RARE EVENTS AND HEAVY TAILS IN STOCHASTIC SYSTEMS

BERT ZWART

ABSTRACT. This is a set of lecture notes on large deviations and heavy tails in stochastic systems. After an introduction, a short and biased review on heavy-tailed distribution is given. Applications are mostly concerned with stochastic networks, and cover single-server queues, regenerative processes, and fluid models.

1. Introduction and Background

In this short course, we focus on several problems related to the following topics:

(1) Rare event analysis (in particular large deviations);
(2) Heavy tails;
(3) Applied Probability models arising in applications such as computer systems, communication networks and actuarial mathematics.

The assumed background is a first course on probability, and basic concepts related to Markov Chains and Poisson Processes. Renewal theory will be reviewed when necessary. In general, some mathematical maturity (especially with respect to analytic methods) is desirable. Advanced concepts such as measure theory, functional analysis, and general topology are not required.

The goal of this course is to get acquainted with several mathematical notions and concepts related to the above-mentioned three topics. Ideally, at the end of this course, you should have a good intuition about how rare events in stochastic systems take place when some of the random variables are heavy tailed. You will also get exposed to a range of proof techniques.

The goal of the first lecture is to give an overview of several models arising in applications, and to give an introduction to rare event analysis and heavy tails.

1.1. **Rare events.** Informally, an event is rare if it has small probability. The concept of "small" is vague, and depends on the application. Often, the event modeled is undesirable: for example an email that

gets lost, a supercomputer computation that unsuccessfully terminates, or an insurance company that is getting bankrupt. Formally, such an event is denoted A_n, where n is a parameter. Thus, we have a sequence of events $A_n, n \geq 1$, such that $P(A_n) \to 0$.

To get some appreciation for rare events, it is important to review the main limit theorems in probability theory, which include the *law of large numbers* (LLN) and the *central limit theorem.* Let $X_i, i \geq 1$ be a sequence of independent and identically distributed (i.i.d.) real-valued random variables, with $\mu = E[X_1]$ well defined. Let $S_0 = 0$ and for $n \geq 1$, let $S_n = X_1 + \ldots + X_n$.

1.1.1. *Laws of large numbers.* The strong law of large numbers (SLLN) states that

$$P(\lim_{n\to\infty} S_n/n = \mu) = 1, \tag{1}$$

we often say that $S_n/n \to \mu$ almost surely (a.s.). The weak law of large numbers (WLLN) states that

$$\lim_{n\to\infty} P(|S_n/n - \mu| > \epsilon) = 0 \tag{2}$$

for every $\epsilon > 0$. A proof of these theorems is not relevant for this course. The main point is that these theorems rigorously establish that

$$S_n \approx \mu n \text{ for } n \text{ large.} \tag{3}$$

1.1.2. *Central limit theorem.* If the variance $\sigma^2 = E[X_1^2] - \mu^2$ of X_1 is assumed to be finite, then it is possible to refine the weak law of large numbers. Let U be a random variable with a standard Normal distribution, i.e. U has density $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$. The CLT states

$$\frac{S_n - \mu n}{\sigma\sqrt{n}} \xrightarrow{d} U, \tag{4}$$

as $n \to \infty$.

(Recall that a sequence of random variables Z_n converges in distribution to Z if $P(Z_n \leq x) \to P(Z \leq x)$ for every x where $P(Z \leq x)$ is continuous; this is abbreviated as $Z_n \xrightarrow{d} Z$.)

Again, a proof of the CLT is beyond the scope of this course. The point which is relevant for us is that the CLT makes the heuristic statement

$$S_n \approx \mu n + O(\sqrt{n}) \text{ for } n \text{ large.} \tag{5}$$

rigorous. In fact, the term $O(\sqrt{n})$ can be written as $\sigma U \sqrt{n}$.

In short the CLT states that **typical deviations from the mean are of the order** $\sqrt{n}$.

1.1.3. *Large deviations; Cramérs theorem.* For random walks, it is natural to think of rare events in the following way: *Rare events are events that have too small probability to be explained by the central limit theorem.*

A key example is the following. Let $a > \mu$ and define $A_n = \{S_n \geq na\}$. It is clear that $P(A_n) \to 0$, but 0 is not a very useful approximation for such a probability. The question is therefore: what can we say about the convergence rate to 0?

Assuming that $P(X_1 > a) > 0$, we observe that

$$P(S_n \geq an) \geq P(X_i > a, i = 1, ..., n) = P(X_1 > a)^n. \tag{6}$$

Therefore, the convergence is at most exponentially fast (so it cannot go to zero at a speed of e^{-n^2} for example). This example is taken from [31]. To get an upper bound, verify that for any $\theta > 0$,

$$P(S_n \geq an) \leq E[e^{\theta X_1}]^n e^{-\theta an}. \tag{7}$$

If we optimize over θ we obtain (verify) the Chernoff bound, which states that

$$P(S_n \geq an) \leq e^{-nI(a)}, \tag{8}$$

with $I(a) = \sup_{\theta \geq 0}[\theta a - \log E[\exp\{\theta X_1\}]]$. That this upper bound is, in some sense, sharp follows from Cramér's theorem, which can be seen as a third main limit theorem (apart from the LLN and CLT), and can be seen as a starting point of large deviations theory. The term large deviations comes from the fact that we study deviations from the mean that are not of $O(\sqrt{n})$ (which is a normal deviation), but $O(n)$. Examples of texts on classical large deviations are [12, 18, 31].

In a simplified form, Cramérs theorem states that

$$\lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq an) = -I(a). \tag{9}$$

A proof of this theorem is covered later in this course. At this point, it only serve to illustrate that **often, large deviation probabilities converge to zero at exponential rate**. In this case, this rate is $I(a)$.

This is only meaningful if $I(a) > 0$, otherwise the bound $P(S_n \geq an) \leq e^{-nI(a)}$ is not informative. Using elementary convex analysis, it can be shown that $I(a) > 0$ if $a > \mu$ (a natural condition required to make the event A_n rare) and

$$E[e^{\theta X_1}] < \infty \text{ for some } \theta > 0. \tag{10}$$

This is equivalent to the condition that

$$P(X_1 > x) \leq Me^{-\epsilon x}, \tag{11}$$

for some $M < \infty$ and $\epsilon > 0$.

Therefore, if X_1 satisfies (10), we say that X_1 has a light (right) tail, or more informally (the distribution of) X_1 is light tailed. Sums and mixture of exponential distributions, the more general class of phase-type distributions, and distributions with bounded support are all light-tailed.

1.2. **Heavy tails.** If

$$E[e^{\theta X_1}] = \infty \text{ for all } \theta > 0, \tag{12}$$

we say that X_1 has a heavy (right) tail. More informally, we say that (the distribution of) X_1 is heavy tailed. An example of a heavy-tailed distribution is the Pareto distribution, determined by

$$P(X_1 > x) = x^{-\alpha}, x \geq 1, \tag{13}$$

with $\alpha > 0$ a constant. One can easily verify that (10) is not satisfied.

If X_1 is heavy-tailed, then $I(a) = 0$, and therefore the Chernoff bound is not meaningful. Large deviations theory for light tails heavily relies on the non-triviality of $I(a)$, and is therefore not applicable.

One of the first main results we shall establish is that in the Pareto example (13) $P(A_n) = O(n^{1-\alpha})$. This is fundamentally different from the exponential convergence rate for light tails.

Another fundamental difference is the most likely way the rare event A_n happens, given that it occurs. We shall analyze this most likely way, and will compare it with the light-tailed case.

1.2.1. *Relevance.* Why do we care about heavy tails? The answer is simple from a practical standpoint: statistical evidence shows that many real-life quantities can be modeled as a heavy-tailed random variable. Examples are:

- Income of an arbitrary household,
- File sizes, session lengths and other quantities in computer-communication networks,
- Claim sizes (think of the hurricane Andrew in 1992 or the September 11 attacks in 2001).
- Daily log-returns in financial markets.

Stochastic systems with heavy tails are difficult to simulate since rare events are do not have as small probability in conventional stochastic systems with light tails. In addition, heavy tails typically do not allow an explicit mathematical analysis of stochastic models, while light tails do (using sums and mixtures of exponential distributions). It is sometimes even hard to find a convenient Markovian description of the system of interest. Therefore, asymptotic techniques that analyze the frequency and nature of rare events in stochastic systems with heavy tails are not only of theoretical relevance. A key text on heavy tails with more real-life examples and a wide range of theoretical results is the monograph [15], focusing on applications in finance and insurance. The present lecture notes are more focused on stochastic networks. More references and background information can be found in [35].

1.3. **Some canonical models.**

1.3.1. *Ruin models.* Consider an insurance company with initial capital x. The premium rate of the insurance company is fixed (say c Euro per time unit). Thus, the total capital at time t would be $x + ct$ if no claims occur. We assume however, that claims do occur at a Poisson process with rate λ. The size of the ith claim is B_i. If $N(t)$ is the total number of claims up to time t, then the amount of capital $R(t)$ at time t is

$$R(t) = x + ct - \sum_{i=1}^{N(t)} B_i. \tag{14}$$

A key performance measure in actuarial mathematics is the function $p(x) = P(\inf_{t>0} R(t) < 0 \mid R(0) = x)$. The quantity $\tau(x) = \inf\{t : R(t) < 0\}$ is called the time until ruin. Thus, $p(x) = P(\tau(x) < \infty)$ is the probability that the insurance company gets ever ruined.

Let E_i be the time between two claims and set $X_i = B_i - cE_i$. From a practical standpoint, it is realistic to assume that the premium rate c is large enough so that $E[X_1] < 0$. If we define $S_0 = 0$ and, for $n \geq 1$, $S_n = X_1 + \ldots + X_n$, then it can be verified that

$$p(x) = P(\sup_{n\geq 0} S_n > x).$$

(Hint: ruin can only occur at claim arrival epochs.)

The distribution of the quantity $M = \sup_{n\geq 0} S_n$ can be explicitly computed if B_i has an exponential or Erlang distribution. For general distributions of the claim size (particularly heavy-tailed distributions), only the quantity $E[e^{-sM}]$ (i.e. the Laplace-Stieltjes Transform of the distribution of M) can be computed (assuming the claim sizes are i.i.d. and independent of the arrival process).

As we shall see during this course, it is possible to analyze the behavior of $p(x)$ as $x \to \infty$. A typical problem is how to choose x such that $p(x)$ is smaller than some acceptable level (say ϵ). If $p(x) \approx e^{-\gamma x}$, then one should take $x = \frac{-\log \epsilon}{\gamma}$. If $p(x) \approx x^{-\alpha}$, then one needs to choose $x = \epsilon^{-1/\alpha}$. It is not hard to see that $\epsilon^{-1/\alpha}$ is much larger than $-\log \epsilon$ is ϵ is small. Thus, if ϵ is small, then one needs a much higher safety capital when the claim sizes are not light tailed but heavy tailed. We refer to the textbooks [4, 5, 28] for more background on ruin models.

