X_n - cn,\ n \geq 1\,.$$

We start with the case when the stationary SαS process (X_n) is generated by a dissipative flow. In that case the process is a mixed moving average with a representation (5.104). Define

$$J_\pm(w, s) = \lim_{h\downarrow -\infty} \liminf_{m\to\infty} \sup_{h\leq j\leq m} \left(\sum_{k=j}^{m} g(w, s+k)\right)_\pm$$

for $w \in W$ and $s \in I\!R$.

We have the following general result due to Mikosch and Samorodnitsky (2000a).

Theorem 28. *(a) For any mixed moving average process* (5.104) *the following lower bound for the ruin probability holds:*

$$\liminf_{\lambda\to\infty} \lambda^{\alpha-1} P_{\text{ruin}}(\lambda) \geq \frac{1}{2\alpha(\alpha-1)c}\, I(f)\,, \tag{5.105}$$

where

$$I(f) :=$$
$$\int_W \int_0^1 \left([J_+(w,s)]^\alpha + [J_-(w,s)]^\alpha\right)\, \nu(dw)\, ds\,.$$

(b) Assume that for ν-almost every $w \in W$ there is a compact interval $[K_l(w), K_r(w)]$ such that $0 < K_r(w) - K_l(w) \leq L$ for some finite constant L which does not depend on $w \in W$ and that $g(w,s) = 0$ for Leb-almost every $s \notin [K_l(w), K_r(w)]$. Then

$$\lim_{\lambda\to\infty} \lambda^{\alpha-1} P_{\text{ruin}}(\lambda) = \frac{1}{2\alpha(\alpha-1)c}\, I(f) < \infty\,. \tag{5.106}$$

That is, the ruin probability is always at least of the same order of magnitude as that in the case of iid claims (X_n) (see the Embrechts and Veraverbeke result (5.46)).

On the other hand, if s-sections of the kernel g are supported by compact intervals uniformly in w, then the ruin probability is exactly of the same order of magnitude as that in the case of iid claims (X_n). We, obviously, view it as saying that such stationary SαS processes have short memory. Below we will see what may happen if the s-sections of the kernel g are not compactly supported.

Let us go back to the example of one-sided linear fractional α-stable noise given by (5.75):

$$X_n = \int_{\mathbb{R}} g(s-n)\, M(ds),\ n = 1, 2, \dots ,$$

with

$$g(s) = (-s)_+^{H-1/\alpha} - (-s-1)_+^{H-1/\alpha},\ s \in \mathbb{R},$$

$0 < H < 1$. This is a moving average process; hence, it is generated by a dissipative flow. We saw previously that the memory of this process seems to become longer as H increases. It was not clear, however, where the boundary between long and short memory was. Things become clearer when we look at the ruin probabilities.

The following theorem, due to Mikosch and Samorodnitsky (2000a), describes the behavior of the ruin probabilities for linear fractional α-stable noises.

Theorem 29. *(a) If* $1/\alpha < H < 1$ *then*

(5.107) $$P_{\text{ruin}}(\lambda) \sim \frac{K(\alpha, H)}{c^{\alpha H}} \lambda^{-\alpha(1-H)}, \quad \lambda \to \infty .$$

(b) If $0 < H \le 1/\alpha$ *then*

(5.108) $$P_{\text{ruin}}(\lambda) \sim \frac{K(\alpha, H)}{c} \lambda^{-(\alpha-1)}, \quad \lambda \to \infty .$$

In both cases $K(\alpha, H)$ *is a finite positive constant.*

This result is an easy consequence of the equivalence (5.56). Note that the requirement (5.96) is satisfied, as in this case $m_n = \text{const}\, n^H$.

We see that Theorem 29 shows that

- If $0 < H \le 1/\alpha$, then the order of magnitude of the ruin probability does not change as H changes, and it is the same as in the case of iid claims (X_n).
- If $1/\alpha < H < 1$ then the order of magnitude of the ruin probability is large than that n the case of iid claims, and this order of magnitude changes as H changes.

Therefore, we have an indication of a phase transition as H crosses the boundary $1/\alpha$:

we view linear fractional α-stable noises with $0 < H \le 1/\alpha$ as having short memory

we view linear fractional α-stable noises with $1/\alpha < H < 1$ as having long memory

We conclude with examining what may happen with the ruin probability when a stationary SαS process is generated by a conservative flow. We will consider a class of such stationary SαS processes, described below. Consider an irreducible null-recurrent Markov chain on $\mathbb{Z} := \{-\ldots,-1,0,1,2,\ldots\}$ with law $P_i(\cdot)$ on

$$S = \{\mathbf{s} = (s_0, s_1, s_2, \ldots) : s_n \in \mathbb{Z},\ n = 0, 1, \ldots\}$$

corresponding to the initial state $s_0 = i \in \mathbb{Z}$. Let $\pi = (\pi_i)_{i\in\mathbb{Z}}$ be the σ-finite invariant measure corresponding to the family (P_i) satisfying $\pi_0 = 1$.