1.3.2. *The single server queue.* Consider the following model: customers arrive at a service facility according to a Poisson process with rate λ. Suppose that E_i is the time between the ith and $(i+1)$-st arriving customer. The ith customer has job size B_i. The server works at unit speed as long as there is work in the system. Let W_n be the total amount of work when customer n arrives. It can be shown that

$$W_{n+1} = (W_n + X_n)^+, \tag{15}$$

with $y^+ = \max\{0, y\}$ and $X_n = B_n - E_n$. If the sequence $X_n, n \geq 1$ is i.i.d. (which we assume), then it can be shown that the sequence $W_n, n \geq 0$ is a Markov chain. The recursion (15) is known as Lindley's recursion, and also describes the waiting time of the nth customer if the service discipline is First Come First Served (FCFS).

If $E[X_n] < 0$, then it can be shown that $W_n \xrightarrow{d} W$, with $W \stackrel{d}{=} (W + X_1)^+$. This does not depend on W_1. If $W_0 = 0$ (numbering customers from 0 onwards), then it can be shown (using duality theory for random walks) that

$$W_n \stackrel{d}{=} \sup_{k \in \{0,1,\ldots,n\}} S_k. \tag{16}$$

Thus,

$$W \stackrel{d}{=} \sup_{k \in \{0,1,\ldots\}} S_k. \tag{17}$$

We conclude that the waiting time tail probability can be expressed as a ruin probability. [4] is a good source of information on properties of W.

1.3.3. *Fluid models.* Modern communication networks transmit data that typically consists of small *packets*. For example, in ATM networks, files are broken into packets of 56 bits each. Since these packets are small and deterministic in size, it makes sense to model them as *fluid*. This leads to a fluid model, which is known as the Anick-Mitra-Sondhi (AMS) model [1].

An informal description of the model is as follows. Consider a fluid reservoir which can store infinite amount of fluid. At the bottom of the reservoir there is a hole, through which fluid can drain at rate c. The reservoir is being fed by N statistically identical sources. Each source is either "on" or "off", and is therefore called an on-off source. On-times are exponentially distributed with rate μ, and off-times are exponentially distributed with rate λ. Formally, for source n, $I_n(t) = 1$

if source n is on at time t and 0 otherwise. The process $I_n(t), t \geq 0$ is a two-state Markov chain, since the on and off-times are both exponential.

If a source is on, it feeds fluid into the buffer at rate r. The total amount of fluid offered to the buffer between time 0 and t is

$$A(t) = r \sum_{n=1}^{N} \int_0^t I_n(s)ds. \tag{18}$$

A key quantity of interest is the amount of fluid in the reservoir if the system is operating in steady state. It can be shown (this is beyond the scope of these notes) that

$$P(V > x) = P(\sup_{t>0}[A(t) - ct] > x). \tag{19}$$

Moreover, using the fact that $(V(t), I(t)), t \geq 0$, (with $I(t) = \sum_{i=1}^N I_n(t)$) is a Markov chain, the distribution of V can be computed. In fact, it is possible to find constants K_i and γ_i such that

$$P(V > x) = \sum_{i=1}^{N} K_i e^{-\gamma_i x}. \tag{20}$$

Although the method to compute this is interesting, it is not of much relevance for this course. The main point is that (i) light tails (in particular exponential tails) allow for a Markovian modeling and explicit analysis; (ii) $P(V > x)$ is decreasing to 0 at exponential rate (determined by the smallest of the constants γ_i). Both properties will be lost when the on-times will have a heavy-tailed distribution.

1.4. **Concluding remarks.** For all three models, we shall investigate the tail behavior of probability of interest when the underlying distributions are heavy tailed. To do this, we first need to develop the theory of heavy-tailed distributions.

2. Heavy tailed distributions and regular variation

The first main topic of this course is concerned with several subclasses of heavy-tailed distributions. Subclasses are necessary, since the full class of heavy-tailed distributions appears to be too big to build a decent theory. Three main subclasses of heavy-tailed distributions are the so-called long-tailed distributions, subexponential distributions and regularly varying distributions. The theory of regular variation (which is a beautiful piece of analysis) plays a fundamental role in the development of the theory. This chapter is heavily influenced by the monograph [15], and also builds on the monumental treatise [7], as well as the queueing-oriented survey [32].

2.1. **Regularly varying functions.** A real-valued function L is *slowly varying* if for every $a > 0$,

$$L(ax)/L(x) \to 1 \tag{21}$$

as $x \to \infty$. Examples of slowly varying functions are constants, logarithms and iterated logarithms. The function $\sin x$ is not slowly varying. A real-valued function f is called *regularly varying* of index β if

$$f(ax)/f(x) \to a^{\beta} \tag{22}$$

for every $a > 0$. We see that regularly varying functions of index 0 are slowly varying. In fact, in general we can always write a regularly varying function $f(x)$ in the form $x^{\beta}L(x)$, with L a slowly varying function. Some remarks:

- The definitions can be weakened: convergence only needs to hold for some a in an open interval in $(0, \infty)$, and the limit function a^{β} can in principle be replaced by a more general function $g(a)$ (under regularity assumptions such as measurability it can be shown that $g(a) = a^{\beta}$ is the only possible limit.
- The book "regular variation" of Bingham, Goldie, Teugels (BGT) [7] is the main reference text. Since these notes focus on probability and not analysis, we will take some of the basic results as given. Proofs can be found in BGT.

We now list two main fundamental properties of slowly varying functions:

- *Uniform convergence theorem (UCT)*: the convergence (21) is always uniform on compact sets in $(0, \infty)$.
- *representation theorem:* L is slowly varying if and only if there exist a constant x_0, and functions c and δ such that $c(x) \to c \in (0, \infty)$, $\delta(x) \to 0$, and

$$L(x) = c(x) \exp\{\int_{x_0}^{x} \frac{\delta(u)}{u} du\}, x > x_0. \tag{23}$$

It is not difficult to check the if-statement in the second result. The "only if" statement is more involved, and relies on the UCT. It is sometimes useful to rewrite the representation for L as

$$L(x) = c(x) \exp\{\int_{\log x_0}^{\log x} \bar{\delta}(y) dy\}, \tag{24}$$

with $\bar{\delta}(y) = \delta(e^y)$. From this representation, it is not difficult to see that for every $\epsilon > 0$ there exist x_ϵ such that

$$x^{-\epsilon} \leq L(x) \leq x^{\epsilon}, x > x_\epsilon. \tag{25}$$

A third main result is *Karamata's theorem.* For pure power functions, we have (for $\alpha > 1$) the identity

$$\int_x^\infty u^{-\alpha} du = \frac{1}{\alpha - 1} x^{1-\alpha}. \tag{26}$$

Karamata's theorem says that

$$\int_x^\infty L(u) u^{-\alpha} du \sim \frac{1}{\alpha - 1} L(x) x^{1-\alpha}, \tag{27}$$

with $f(x) \sim g(x)$ denoting that $f(x)/g(x) \to 1$ as $x \to \infty$.

A remarkable fact is that, if L satisfies (27), then L must be slowly varying. Thus, slowly varying functions are precisely those functions that can be treated as constants in the asymptotic evaluation of integrals (since we can interchange L and the integral). I encourage you to consult BGT for a proof of this deep fact. In this course, we are mainly concerned with applications of results from regular variation to probability theory.

2.2. **Long-tailed random variables.** From now on, unless stated otherwise, all random variables that are introduced are defined on the same probability space, are non-negative and independent of each other.

Let X be a random variable with distribution function F and tail $\bar{F} = 1 - F$. We say that X is *long-tailed* (and write $X \in \mathcal{L}$) if

$$\frac{P(X > x + y)}{P(X > x)} \to 1 \tag{28}$$

as $x \to \infty$ for every $y > 0$. An equivalent way to write this is

$$P(X > x + y \mid X > x) \to 1. \tag{29}$$

Thus, if a long-tailed random variable exceeds a large value x, it is likely that it exceeds an even larger value $x + y$ as well. Verify yourself, using the memoryless property, that exponential random variables are clearly not long-tailed.

If $X \in \mathcal{L}$, then we can write $\bar{F}(x) = L(e^x)$ for a slowly varying function L. By the representation (24) for slowly varying functions, we can write

$$\bar{F}(x) = L(e^x) = c(e^x) \exp\{\int_{\log x_0}^{x} \bar{\delta}(y) dy\}, \tag{30}$$

from which it follows that

$$e^{\epsilon x}\bar{F}(x) \to \infty \tag{31}$$

for any $\epsilon > 0$. Thus, the class of long-tailed distributions is a subclass of the heavy-tailed distributions. Still, the class of long-tailed random variables seems too large for applications. Therefore, we assume an additional regularity condition, leading to subexponential random variables.

2.3. **Subexponential random variables.** A random variable X is *subexponential* if for two independent copies X_1, X_2 of X,

$$\frac{P(X_1 + X_2 > x)}{P(X_1 > x)} \to 2, \tag{32}$$

as $x \to \infty$. We write $X \in \mathcal{S}$. Note that

$$P(\max\{X_1, X_2\} > x) \sim 2P(X_1 > x), \tag{33}$$

and that $X_1 + X_2 > \max\{X_1, X_2\}$. Thus, X is subexponential, if the inequality in

$$P(X_1 + X_2 > x) \geq P(\max\{X_1, X_2\} > x) \tag{34}$$

can be replace by "$\sim$", and

$$P(X_1 + X_2 > x; \max\{X_1, X_2\} < x) = o(P(X_1 > x)). \tag{35}$$

Thus, under subexponentiality, a large sum is most likely due to a large maximum.

To understand the difference between $\mathcal{L}$ and $\mathcal{S}$, note that

$$\frac{P(X_1 + X_2 > x)}{P(X_1 > x)} = 1 + \int_0^x \frac{\bar{F}(x-t)}{\bar{F}(x)} dF(t). \tag{36}$$

So subexponentiality follows from long-tailedness once it is possible to justify the interchange of limit and integration. There exist examples of random variables that are long-tailed, but not subexponential. These examples are, however, rather pathological.

2.3.1. *Properties.* We now list some properties of subexponential distributions. Let F^{n*} be the n-fold convolution of F with itself. The proofs in this subsection are taken from [15], which also provides a detailed bibliographic account.

Proposition 1. *If $X \in \mathcal{S}$, then $\bar{F}^{n*}(x)/\bar{F}(x) \to n$. Thus,*

$$P(X_1 + \ldots + X_n > x) \sim P(\max\{X_1, \ldots, X_n\} > x). \tag{37}$$

Proof. We use induction in n. Write

$$\frac{\bar{F}^{(n+1)*}(x)}{\bar{F}(x)} = 1 + \frac{F(x) - F^{(n+1)*}(x)}{\bar{F}(x)} = 1 + \int_0^x \frac{1 - F^{n*}(x-t)}{\bar{F}(x)} dF(t). \tag{38}$$

Fix y, and write the last integral as $I_1(x) + I_2(x)$, with

$$I_1(x) = \int_0^{x-y} \frac{\bar{F}^{n*}(x-t)}{\bar{F}(x-t)} \frac{\bar{F}(x-t)}{\bar{F}(x)} dF(t), \tag{39}$$

$$I_2(x) = \int_{x-y}^{x} \frac{\bar{F}^{n*}(x-t)}{\bar{F}(x-t)} \frac{\bar{F}(x-t)}{\bar{F}(x)} dF(t). \tag{40}$$

Observe that

$$I_1(x) = \int_0^{x-y} \left(\frac{\bar{F}^{n*}(x-t)}{\bar{F}(x-t)} - n \right) \frac{\bar{F}(x-t)}{\bar{F}(x)} dF(t) + n \int_0^{x-y} \frac{\bar{F}(x-t)}{\bar{F}(x)} dF(t).$$

The first term can be made arbitrarily small if y is large, by the induction hypothesis, so we have

$$I_1(x) = (o_y(1) + n) \int_0^{x-y} \frac{\bar{F}(x-t)}{\bar{F}(x)} dF(t). \tag{41}$$

The lower limit of the integral (as $x \to \infty$) is at least 1 since $X \in \mathcal{L}$. For the upper limit, observe that

$$\int_0^{x-y} \frac{\bar{F}(x-t)}{\bar{F}(x)} dF(t) = \frac{F(x) - F^{2*}(x)}{\bar{F}(x)} - \int_{x-y}^{x} \frac{\bar{F}(x-t)}{\bar{F}(x)} dF(t) \tag{42}$$

The first term converges to 1 by the definition of $\mathcal{S}$. The second integral is non-negative, so the upper limit is 1 as well. Thus, letting first $x \to \infty$ and then $y \to \infty$ we see that $I_1(x) \to n$. The second integral $I_2(x)$ can be upper bounded by

$$I_2(x) \leq \sup_{t \in [x-y, x]} \left(\frac{\bar{F}^{n*}(x-t)}{\bar{F}(x-t)} \right) \int_{x-y}^{x} \frac{\bar{F}(x-t)}{\bar{F}(x)} dF(t). \tag{43}$$

The first factor is bounded by $1/\bar{F}(y)$, while the second factor is bounded by

$$(F(x) - F(x-y))/\bar{F}(x) = \bar{F}(x-y) - \bar{F}(x)/\bar{F}(x) \to 0.$$

for every y, since $X \in \mathcal{L}$. We conclude that $\bar{F}^{(n+1)*}(x)/\bar{F}(x) \to n + 1$. □

The next proposition establishes that $\mathcal{S}$ is included in $\mathcal{L}$.

Proposition 2. *If X is subexponential, it is also long-tailed.*

Proof. Take y fixed.

$$\frac{\bar{F}^{2*}(x)}{\bar{F}(x)} = 1 + \int_0^y \frac{\bar{F}(x-t)}{\bar{F}(x)} dF(t) + \int_y^x \frac{\bar{F}(x-t)}{\bar{F}(x)} dF(t).$$

The first integral can be lower bounded by $F(y)$, while the second integral can be lower bounded by $(\bar{F}(x-y)/\bar{F}(x))(F(x)-F(y))$. Combining these bounds we obtain

$$1 \leq \frac{\bar{F}(x-y)}{\bar{F}(x)} \leq \left(\frac{\bar{F}^{2*}(x)}{\bar{F}(x)} - 1 - F(y)\right)(F(x)-F(y))^{-1}.$$

The right-hand side converges to 1 as $x \to \infty$ since X is subexponential. □

The third proposition can be a useful bound in proofs.

Proposition 3. *If X is subexponential, then its distribution function F satisfies the following: for every $\epsilon > 0$ there exists $K < \infty$ such that for every $n \geq 2$ and $x \geq 0$:*

$$\frac{\bar{F}^{n*}(x)}{\bar{F}(x)} \leq K(1+\epsilon)^n \tag{44}$$

Proof. Define $\alpha_n = \sup_{x \geq 0} \frac{\bar{F}^{n*}(x)}{\bar{F}(x)}$ and write

$$\alpha_{n+1} = 1 + \sup_{x \geq 0} \int_0^x \frac{\bar{F}^{n*}(x-t)}{\bar{F}(x)} dF(t).$$

Fix T and observe that the right-hand side of the above display is smaller than

$$1 + \sup_{x \in [0,T]} \int_0^x \frac{\bar{F}^{n*}(x-t)}{\bar{F}(x)} dF(t) + \sup_{x > T} \int_0^x \frac{\bar{F}^{n*}(x-t)}{\bar{F}(x)} dF(t).$$

The first supremum is smaller than $1/\bar{F}(T)$. For the second supremum, observe that

$$\begin{aligned}\sup_{x>T} \int_0^x \frac{\bar{F}^{n*}(x-t)}{\bar{F}(x)} dF(t) &= \sup_{x>T} \int_0^x \frac{\bar{F}^{n*}(x-t)}{\bar{F}(x-t)} \frac{\bar{F}(x-t)}{\bar{F}(x)} dF(t) \\ &\leq \alpha_n \sup_{x>T} \frac{F(x) - F^{2*}(x)}{\bar{F}(x)}.\end{aligned}$$

By subexponentiality, we can choose T so large that the last supremum is smaller than $1+\epsilon$. Putting everything together, we obtain

$$\alpha_{n+1} \leq 1 + \frac{1}{\bar{F}(T)} + \alpha_n(1+\epsilon).$$

From this inequality one can verify inductively that

$$\alpha_n \leq \left(1 + \frac{1}{\bar{F}(T)}\right)\frac{1}{\epsilon}(1+\epsilon)^n,$$

which completes the proof by putting $K = \left(1 + \frac{1}{\bar{F}(T)}\right)\frac{1}{\epsilon}$. □

The next property shows that the class $\mathcal{S}$ is closed under tail-equivalence.

Proposition 4. *Let $X \in \mathcal{S}$ and let Y be such that $P(Y > x) \sim CP(X > x)$ for some $C > 0$. Then $\mathcal{Y} \in \mathcal{S}$ as well.*

Proof. Let G be the distribution function of Y and let Y_1, Y_2 be two iid copies of Y. Note that $Y \in \mathcal{L}$ since

$$\bar{G}(x+y) \sim C\bar{F}(x+y) \sim C\bar{F}(x) \sim C\bar{G}(x).$$

Also, note that for $x > 2v$ and v a constant,

$$\begin{aligned}\{Y_1 + Y_2 > x\} &= \{Y_1 < v, Y_1 + Y_2 > x\} \cup \{Y_2 < v, Y_1 + Y_2 > x\} \\ &\cup \{v < Y_1 < x - v, Y_1 + Y_2 > x\} \cup \{Y_2 > v, Y_1 > x - v\}.\end{aligned}$$

The four events on the r.h.s. are disjoint. Therefore

$$\frac{\bar{G}^{2*}(x)}{\bar{G}(x)} = 2\int_0^v \frac{\bar{G}(x-y)}{\bar{G}(x)}dG(y) + \int_v^{x-v}\frac{\bar{G}(x-y)}{\bar{G}(x)}dG(y) + \frac{\bar{G}(x-v)}{\bar{G}(x)}\bar{G}(v). \tag{45}$$

Number the integrals as I, II, III. We see that the first integral converges to $2G(v)$ and the last integral converges to $\bar{G}(v)$. To bound the second integral, note that for any $\epsilon > 0$ there exists v such that $\bar{G}(x) \geq (c-\epsilon)\bar{F}(x)$ and $\bar{G}(x-y) \leq (c+\epsilon)\bar{F}(x-y)$ for all y in the integration range. Thus,

$$II \leq \frac{c+\epsilon}{c-\epsilon}\int_v^{x-v}\frac{\bar{F}(x-y)}{\bar{F}(x)}dG(y).$$

Write $dG(y) = -d\bar{G}(y)$ and apply partial integration. The integral on the right-hand side (without the prefactor) is equal

$$\frac{\bar{F}(x-y)}{\bar{F}(x)}\bar{G}(v) - \frac{\bar{F}(v)}{\bar{F}(x)}\bar{G}(x-v) + \int_v^{x-v}\frac{\bar{G}(x-t)}{\bar{F}(x)}dF(t).$$

The first term converges to $\bar{G}(v)$ and the second to $C\bar{F}(v)$. In the third term, bound $\bar{G}(x-t)$ with $(c+\epsilon)\bar{F}(x-t)$, and conclude from subexponentiality of X that

$$\int_v^{x-v}\frac{\bar{F}(x-t)}{\bar{F}(x)}dF(t) \to (2 - 2F(v)).$$

Putting everything together, we see that

$$\limsup_{x\to\infty} II \leq \bar{G}(v) + C\bar{F}(v) + (C+\epsilon)(2-2F(v)).$$

This term can be made arbitrarily small by letting v become arbitrarily large. Since also term III vanishes for large v, the only contribution that remains in (45) is term I, which converges to 2 for large v. We conclude that

$$\limsup_{x\to\infty} \frac{\bar{G}^{2*}(x)}{\bar{G}(x)} \leq 2.$$

□

2.3.2. *Application to random sums.* Let N be a random variable independent of the *i.i.d.* sequence $X_1, X_2, \ldots$. Define the random sum $Z = X_1 + \ldots X_N$. What can we say about the tail behavior of Z if $X_1 \in \mathcal{S}$?

Proposition 5. *Under the above conditions, if N is also such that $E[(1+\epsilon)^N] < \infty$ for some $\epsilon > 0$, then*

$$\lim_{x\to\infty} \frac{P(Z>x)}{P(X_1>x)} = E[N].$$

Proof. Write $p_n = P(N=n)$ and observe that

$$\frac{P(Z>x)}{P(X_1>x)} = \sum_{n=0}^{\infty} p_n \frac{\bar{F}^{n*}(x)}{\bar{F}(x)}.$$

To evaluate the limit as $x \to \infty$, it is tempting to interchange the summation and the limit, since $\frac{\bar{F}^{n*}(x)}{\bar{F}(x)} \to n$ for every fixed n. By Proposition 3, there exists a constant K such that $\frac{\bar{F}^{n*}(x)}{\bar{F}(x)} \leq K(1+\epsilon)^n$, and by assumption, $E[(1+\epsilon)^N] < \infty$. This enables us to apply the dominated convergence theorem. □

2.3.3. *First example: regularly varying tails.* So far, we have made several additional assumptions, restricting heavy tails to long tails, and even to subexponential tails. A first piece of evidence that the class of subexponential random variables is still big enough to include many appealing examples is provided here. Assume that $P(X>x) = \bar{F}(x) = L(x)x^{-\alpha}$, with L a slowly varying function.