Let $\mathcal{A}$ be the cylindrical σ-field on S, and define a σ-finite measure on $(S, \mathcal{A})$ by

$$m(\cdot) = \sum_{i=-\infty}^{\infty} \pi_i\, P_i(\cdot)\,.$$

One can view m as the measure generated on the path space by the Markov chain starting according to the (infinite) initial invariant measure π. Observe that the measure m is invariant under the shift $\varphi:\ S \to S$:

$$\varphi((s_0, s_1, s_2, \ldots)) = (s_1, s_2, \ldots)$$

$\mathbf{s} = (s_0, s_1, s_2, \ldots) \in S$.

We consider a family of stationary SαS processes defined by (5.99), with the flow (φ^n) being the shift flow, the cocycle $a_n \equiv 1$ for all n, and a particularly simple choice of the kernel f. Since the measure m is invariant under the flow, the Radon-Nykodim derivative in (5.99) disappears, and we have

$$X_n = \int_S f \circ \varphi^n(s)\, M(ds) \tag{5.109}$$

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

We choose the kernel

$$f(\mathbf{s}) = \mathbf{1}_{\{s_0=0\}}(\mathbf{s})\,, \quad \mathbf{s} = (s_0, s_1, s_2, \ldots) \in S\,. \tag{5.110}$$

Then we obtain a stationary SαS process

$$X_n = \int_S \mathbf{1}\,(s_n = 0)\ M(ds)\,, \tag{5.111}$$

$n = 1, 2, \ldots$. This process is generated by a conservative flow. This follows from the fact that the Markov chain is recurrent. Moreover, this process is mixing.

However, if the Markov chain were positive recurrent, then the resulting SαS process would not even be ergodic (intuitively, the Markov chain would then return too often to its initial state). See Rosiński and Samorodnitsky (1996) for all of the above statements.

One can view the stationary SαS process in (5.111) as parameterized by a null-recurrent Markov chain. It is, then, natural to try to relate the length of memory of the SαS process to the properties of the Markov chain. By now we know that it is important to look at how quickly the Markov chain returns to its initial state. For a given $\mathbf{s} = (s_0, s_1, s_2, \ldots) \in S$, let

$$\tau = \tau(\mathbf{s}) = \inf\{n \geq 1 : s_n = 0\} \tag{5.112}$$

be the first return time to 0. Since the Markov chain is null recurrent, we must have $E_0\tau = \infty$.

Assume that there are $\gamma \in (0,1]$ and a slowly varying function L such that

$$P_0(\tau \geq n) = n^{\gamma-1}L(n)\,. \tag{5.113}$$

The parameter γ in (5.113) shows how long it takes the Markov chain to come back to its initial state.

Intuitively, this parameter also determines the length of memory of the corresponding stationary SαS process. From this point of view, small values of γ (close to 0) correspond to more frequent returns of the Markov chain and to longer memory of the SαS process. Note that, if γ crosses zero, then the Markov chain becomes positive recurrent, and the SαS process stops being ergodic.

It turns out that, at least as far as the rate of decay of the ruin probability is concerned, this intuition is correct. We have the following result.

Theorem 30. *Under the assumption (5.113),*

$$P_{\text{ruin}}(\lambda) \sim K(\alpha,\gamma)c^{\gamma(\alpha-1)-\alpha}\lambda^{-\gamma(\alpha-1)}L(\lambda)^{-(\alpha-1)} \tag{5.114}$$

as $\lambda \to \infty$. *Here* $K(\alpha,\gamma)$ *is a finite positive constant.*

This result is due to Mikosch and Samorodnitsky (2000a). From this result we see that:

- The ruin probability decays at a slower rate than in the case of iid claims (X_n).
- The rate of decay of the ruin probability decreases as γ decreases to zero.

This allows us to say that the class of stationary SαS processes in (5.111) has, under the assumption (5.113), long range dependence.

In fact, it may well be the case that under a proper point of view all (non-degenerate) ergodic stationary SαS processes generated by conservative flows have long range dependence. We do not know how to approach this question yet.

Notice, further, that stationary SαS processes generated by dissipative flows may or may not have long range dependence. In that case the length of memory appears to be determined by the asymptotic behavior of the kernel g in (5.104), as a function of s, for large values of s. In contrast, we saw that a stationary SαS process generated by a conservative flow can have long range dependence even if the kernel is as "nice" as that in (5.110).

Conclusion

The problem of long range dependence is an exciting and difficult one, and the approach of looking for phase transitions in the way rare events happen may bring new insights to this problem. Lots of work lies ahead.