We will show that X is also subexponential. Since $X_1 + X_2 \geq \max\{X_1, X_2\}$, it always holds that

$$\liminf_{x\to\infty} \frac{P(X_1+X_2>x)}{P(X_1>x)} \geq 2, \tag{46}$$

so to prove subexponentiality, it suffices to show that

$$\limsup_{x\to\infty} \frac{P(X_1+X_2>x)}{P(X_1>x)} \leq 2. \tag{47}$$

Let X_1 and X_2 be two independent copies of X and observe that for $\delta \in (0,1/2)$,

$$\{X_1+X_2>x\} \subseteq \{X_1>(1-\delta)x\} \cup \{X_2>(1-\delta)x\} \cup \{X_1>\delta x, X_2>\delta x\}.$$

Thus,

$$P(X_1+X_2>x) \leq 2P(X>(1-\delta)x) + P(X>\delta x)^2.$$

Using this bound we obtain

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

This holds for any $\delta \in (0,1/2)$. The property (47) now follows by taking $\delta \downarrow 0$. This example is taken from [17], which contains a section on regular variation in probability theory.

2.3.4. *How to detect if a distribution is subexponential?* For regularly varying tails, we saw that it was possible to show that subexponentiality holds by verifying the condition directly. In general this can be problematic though, and it is therefore convenient to have other means. The proposition below, due to Pitman [26] is therefore very useful.

To get into Pitman's framework, we need to assume that X is a continuous random variable with a density f. Recall that $q(x) = f(x)/\bar{F}(x)$ is the hazard rate, and note that

$$\bar{F}(x) = \exp\{-\int_0^x q(u)du\}.$$

So $q(u) \to 0$ is F is long-tailed.

Proposition 6. *Suppose that $q(x)$ is eventually decreasing and that $q(x) \to 0$ if $x \to \infty$.*
(a) X is subexponential if and only if $\int_0^x e^{yq(x)} f(y)dy \to 1$.
(b) If $e^{xq(x)} f(x)$ is integrable at ∞ then X is subexponential.

A proof of this proposition can be found in, for example [15]. Below, we list some applications.

I. Weibull
Suppose $\bar{F}(x) = e^{-cx^\tau}$ with $c>0$ and $0<\tau<1$. The density is

$f(x) = c\tau x^{\tau-1}e^{-cx^\tau}$. The hazard rate is therefore $q(x) = c\tau x^{\tau-1}$.

We now have all data to apply part (b) of the above proposition. Note that, in our case

$$e^{xq(x)}f(x) = e^{c(\tau-1)x^\tau}c\tau x^{\tau-1}.$$

This function is integrable since $\tau < 1$.
Question: what can we say about the example $\bar{F}(x) = e^{-c\lfloor x\rfloor^\tau}$?

II. "Almost exponential"
We know that the exponential distribution is not even long-tailed, so it cannot be subexponential. Consider the example

$$\bar{F}(t) = e^{-t/\ln t}.$$

The density in this case is

$$f(x) = \left(\frac{1}{\ln x} - \frac{1}{(\ln x)^2}\right)e^{-x/\ln x}.$$

The hazard rate is therefore equal to

$$q(x) = \frac{1}{\ln x} - \frac{1}{(\ln x)^2}.$$

This is clearly converging to 0 and also can be shown to be eventually decreasing. We see that

$$e^{xq(x)}f(x) = e^{-x/(\ln x)^2},$$

which is integrable.

III. Lognormal
Another example that can be verified is the lognormal distribution. In this case, $\bar{F}(x) = \Phi(\exp\{x\})$. Unfortunately, Φ is not very explicit, but one can show that $\Phi(x) \sim \phi(x)/x$, with $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$. One can combine this with Proposition 4 to show that the lognormal is subexponential. This will be left as an exercise.

2.3.5. *Integrated tails.* Suppose that $X \geq 0$ has finite mean μ. We can define a random variable X^r which has density $P(X > x)/\mu$. The distribution tail is given by

$$\bar{F}^r(x) = \frac{1}{\mu}\int_x^\infty P(X > u)du,$$

and it called the integrated tail.

What can we say about X^r is X is long-tailed, subexponential or regularly varying?

It turns out that X^r is long-tailed if X is long-tailed, in which case $\bar{F}^r(x)/\bar{F}(x) \to \infty$. By using Karamata's theorem, one can show that X^r is regularly varying with index $1-\alpha$ if X is regularly varying with index $-\alpha$ ($\alpha > 1$).

Does the same property holds for $\mathcal{S}$? The answer is no, but it took until 2004 until a counterexample has been published by Denisov, Foss and Korshunov [14].

Claudia Klüppelberg [23] proposed the subclass $\mathcal{S}^*$ of $\mathcal{S}$, defined as follows: $X \in \mathcal{S}^*$ if $\mu = E[X] < \infty$ and

$$\lim_{x\to\infty} \int_0^x \frac{\bar{F}(x-y)\bar{F}(y)}{\bar{F}(x)} = 2\mu.$$

The following results are stated without proof, for a proof, see [23].

Proposition 7. *If $X \in \mathcal{S}^*$, then $X \in \mathcal{S}$ and $X^r \in \mathcal{S}$*

Proposition 8. *Let X have a hazard rate $q(u)$ which eventually decreasing to 0. Then $X \in \mathcal{S}$ if and only if*

$$\lim_{x\to\infty} \int_0^x e^{yq(x)} \bar{F}(y)dy = 2\mu.$$

There exist much more properties and examples of subexponential distributions, for which we refer to [15] and references therein. We now move on to some concrete applications.

3. Random walks

Let $S_n, n \geq 0$ be the standard random walk with increments X_n. Let X_1 be regularly varying with index $-\alpha$, $\alpha > 1$. We know that, if $x \to \infty$ and n fixed, then $P(S_n > x)/nP(X_1 > x) \to 1$. The question is whether n can be allowed to be varied with x. In particular, what can we say about the behavior of $P(S_n > an)$ as $n \to \infty$ with $a > \mu$?

Heuristics

It is reasonable to expect that one of the n X_i's is large, while the other ones have a normal value. By symmetry, we get

$$\begin{aligned} P(S_n > an) &\approx nP(X_1 > cn, X_2 + \ldots + X_n \approx (n-1)\mu) \\ &\approx nP(X_1 > cn) \end{aligned}$$

where c should be such that $cn + n\mu = an$, so $c = \mu - a$. That this intuition is correct will be proven, following the proof in [29].

Theorem 1. *Suppose* $P(X_1 > x) = L(x)x^{-\alpha}$, $\alpha > 1$. *For* $\epsilon > 0$, *as* $n \to \infty$,

$$P(S_n > (\mu + \epsilon)n) \sim nP(X_1 > \epsilon n). \tag{48}$$

Proof. The proof consist of the construction of an asymptotic lower bound and an asymptotic upper bound, which coincide. We start with the lower bound. Let $\delta > \epsilon$. Observe that

$$P(S_n > (\mu+\epsilon)n) \geq P(\cup_{i=1}^n \{X_i > \delta n, \sum_{j=1, j\neq i}^{n} X_j > n(\mu+\epsilon-\delta)\}) \tag{49}$$

Define $B_i = \{X_i > \delta n, \sum_{j=1,j\neq i}^n X_j > n(\mu + \epsilon - \delta)\}$. Observe that $P(B_i) = P(B_j)$. From (49) we obtain that

$$\begin{aligned} P(S_n > (\mu+\epsilon)n) &\geq P(\cup_{i=1}^n B_i) \\ &\geq \sum_{i=1}^n P(B_i) - 2\sum_{i<j} P(B_i \cap B_j) \\ &= nP(B_1) - n(n-1)P(B_1 \cap B_2). \end{aligned}$$

Since

$$P(B_1 \cap B_2) \leq P(X_1 > \epsilon n \cap X_2 > \epsilon n) = P(X_1 > \epsilon n)^2,$$

we see that, by decomposing $P(B_1)$,

$$P(S_n > (\mu+\epsilon)n) \geq nP(X_1 > \delta n)P(\sum_{j=2}^n X_j > n(\mu+\epsilon-\delta)) - n(n-1)P(X_1 > \epsilon n)^2. \tag{50}$$

By applying the weak law of large numbers, we see that

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

since $\delta > \epsilon$. In addition, it can be shown that $n(n-1)P(X_1 > \epsilon n)^2$ is regularly varying with index $2 - 2\alpha$, while $nP(X_1 > \delta n)$ is regularly varying with index $1 - \alpha$. Since $\alpha > 1$ it follows that $n(n-1)P(X_1 > \epsilon n)^2 = o(nP(X_1 > \delta n))$. Putting everything together we see from (50) that

$$P(S_n > (\mu + \epsilon)n) \geq nP(X_1 > \delta n)(1 - o(1)) - o(nP(X_1 > \delta n)).$$

Therefore

$$\liminf_{x\to\infty} \frac{P(S_n > (\mu+\epsilon)n)}{P(X_1 > \epsilon n)} \geq \left(\frac{\epsilon}{\delta}\right)^\alpha .$$

This holds for every $\delta > \epsilon$. By letting $\delta \downarrow \epsilon$, we conclude that

$$\liminf_{x\to\infty} \frac{P(S_n > (\mu+\epsilon)n)}{P(X_1 > \epsilon n)} \geq 1,$$

concluding the proof of the lower bound.

The proof of the upper bound is more involved. Define $A_n = \{S_n > (\mu+\epsilon)n\}$ and let $\tau \in (0,\epsilon)$. Write

$$\begin{aligned} P(A_n) &= P(A_n; \max_{i=1,\dots,n} X_i > \tau n) + P(A_n; \max_{i=1,\dots,n} X_i \leq \tau n) \\ &:= P(A_{1,n}) + P(A_{2,n}). \end{aligned}$$

Suppose we can prove that

$$\lim_{n\to\infty} \frac{P(A_{2n})}{nP(X_1 > \epsilon n)} = 0 \tag{51}$$

for a suitable choice of $\tau \in (0,\epsilon)$. We will prove this separately in a lemma below. Then we can focus on $P(A_{1,n})$. Take $\delta \in (\tau,\epsilon)$ and observe that, since $\tau < \delta$,

$$\begin{aligned} P(A_{1,n}) &\leq P(\cup_{i=1}^n \{X_i > \delta n\}) + P(\cup_{i=1}^n \{X_i > \tau n, \sum_{j=1, j\neq i}^n X_j > n(\mu+\epsilon-\delta)\}) \\ &\leq nP(X_1 > \delta n) + nP(X_1 > \tau n) P(\sum_{j=2}^n X_j > n(\mu+\epsilon-\delta)). \end{aligned}$$

By the weak law of large numbers, we see that $P(\sum_{j=2}^n X_j > n(\mu+\epsilon-\delta)) \to 0$, so that

$$\limsup_{n\to\infty} \frac{P(A_{1n})}{nP(X_1 > \delta n)} \leq \left(\frac{\epsilon}{\delta}\right)^\alpha. \tag{52}$$

Now take $\delta \uparrow \epsilon$ and use (51) to conclude that

$$\limsup_{n\to\infty} \frac{P(A_n)}{nP(X_1 > \delta n)} \leq 1. \tag{53}$$

□

Thus, our theorem is proven once we have shown the validity of (51). For this, we need the following Theorem, which comes from Prohorov [27].

Theorem 2. *Let $Y_i, i \geq 1$ be an i.i.d. sequence such that $E[Y_i] = 0$ and $P(Y_i \leq c) = 1$ for some $c \in (0,\infty)$. For any $t > 0$ and any $n \geq$*

the following inequality holds:

$$P(Y_1+\ldots+Y_n>t)\le \exp\{-\frac{t}{2c}\log\left(\frac{c\lambda}{nVar(Y_1)}\right)\}.$$

We are going to apply this result to prove (51). Although the original random variables X_i are unbounded, they are bounded from above by τx in the event A_{2n}, so we can apply Prohorov's result with $c=\tau x$.

of (51). We first rewrite/bound $P(A_{2,n})$ into a conditional probability.

Let $X_i^*, i\ge 1$ be an i.i.d. sequence such that $P(X_1^*\le y)=P(X_i\le y\mid X_i\le \tau n)$. Observe that $\mu^*=E[X_1^*]\le\mu$. Write

$$P(A_{2n})=P(S_n>(\mu{+}\epsilon)n\mid X_1\le\tau n,\ldots,X_n\le\tau n)/P(X_1\le\tau n,\ldots,X_n\le\tau n).$$

Observe that $P(X_1\le\tau n,\ldots,X_n\le\tau n)\to 1$ as $n\to\infty$. So, to prove (51), it suffices to show that

$$P(S_n>(\mu+\epsilon)n\mid X_1\le\tau n,\ldots,X_n\le\tau n)=o(nP(X_1>\epsilon n)). \quad (54)$$

The probability on the left hand side is equal to

$$P(X_1^*+\ldots+X_n^*>(\mu+\epsilon)n)\le P((X_1^*-\mu^*)+\ldots+(X_n^*-\mu_*)>\epsilon n).$$

Apply now the theorem of Prohorov with $Y_i=X_i^*-\mu^*$. Note that $Y_i\le\tau n=:c$

To determine an upper bound for $Var(Y_1)$ we proceed as follows: since there exists a $\beta\in(1,\alpha)$ such that $\mu_\beta=E[X_1^\beta]<\infty$ we have $P(X_1>u)\le\mu_\beta u^{-\beta}$. Now, observe that

$$\begin{aligned} Var(Y_1) &= Var(X_1^*)\\ &\le E((X_1^*)^2)\\ &= \frac{1}{P(X_1<\tau n)}\int_0^{\tau n} 2uP(u<X_1<\tau n)du\\ &\le \frac{2\mu_\beta}{P(X_1<\tau n)}\int_0^{\tau n} uu^{-\beta}du\\ &= \frac{2\mu_\beta}{(2-\beta)P(X_1<\tau n)}(\tau n)^{2-\beta}. \end{aligned}$$

Given τ there is n_τ such that $P(X_1<\tau n)>1/2$ if $n>n_\tau$, so that

$$Var(Y_1)\le\frac{4\mu_\beta\tau^{2-\beta}}{(2-\beta)}n^{2-\beta}=:C_{\beta,\tau}n^{2-\beta}$$

Putting everything together, we can apply Prohorov's theorem with $t = \epsilon n$ and the other constants specified above. This yields

$$\begin{aligned} P(X_1^* + \ldots + X_n^* > (\mu+\epsilon)n) &\leq \exp\{-\frac{n\epsilon}{2\tau n} \log\left(\frac{\epsilon\tau n^2}{nC_{\beta,\tau}n^{2-\beta}}\right) \\ &= \left(n\frac{\epsilon}{C_{\tau,\beta}}\right)^{-(\beta-1)\epsilon/(2\tau)}. \end{aligned}$$

Since this holds for any $\tau > 0$, we can take τ so small that $(\beta - 1)\epsilon/(2\tau) > \alpha - 1$, which implies (54). As explained before, this implies (51).

□

A natural question is whether the proof, or at least the statement of the theorem carries over to subexponential tails. One key property that is used in the proof is

$$\lim_{\delta\to\epsilon}\lim_{x\to\infty}\frac{P(X_1 > \epsilon x)}{P(X_1 > \delta x)} = 1. \tag{55}$$

This is true for regularly varying tails, but not for Weibull tails of the form $\exp\{-x^\eta\}$, in which case the above ratio equals:

$$\frac{P(X_1 > \epsilon x)}{P(X_1 > \delta x)} = \exp\{-x^\eta(\epsilon^\eta - \delta^\eta)\}.$$

This either converges to 0 or to ∞ is $\delta \neq \epsilon$. So to extend the theorem to Weibull distributions, it is at least necessary to use different methods.

To see that the statement $P(S_n > (\mu+\epsilon)n) \sim nP(X_1 > \epsilon n)$ may not even be true, consider the following argument. Assume that $E[X_1^2] < \infty$. From the symmetry, we have the following lower bound:

$$P(S_n > (\mu+\epsilon)n) \geq nP(X_1 > \epsilon n - \sqrt{n})P(X_2 + \ldots + X_n > \mu n + \sqrt{n}). \tag{56}$$

Using the central limit theorem, we have that $P(X_2 + \ldots + X_n > \mu n + \sqrt{n}) \to K$ for some $K > 0$. Thus, if

$$\frac{P(X_1 > \epsilon n - \sqrt{n})}{P(X_1 > \epsilon n)} \to \infty, \tag{57}$$

then also

$$\frac{P(S_n > (\mu+\epsilon)n)}{nP(X_1 > \epsilon n)} \to \infty. \tag{58}$$

For the Weibull case, it can be shown that (57) holds if $\eta > 1/2$. On the other hand, if $\eta < 1/2$, then the tail of X is *square-root insensitive*, i.e.

$$P(X_1 > x + \sqrt{x}) \sim P(X_1 > x) \tag{59}$$

and it can be shown that $P(S_n > (\mu + \epsilon)n) \sim nP(X_1 > \epsilon n)$ indeed holds in this case. A recent study of these type of problems is a paper by Denisov, Dieker and Shneer (2007) [13]

3.1. **The all-time maximum of a random walk.** In the example above, we saw that, at least for regularly varying random variables, rare event probabilities for random walks tend to happen by a single big jump of one of the increments of the random walk. As discussed in Chapter 1, the all-time maximum of a random walk plays a crucial role in several applications, particularly in queueing and insurance.

We consider the same setting as above and set $a = -E[X_1]$. Assume that $a > 0$. Define

$$M_n = \sup_{0 \leq k \leq n} S_k, \qquad M = M_\infty,$$

$$\bar{F}^s(x) = \max\{1, \int_x^\infty \bar{F}(t)dt\}.$$

The main question is, what can be said about the behavior of $P(M > x)$ as $x \to \infty$. Many researchers have investigated this problem. Key references are Borovkov (1971) [**?**], Cohen (1973) [10], Pakes (1975) [11], Veraverbeke (1977) [33], Embrechts & Veraverbeke (1982), [16], Korshunov (1997) [**?**], Baccelli & Foss (2002) [**?**] and Zachary (2004) [34]; this list is not exhaustive. The formulation and proof in these notes follow [34]. The proof techniques in the earlier papers are based on Wiener-Hopf factorization techniques.

Let X^s be a random variable such that $P(X^s > x) = \bar{F}^s(x)$.

Theorem 3. *(i) If $X^s \in \mathcal{L}$, then*

$$\liminf_{x\to\infty} \frac{P(M > x)}{\bar{F}^s(x)} \geq \frac{1}{a}. \tag{60}$$

(ii) If $X^s \in \mathcal{S}$, then

$$\lim_{x\to\infty} \frac{P(M > x)}{\bar{F}^s(x)} = \frac{1}{a}. \tag{61}$$

We see that the statement applies much more generally than before. In fact, Korshunov (1997) has shown that, if (61) holds, then X^s is subexponential. We will not treat his proof, but instead focus on proving the above theorem. Before that, we first give some heuristics.

One way to obtain the event $M > x$ is that at some time n, the random walk S_n has the typical value $-an$, and X_{n+1} is so large that $S_{n+1} > x$. For this to happen, we need $X_n > an + x$. This can happen at any time n, and since the random walk needs to be behaving properly until time n, these events (indexed by n) are mutually exclusive.

$$\begin{aligned} P(M > x) &\approx P(\cup_{n=1}^{\infty}\{S_n \approx -an; X_{n+1} > an + x\}) \\ &\approx \sum_{n=0}^{\infty} P(X_{n+1} > an + x) \\ &= \sum_{n=0}^{\infty} \bar{F}(an + x) \\ &\sim \frac{1}{a}\int_x^{\infty} \bar{F}(u)du \end{aligned}$$

(To show the last equivalence will be an exercise).

We now show that these heuristics provide the correct answer.

Proof. Let $\epsilon, \delta > 0$. By the weak law of large numbers, there exists a constant $L = L_{\epsilon,\delta}$ such that $P(S_n > -L - n(a + \epsilon)) \geq 1 - \delta$.

$$\begin{aligned} P(M > x) &= \sum_{n=0}^{\infty} P(M_n \leq x, S_{n+1} > x) \\ &\geq \sum_{n=0}^{\infty} P(M_n \leq x, S_n > -L - n(a + \epsilon), X_{n+1} > x + L + n(a + \epsilon)) \\ &= \sum_{n=0}^{\infty} P(M_n \leq x, S_n > -L - n(a + \epsilon))P(X_{n+1} > x + L + n(a + \epsilon)) \\ &\geq \sum_{n=0}^{\infty} (1 - \delta - P(M_n > x))P(X_{n+1} > x + L + n(a + \epsilon)) \\ &\geq (1 - \delta - P(M > x)) \sum_{n=0}^{\infty} \bar{F}(x + L + n(a + \epsilon)) \\ &\geq (1 - \delta - P(M > x)) \frac{\bar{F}^s(x + L)}{a + \epsilon}, \end{aligned}$$

with the last inequality following from the inequality

$$c\sum_{n=0}^{\infty} \bar{F}(y + nc) \geq \int_x^{\infty} \bar{F}(u)du.$$

Thus,

$$\liminf_{x\to\infty} \frac{P(M > x)}{\bar{F}^s(x)} \geq \frac{1 - \delta}{a + \epsilon} \liminf_{x\to\infty} \frac{\bar{F}^s(x + L)}{\bar{F}^s(x)} = \frac{1 - \delta}{a + \epsilon},$$

where we used the assumption that $X^s \in \mathcal{L}$. Since this holds for any $\delta, \epsilon > 0$ we arrive at part (i).