References

R. Agrawal, A. Makowski and P. Nain (1999): On a reduced load equivalence for fluid queues under subexponentiality. *Queueing Systems. Theory and Applications* 33:5–41.

S. Asmussen (1987): *Applied Probability and Queues*. Wiley, Chichester, West Sussex, UK.

S. Asmussen, H. Schmidli and V. Schmidt (1999): Tail probabilities for non–standard risk and queueing processes with subexponential jumps. *Advances in Applied Probability* 31:422–447.

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.

F. Baccelli and P. Brémaud (1994): *Elements of Queueing Theory. Palm–Martingale Calculus and Stochastic Recurrences*. Springer–Verlag, Berlin.

S. Borst and B. Zwart (2001): Fluid queues with heavy-tailed $M/G/\infty$ input. Preprint.

M. Braverman (2002): Tail probabilities of subadditive functionals on stable processes with continuous and discrete time. Preprint.

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

D. Cline and G. Samorodnitsky (1994): Subexponentiality of the product of independent random variables. *Stochastic Processes and Their Applications* 49:75–98.

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

R. Dobrushin and P. Major (1979): Non-central limit theorems for non-linear functions of Gaussian fields. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 50:27–52.

P. Embrechts and C. Goldie (1980): On closure and factorization properties of subexponential distributions. *Journal of Australian Mathematical Society, Series A* 29:243–256.

P. Embrechts, C. Goldie and N. Veraverbeke (1979): Subexponentiality and infinite divisibility. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 49:335–347.

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

P. Embrechts and N. Veraverbeke (1982): Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance: Mathematics and Economics* 1:55–72.

W. Feller (1952): The asymptotic distribution of the range of sums of independent random variables. *Annals of Mathematical Statistics* 22:427–432.

D. Heath, S. Resnick and G. Samorodnitsky (1998): Heavy tails and long range dependence in on/off processes and associated fluid models. *Mathematics of Operations Research* 23:145–165.

D. Heath, S. Resnick and G. Samorodnitsky (1999): How system performance is affected by the interplay of averages in a fluid queue with long range dependence induced by heavy tails. *The Annals of Applied Probability* 9:352–375.

P. Jelenković, P. Momčilović and B. Zwart (2002): Reduced load equivalence under subexponentiality. Technical report.

U. Krengel (1985): *Ergodic Theorems*. De Gruyter, Berlin, New York.

J. Leslie (1989): On the non-closure under convolution of the subexponential family. *Journal of Applied Probability* 26:58–66.

B. Mandelbrot (1975): Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 31:271–285.

B. Mandelbrot and M. Taqqu (1979): Robust R/S analysis of long-run serial correlation. In *Proceedings of the 42nd Session of the International Statistical Institute*. Bulletin of the I.S.I., Manila, pp. 69–104. Vol.48, Book 2.

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.

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.

P. Moran (1964): On the range of cumulative sums. *Ann. Inst. Stat. Math.* 16:109–112.

V. Petrov (1995): *Limit Theorems of Probability Theory*. Oxford University Press, Oxford.

E. Pitman (1980): Subexponential distribution functions. *Journal of Australian Mathematical Society, Series A* 29:337–347.

N. Prabhu (1998): *Stochastic Storage Processes: Queues, Insurance risk, Dams and Data Communication*. Springer, New York.

Y. Prokhorov (1959): An extremal problem in probability theory. *Theory of Probability and Its Applications* 4:201–204.

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.

S. Resnick and G. Samorodnitsky (2001): Limits of on/off hierarchical product models for data transmission. Technical Report 1281, School of ORIE, Cornell University. Available at www.orie.cornell.edu/trlist/trlist.html.

M. Rosenblatt (1962): *Random Processes*. Oxford University Press, New York.

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

J. Rosiński (1995): On the structure of stationary stable processes. *The Annals of Probability* 23:1163–1187.

J. Rosiński and G. Samorodnitsky (1993): Distributions of subadditive functionals of sample paths of infinitely divisible processes. *Annals of Probability* 21:996–1014.

J. Rosiński and G. Samorodnitsky (1996): Classes of mixing stable processes. *Bernoulli* 2:3655–378.

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

D. Surgailis, J. Rosiński, V. Mandrekar and S. Cambanis (1993): Stable mixed moving averages. *Probab. Theory Related Fields* 97:543–558.

M. Taqqu (1975): Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 31:287–302.

M. Taqqu (1979): Convergence of integrated processes of arbitrary Hermite rank. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 50:53–83.

W. Whitt (1999): The reflection map is Lipschitz with appropriate Skorohod M-metrics. Preprint, AT&T Labs.

A. Zwart, B. Zwart and M. Mandjes (2000): Exact asymptotics for fluid queues fed by multiple heavy-tailed on-off flows. Preprint.

School of Operations Research and Industrial Engineering, Cornell University
E-mail address: gennady@orie.cornell.edu