The proof of part (ii) of the theorem is more involved, and requires the introduction of several definitions. For given constants $\epsilon > 0$ and $R < \infty$, we define the sequence of random times $\tau_n, n \geq 1$ as follows.

$$\begin{aligned}\tau_1 &= \min\{n \geq 1 : S_n > R - n(a-\epsilon)\},\\ \tau_m &= \tau_{m-1} + \min\{n \geq 1 : S_{\tau_{m-1}+n} - S_{\tau_{m-1}} > R - n(a-\epsilon)\}.\end{aligned}$$

Observe that τ_m may be ∞, and observe that $((\tau_m - \tau_{m-1}, (S_{\tau_m} - S_{\tau_{m-1}})), n \geq 1$ is a sequence of i.i.d. random vectors. Define $\gamma = \gamma(\epsilon, R) = P(\tau_1 < \infty)$. Observe that $\gamma \to 0$ as $R \to \infty$ for every $\epsilon > 0$ (this will be another exercise). For convenience, set $S_\infty = -\infty$.

A key step in the proof of the upper bound is to develop an upper bound for $P(S_{\tau_1} > x)$. Write

$$\begin{aligned}P(S_{\tau_1} > x) &= \sum_{n=1}^{\infty} P(\tau_1 = n, S_n > x)\\ &\leq \sum_{n=1}^{\infty} P(S_{n-1} \leq R - (n-1)(a-\epsilon), S_n > x)\\ &\leq \sum_{n=1}^{\infty} P(X_n > x - R + (n-1)(a-\epsilon))\\ &\leq \frac{1}{a-\epsilon}\bar{F}^s(x - R - a + \epsilon).\end{aligned}$$

The last inequality follows from the fact that $\int_y^\infty \bar{F}(u)du \geq c\sum_{n=1}^\infty \bar{F}(y+nc)$.

Define a sequence of i.i.d. random variables $\phi_m, m \geq 1$ as follows:

$$P(\phi_m > x) = P(S_{\tau_m} - S_{\tau_{m-1}} > x \mid \tau_m < \infty)$$

Observe that, since $S_\infty = -\infty$,

$$\begin{aligned}P(\phi_1 > x) &= P(S_{\tau_1} > x)/\gamma\\ &\leq \frac{1}{\gamma(a-\epsilon)}\bar{F}^s(x - R - a + \epsilon)\\ &=: \bar{G}(x).\end{aligned}$$

Since $\bar{G}(x) \sim \frac{1}{\gamma(a-\epsilon)}\bar{F}^s(x)$, $\bar{G}(x)$ is the tail of a subexponential random variable with distribution function $G(x) = 1 - \bar{G}(x)$.

We are now ready to complete the proof. Observe that for x larger than R, the event $M > x$ can only be realized at the random times τ_m.

Thus, if $R > a - \epsilon$,

$$\begin{aligned} P(M > x) &\leq \sum_{m=1}^{\infty} P(M > x; S_{\tau_m} > x - R + a - \epsilon) \\ &\leq \sum_{m=1}^{\infty} P(S_{\tau_m} > x - R + a - \epsilon) \\ &= \sum_{m=1}^{\infty} \gamma^m P(\phi_1 + \ldots + \phi_m > x - R + a - \epsilon). \end{aligned}$$

The last inequality follows by conditioning and noting that $P(\tau_m < \infty) = \gamma^m$. Since $P(\phi_m > x) \leq \bar{G}(x)$, we obtain that

$$P(M > x) \leq \sum_{m=1}^{\infty} \gamma^m \bar{G}^* m(x - R + a - \epsilon). \tag{62}$$

Observe that $\bar{G}^* m(x - R + a - \epsilon) \sim m\bar{G}(x - R + a - \epsilon)$ as $x \to \infty$. Using Proposition 3, we can interchange the sum and the limit $x \to \infty$ to conclude that

$$\sum_{m=1}^{\infty} \gamma^m \bar{G}^* m(x - R + a - \epsilon) \sim \sum_{m=1}^{\infty} m\gamma^m \bar{G}(x - R + a - \epsilon). \tag{63}$$

Since $\sum_{m=1}^{\infty} m\gamma^m = \gamma/(1-\gamma)^2$, and since $\bar{G}(x - R + a - \epsilon) \sim \bar{G}(x)$, we conclude that

$$\begin{aligned} \limsup_{x\to\infty} \frac{P(M > x)}{\bar{F}^s(x)} &\leq \left(\limsup_{x\to\infty} \frac{P(M > x)}{\bar{G}(x)}\right)\left(\limsup_{x\to\infty} \frac{\bar{G}(x)}{\bar{F}^s(x)}\right) \\ &= \frac{\gamma}{(1-\gamma)^2}\frac{1}{\gamma(a-\epsilon)} = \frac{1}{(a-\epsilon)(1-\gamma)^2}. \end{aligned}$$

The desired limsup result now follows by first taking $R \to \infty$ (so that $\gamma \downarrow 0$) and then $\epsilon \downarrow 0$. □

3.2. **Comparison with light tailed random walks.** The rare events we investigated so far concern random walks, and reveal that, in the heavy-tailed case, a rare event is due to a single big realization of one increment. In order words: in heavy-tailed random walks, rare events often occur by a single big jump.

To put this into perspective, it is important to gain some understanding about rare events for random walks with light-tailed increments. In this section, we analyze the behavior of $P(S_n > an)$ as $n \to \infty$ and $P(M > x)$ as $x \to \infty$ under the assumption that $E[e^{\theta X}] < \infty$ for a $\theta > 0$, with $X \stackrel{d}{=} X_1$.

Define the so-called cumulant generating function of a random variable X $\Lambda(\theta) = \log E[e^{\theta X}]$. A useful property is that $\Lambda(\theta)$ is convex on the interior of its effective domain (consisting of all values θ for which $\Lambda(\theta) < \infty$. To see this, take $\theta_1 < \theta_2$ and $\alpha \in (0,1)$. Observe that, by Hölders inequality,

$$\begin{aligned} E[e^{(\alpha\theta_1+(1-\alpha)\theta_2)X}] &= E[e^{\alpha\theta_1 X}e^{(1-\alpha)\theta_2 X}] \\ &\leq E[e^{\theta_1 X}]^{\alpha}E[e^{\theta_2 X}]^{(1-\alpha)}. \end{aligned}$$

Taking logarithms on both sides, we obtain

$$\Lambda(\alpha\theta_1 + (1-\alpha)\theta_2) \leq \alpha\Lambda(\theta_1) + (1-\alpha)\Lambda(\theta_2).$$

This implies convexity of Λ. If X is not deterministic, Hölders inequality is strict, in which case Λ is strictly convex.

We are now ready to investigate the rare event problems mentioned above. We will focus on so-called *logarithmic asymptotics*, i.e., we will analyse the asymptotic behavior of $\log P(S_n > an)$ and $\log P(M > x)$. The presentation is almost identical to the one in [18].

3.2.1. *Cramérs theorem.* Given a random variable X and its cumulant generating function Λ, we define the convex conjugate $\Lambda^*(a)$ of $\Lambda(\theta)$ as

$$\Lambda^*(a) = \sup_{\theta\geq 0}[a\theta - \Lambda(\theta)].$$

In the first lecture we saw that, for $a > \mu = E[X_1]$,

$$P(S_n > an) \leq e^{-n\Lambda^*(a)}. \tag{64}$$

We will now show that this bound is asymptotically sharp.

Theorem 4. *Let $a > \mu$.*

$$\lim_{n\to\infty} \frac{1}{n} \log P(S_n > an) = -\Lambda^*(a). \tag{65}$$

We will prove this theorem under the additional condition that $P(X > a) > 0$, which guarantees that $\Lambda^*(a) < \infty$ (verify this yourself). This implies that X is not deterministic, since $a > \mu$. Thus, $\Lambda(\theta)$ is strictly convex, implying that $\theta a - \Lambda(\theta)$ is strictly concave. Thus, the supremum in $\Lambda^*(a)$ is attainted by a unique $\theta^* > 0$. Strict positivity follows since the derivative of $\theta a - \Lambda(\theta)$ at 0 equals $a - \mu > 0$.

A second assumption we make is that $E[exp^{(\theta^*+\epsilon)X}] < \infty$ for some $\epsilon > 0$.

For a proof without all these regularity assumptions, we refer to Dembo & Zeitouni (1998).

Proof. Since, by (64), $\log P(S_n > an) \leq -\Lambda^*(a)n$, it suffices to show that

$$\liminf_{n\to\infty} \frac{1}{n} \log P(S_n > an) \geq -\Lambda^*(a). \tag{66}$$

For this, we will develop a lower bound of $P(S_n > an)$. Define the subset A_n of $\mathbb{R}^n$ by

$$A_n = \{(x_1, \ldots, x_n) : x_1 + \ldots + x_n > an\}.$$

Observe that

$$P(S_n > an) = \int_{A_n} dF(x_1) \ldots dF(x_n),$$

with $F(x) = P(X \leq x)$. Given θ^*, we define the so-called *tilted* distribution $\tilde{F}$ of F as follows:

$$d\tilde{F}(x) = \frac{e^{\theta^* x}}{E[e^{\theta^* X}]} dF(x).$$

Let $\tilde{X}_i, i \geq 1$ be an i.i.d. sequence with distribution function $\tilde{F}$ and define $\tilde{S}_n = \tilde{X}_1 + \ldots \tilde{X}_n$. We see that

$$\begin{aligned}
P(S_n > an) &= \int_{A_n} dF(x_1) \ldots dF(x_n) \\
&= \int_{A_n} e^{n\Lambda(\theta^*)} e^{-\theta^*(x_1+\ldots+x_n)} d\tilde{F}(x_1) \ldots d\tilde{F}(x_n) \\
&= \int_{\mathbb{R}^n} I(x_1 + \ldots + x_n > an) e^{n\Lambda(\theta^*)} e^{-\theta^*(x_1+\ldots+x_n)} d\tilde{F}(x_1) \ldots d\tilde{F}(x_n) \\
&= E[e^{-\theta^* \tilde{S}_n + n\Lambda(\theta^*)} I(\tilde{S}_n > an)].
\end{aligned}$$

Observe that $E[e^{\theta \tilde{X}}] = E[e^{(\theta^*+\theta)X}]/E[e^{\theta^* X}]$. We see that, due to the assumption we made, $E[e^{\epsilon \tilde{X}}] < \infty$, so that all moments of $\tilde{X}$ are finite. Moreover, due to the definition of θ^*, it can be shown that $E[\tilde{X}] = a$. Thus, the tilted distribution $\tilde{F}$ is designed in such a way that the mean has increased from μ to a, which makes the event of interest more likely. In particular, we see that

$$P(an < \tilde{S}_n < an + \sqrt{n}) \to P(0 < U < 1) > 0, \tag{67}$$

with U a normal random variable with zero mean and the same variance as $\tilde{X}$.

We are now ready to derive a lower bound and the corresponding lim inf result.

$$\begin{aligned} P(S_n > an) &= E[e^{-\theta^* \tilde{S}_n + n\Lambda(\theta^*)} I(\tilde{S}_n > an)] \\ &\geq E[e^{-\theta^* \tilde{S}_n + n\Lambda(\theta^*)} I(an < \tilde{S}_n < an + \sqrt{n})] \\ &\geq e^{-\theta^*(an+\sqrt{n})+n\Lambda(\theta^*)} P(an < \tilde{S}_n < an + \sqrt{n}). \end{aligned}$$

Property (66) easily follows by combining this lower bound with (67). □

From the proof of the lower bound, it is apparent that a typical way in which the event $\{S_n > an\}$ occurs is by *conspirary*: all n random variables are sampled from the distribution $\tilde{F}$ instead of F.

One can show that, if F is exponentially distributed with rate λ, then $\tilde{F}$ is exponentially distributed with rate $1/a$.

3.2.2. *The Cramer-Lundberg estimate.* We now turn our attention to $P(M > x) = P(\sup_n S_n > x)$, with $\mu < 0$.

The following seemingly crude bounds will be instrumental in our approach:

$$\sup_n P(S_n > x) \leq P(\sup_n S_n > x) \leq \sum_{n=0}^{\infty} P(S_n > x). \tag{68}$$

It turns out that these bounds are sharp enough to obtain the logarithmic asymptotics of $P(M > x)$ (for precise asymptotics, different ideas are required).

Assume for convenience that $E[e^{\theta X}] < \infty$ for all $\theta < \infty$. Define θ^* as the strictly positive solution of $E[e^{\theta X}] = 1$.

Theorem 5. *Under the above assumption, as $x \to \infty$,*

$$\lim_{x\to\infty} \frac{1}{x} \log P(M > x) = -\theta^*. \tag{69}$$

Proof. We start with an upper bound. Observe that, for any $\theta \in (0, \theta^*)$, we have $E[e^{\theta X}] < 1$, so that

$$\begin{aligned} P(M > x) &\leq \sum_{n=0}^{\infty} P(S_n > x) \\ &= \sum_{n=0}^{\infty} P(e^{\theta S_n} > e^{\theta x}) \\ &\leq \sum_{n=0}^{\infty} e^{-\theta x} E[e^{\theta S_n}] \end{aligned}$$

$$= e^{-\theta x} \frac{1}{1 - E[e^{\theta X}]} < \infty.$$

This holds for any $\theta < \theta^*$, ($E[e^{\theta X}]$ is strictly increasing in a neighborhood of θ^*) so that

$$\limsup_{x \to \infty} \frac{1}{x} \log P(M > x) \leq -\theta^*.$$

For the lower bound, we can use $P(S_n > x)$ with n chosen cleverly (here our path departs from that in [18]). It turns out that $n = \lceil bx \rceil$ with $b = 1/\Lambda'(\theta^*)$ is useful. Intuitively, this makes sense, since under the exponential tilting with this particular θ^*, we have $E[\tilde{X}_1] = \Lambda'(\theta^*)$: under this new probability distribution, the random walk S_n reaches level x at time $x/\Lambda'(\theta^*)$.

We see that

$$\liminf_{x \to \infty} \frac{1}{x} \log P(M > x) \geq \liminf_{x \to \infty} \frac{1}{x} \log P(S_{\lceil bx \rceil} > x).$$

The latter liminf can be analyzed by transforming it in a problem of the type we have seen before. Define $n = \lceil bx \rceil$ and observe that

$$P(S_{\lceil bx \rceil} > x) \geq P(S_n > n/b).$$

From Cramers theorem, we conclude that

$$\begin{aligned} \liminf_{x \to \infty} \frac{1}{x} \log P(S_{\lceil bx \rceil} > x) &\geq b \liminf n \to \infty \frac{1}{n} \log P(S_n > n/b) \\ &= b \sup_{\theta \geq 0} [\theta/b - \Lambda(\theta)] \\ &= \sup_{\theta \geq 0} [\theta - \Lambda(\theta)/\Lambda'(\theta^*)]. \end{aligned}$$

It can be shown that the optimal value of this optimization problem is θ^*. Since $\Lambda(\theta^*) = 0$, the corresponding value is θ^*. This implies that the asymptotic lower and upper bounds are equal, as required. □

4. Sampling at a subexponential time

4.1. **Background: GI/GI/1 performance measures.** In the previous section we derived the tail asymptotics of the supremum of a random walk.

We can specialize this result to the single server queue with renewal arrivals and i.i.d. service times as follows. Let A be a generic interarrival time and let B be a generic service time. Let the arrival rate $\lambda = 1/E[B]$, and let B^e be a residual service time. Let W be the steady-state waiting time of a customer, and let Q be the number of waiting customers on an arrival epoch. Assume the service discipline is FIFO.

It is well known that $W \stackrel{d}{=} M$, with M the supremum of a random walk with increments distributed as $X = B - A$. Let the traffic load $\rho = \lambda E[B]$. For stability, it is required that $\rho < 1$ and this is always assumed.

Theorem 6. *If B^r is subexponential, then*

$$P(W > x) \sim \frac{\rho}{1-\rho} P(B^r > x)$$

This follows from the fact that the tail of B^r is equivalent (up to a constant) of $\int_x^\infty P(X > u)du$, and from Theorem 3.

Another performance measure is the *sojourn time*, which is the total time a customer spends in the system. This equals $W + B$, with W and B independent.

If B^r is subexponential, it can be shown that the tail of B is lighter than the tail of B^r (verify this yourself). Since W is subexponential, and B has a lighter tail than W ($P(B > x) = o(P(W > x))$, it follows that

$$P(W + B > x) \sim P(W > x).$$

We now turn to the queue length Q. Let $S_n^A = A_1 + \ldots + A_n$ be the sum of the first n interarrival times and let $N(t) = \max\{n : S_n^A \leq t\}, t \geq 0$ be the associated renewal process. A well known property linking W and Q is *distributional Little's law*. Take W such that it is independent of $N(\cdot)$. Then $Q \stackrel{d}{=} N(W)$.

We are interested in this identity since it is helpful to establish to obtain the tail asymptotics of Q. Since $N(t)/t \to \lambda$ it is natural to expect that the tail behavior of Q is similar to the tail behavior of λW. We will develop several conditions to verify this conjecture. This problem, and its motivation, has been considered first in [6].

4.2. **General setup and a necessary condition.** The central focus of attention in this chapter is the following: consider a stochastic process $X(\cdot)$ and an independent random variable T, which is heavy tailed. Assume that $X(t)/t \to \mu$ for some $\mu \in (0, \infty)$. We are interested whether

$$P(X(T) > x) \sim P(\mu T > x), x \to \infty. \tag{70}$$

To see what types of conditions we can expect, consider the case where $X(t) = \mu t + B(t)$, with $B(\cdot)$ a standard Brownian motion. In this case,

$$P(X(T) > x) = P(\mu T + \sqrt{T} U > x),$$

with U a standard-normal random variable. By solving a quadratic equation with $\sqrt{T}$ as unknown, it can be shown that

$$P(X(T) > x) = P(T > \frac{x}{\mu} - 2\sqrt{\frac{x}{\mu}} + O_P(1)),$$

with $O_P(1)$ a constant that is an a.s. bounded function of U. From this expression it can be shown that the condition

$$P(T > x) \sim P(T > x - \sqrt{x}),$$

known as square root insensitivity, is necessary for (70) to hold. The error $\sqrt{x}$ comes from the central limit theorem. If $X(t) = \mu t + Y(t)$, and $-M < Y(t) < M$, then the bounds

$$P(\mu T > x + M) \leq P(X(T) > x) \leq P(\mu T > x - M)$$

show that it is sufficient to have that T is long-tailed for (70) to hold.

4.3. **Sufficient conditions under regular variation.** If the tail of T is regularly varying, it is also square root insensitive, so one may expect that (70) holds for a wide class of processes $X(t), t \geq 0$.

For convenience, we assume that $X(\cdot)$ has increasing sample paths; so we can define its inverse $R(\cdot)$. Note that

$$P(X(T) > x) = P(T > R(x)).$$

We introduce the following three conditions.

$$R(x)/x \to \gamma \text{ in probability.} \tag{71}$$

This is a weak law of large numbers, which is not a stringent condition, and holds if a similar condition holds for $X(\cdot)$ with $\mu = 1/\gamma$. The next condition is that the tail of T is regularly varying:

$$P(T > x) = L(x)x^{-\alpha}, \alpha > 0 \tag{72}$$

(if one wishes, one could extend this to intermediate regular variation). The third condition is a more technical one, ruling out unusual behavior of the process $R(\cdot)$. Assume that there exist an $\epsilon > 0$ such that

$$P(R(x) < \epsilon x) = o(P(T > x)). \tag{73}$$

Theorem 7. *If (71) and (72) hold, then*

$$\liminf_{x\to\infty} \frac{P(X(T) > x)}{P(T > \gamma x)} \geq 1. \tag{74}$$

If in addition (73) holds, then

$$\lim_{x\to\infty} \frac{P(X(T) > x)}{P(T > \gamma x)} = 1. \tag{75}$$

This theorem has first been stated in [19]. The proof is taken from [8].

Proof. We first prove the first part of the statement. Let $\delta > 0$ and note that

$$\begin{aligned} P(T > R(x)) &\geq P(T > R(x); R(x) \leq (\gamma+\delta)x) \\ &\geq P(T > (\gamma+\delta)x)P(R(x) \leq (\gamma+\delta)x). \end{aligned}$$

Applying the first two conditions, we obtain

$$\liminf_{x\to\infty} \frac{P(X(T) > x)}{P(T > \gamma x)} \geq \left(\frac{\gamma}{\gamma+\delta}\right)^{\alpha}. \tag{76}$$

To prove the limsup, take $\delta > 0$ and write

$$\begin{aligned} P(T > R(x)) &= P(T > R(x); T > (\gamma-\delta)x) + P(T > R(x); \epsilon x < T < (\gamma-\delta)x) \\ &\quad + P(T > R(x); T < \epsilon x) =: I + II + III. \end{aligned}$$

Note that

$$\begin{aligned} I &\leq P(T > (\gamma-\delta)x), \\ II &\leq P(T > \epsilon x)P(T < (\gamma-\delta)x), \\ III &\leq P(R(x) < \epsilon x). \end{aligned}$$

From these bounds, we conclude that $II = o(P(T > \gamma x))$ and also (by assumption) $III = o(P(T > \gamma x))$. Thus

$$\limsup_{x\to\infty} \frac{P(T > R(x))}{P(T > \gamma x)} \leq \left(\frac{\gamma}{\gamma-\delta}\right)^{\alpha}.$$

Take $\delta \downarrow 0$ to reach the desired conclusion. □

4.4. **Some applications.** Condition (73) is equivalent to the condition that there exists a constant $M < \infty$ such that

$$P(X(t) > Mx) = o(P(T > x)) \tag{77}$$

We now consider some applications to which the above theorem applies.

(1) *GI/GI/1 FCFS Queue length:* In this case, $R(x) = S_{\lfloor x \rfloor}$. Set $n = \lfloor x \rfloor$ and observe that

$$P(S_n < \epsilon x) = P(-S_n > -\epsilon x). \tag{78}$$

Since $-S_1$ has a light right tail, we can apply Chernoff bound to conclude that this probability is decreasing exponentially fast in X. Hence, for the $GI/GI/1$ queue length, assume that B is regularly varying. In this case W is regularly varying as well, and we conclude

$$P(Q > x) \sim P(W > x/\lambda) \sim \frac{\rho}{1-\rho} P(B^e > x/\lambda). \tag{79}$$

(2) *Random sums.* Consider again the random variable $Z = \sum_{i=1}^{N} Y_i$, with $Y_i, i \geq 1$ an i.i.d. sequence, and suppose now that the tail of N is regularly varying with index $-\alpha$. If (77) holds, we conclude that

$$P(Z > x) \sim P(N > xE[Y_1]).$$

If Y is regularly varying with index $-\beta$, then (77) if $\beta > \alpha + 1$, as can be shown by applying Theorem 1. If the Y are light-tailed, one can easily apply Chernoff's bound to verify condition (77).

(3) *Processor Sharing queues.* Consider a single server queue with Poisson arrivals, i.i.d. service times, operating under the processor sharing (PS) discipline, where each customer in the system is served at the same rate (the reciprocal of the total number of customers in the system). For PS, $R(x)$ can be defined as the total amount of service a customer receives between time 0 and x. If B is the service time of a customer, and V is the sojourn time of a customer, it follows that

$$P(V > x) = P(B > R(x)).$$

It can be shown that $R(x)/x \to 1 - \rho$ a.s. so (71) is satisfied. To verify condition (73) is rather technical, and involves several estimates for truncated heavy-tailed random variables. Details can be found in the paper [19]. For PS with a finite waiting room of size K, one always has $R(x) > x/K$, so (73) is easily satisfied in this case.

4.5. **Results under square root insensitivity.** If the tail of T is not regularly varying, it is still possible to obtain asymptotic results. In this section, we give a taste of these results. For full proofs, we refer to the paper [22]. An earlier paper on this subject was [6].

We list a number of properties (without stating proofs). We always assume $T \geq 0$.

(1) If T is square root insensitive, then $P(T > x+c\sqrt{x}) \sim P(T > x)$ for any constant c.
(2) If T is square root insensitive and U is standard normal, then $P(T > x + U\sqrt{x}) \sim P(T > x)$.
(3) T is square root insensitive if and only if $\sqrt{T}$ is long-tailed (a property due to S. Foss).
(4) If T is square root insensitive, then $e^{\epsilon\sqrt{x}}P(T > x) \to \infty$ for any $\epsilon > 0$.

There are several conditions one could invoke on the process $X(\cdot)$ to obtain sufficient conditions for (70). A condition from [6] involves large deviations principles and additional second order conditions on the tail of T which come from extreme value theory. If one restricts to either *Gaussian* or *Regenerative* processes, then such conditions are not necessary. For example, assume that the increments of $X((\cdot)$ are regenerative with $U_n, n \geq 0$ regeneration times, $U_0 = 0$. This means that $(X(U_n + t) - X(U_n), t \geq 0)$ is independent of the collection of random variables $(X(s), s \leq U_n)$.

It is not necessary to assume that the process $X(\cdot)$ is increasing. Set $M_1 = \sup_{0<t<U_1} X(t)$.

Theorem 8. *If $E[U_1^2] < \infty$, if $E[e^{\sqrt{M_1}}] < \infty$, and if T is square root insensitive, then (70) holds with $\mu = E[X(T_1)]/E[T_1]$, i.e.*

$$P(X(T) > x) \sim P(\mu T > x).$$

This theorem is based on the following bound. Let μ be defined as above.

Theorem 9. *Under the above conditions, there exist strictly positive and finite constant c, C such that*

$$P(\sup_{0<t<x}[X(t) - \mu t] > u) \leq C\left(e^{-c\frac{u^2}{x}} + e^{-cx} + xe^{-c\sqrt{u}}\right).$$

A proof of both results is given in [22].

5. Regenerative processes and fluid models

Consider a server which is able to process work at rate c. Excess work is stored in an infinite-sized buffer. Let $A(s, t)$ be the amount of traffic offered to the buffer in time $[s,t]$. Set $A(t) = A(0,t)$. Assume that the increments of $A(\cdot)$ are stationary, and let $\rho = E[A(0,1)]$. If $\rho < c$, the system is stable, and one can show that the stationary amount of work in the system V_c has the property $V_c = \sup_{t>0}[A(-t,0) - ct]$.

A can be generated by several so-called on-off sources transmitting fluid, but it can also be a Gaussian process. The point is that the A process need not to have jumps, but can have gradual input. This queueing system is often denoted as a fluid queue.

Typically, we consider a process $S(t), t \geq 0$ with stationary increments, and are interested in $\sup_{t>0} S(t)$. This leads to V_c if we take $S(t) = A(-t,0) - ct$.

We will be interested in deriving tail asymptotics for V_c in case that the input process $S(\cdot)$ exhibits subexponential characteristics. We first develop a framework for regenerative processes, leading to the concept of a perturbed random walk. We then investigate a single on-off source, and after that move on to multiple on-off sources.

5.1. **Regenerative processes and perturbed random walks.** Assume that $S(t), t \geq 0$, is right continuous with left limits drifting to $-\infty$ such that $S(0) = 0$. Suppose that there exists a renewal process with renewal epochs $0 \leq T_0 < T_1 < ...$ such that

$$(S(t))_{0\leq t<T_0}, \qquad (S(T_0+t) - S(T_0))_{0\leq t<T_1-T_0}, \qquad \ldots$$

are independent, and the distribution of $(S(T_k+t) - S(T_k))_{0\leq t<T_{k+1}-T_k}$ is identical for all $k \geq 0$. We call $T_n, n \geq 0$, the regeneration or renewal epochs for $S(t), t \geq 0$. If $T_0 = 0$, we say that $(S(t))$ is zero-delayed. Define

$$M = \sup_{t\geq 0} S(t), \qquad M_{n+1} = \sup_{T_n\leq t<T_{n+1}} S(t) - S(T_n), \quad \text{for } n \geq 0,$$

and denote $X_{n+1} = S(T_{n+1}) - S(T_n), S_n = S(T_n), n \geq 0$. This is strongly related to the setting considered in Asmussen *et al.* (1999) [3] and many other papers. Typically, the distribution of M is too complicated to compute exactly. Therefore, one is often concerned with the tail behavior of M, i.e. the behavior of $P(M > x)$ as x grows large.

We focus on the zero-delayed case. Under this assumption, we have the identity

$$M = \sup_{n\geq 1}[S_{n-1} + M_n]. \tag{80}$$

The sequence $M_n, n \geq 1$, is i.i.d. but depends on the random walk $S_n, n \geq 1$, since M_n and X_n are dependent. Note that the sequence of pairs $(M_n, X_n), n \geq 1$, is i.i.d. Thus, the regenerative setting can be viewed a special case of the more general framework of the *perturbed random walk*, which is considered in a recent paper by Araman & Glynn (2006) [2]. The authors investigate the tail behavior of M in a variety of cases.

In this section, we focus on the general perturbed random walk case (so we do not necessarily assume that the pair (M_1, X_1) come from the regenerative process). We are interesting in the case that $M_1^* = \max\{M_1, X_1\}$ is heavy-tailed. Since M is stochastically larger than M_1^*, M will be heavy-tailed as well in this case. If we take $M_1 = 0$, we get back at the random walk case, so the theorem we aim to derive will be an extension of Theorem 3. We assume that the joint distribution of (X_1, M_1) satisfies $E[X_1] \in (-\infty, 0)$, $E[M_1] < \infty$, and $P(M_1 = -\infty) = 0$. The next theorem appeared in [25].

Theorem 10. *Suppose that* $E[M_1^*] < \infty$ *and that* $\min\{1, \int_x^\infty P(M_1^* > u)\mathrm{d}u\}$ *is subexponential. Then*

$$P(M > x) \sim \frac{1}{a}\int_x^\infty P(M_1^* > u)\mathrm{d}u, \tag{81}$$

as $x \to \infty$*, with* $a = -E[X_1]$.

Proof. The proof consists in deriving lower and upper bounds, which asymptotically coincide.

We start with the lower bound. The idea of the lower bound is to identify a way in which the event $M > x$ occurs. Informally speaking, we choose an event on which $S_{n-1} - M_n, n \geq 1$, behaves in a typical way up to some time k for which $M_{k+1}^* = \max\{M_{k+1}, X_{k+1}\}$ is large. By also including the event that M_{k+2} is not too small, we ensure that $M > x$.

Let $0 < \delta < a$ be given and define for $n \geq 1$, the event $E_n = E_n(\delta, K)$ as

$$E_n = \{S_k \in (-k(a+\delta) - K, -k(a-\delta) + K), k \leq n\}.$$

In addition, consider the event $F_n = F_n(\delta, K)$ which is defined by

$$F_n = \{M_k < \delta k + K, k \leq n\}.$$

Also define $G(x) = P(M_n < x) = P(M_1 < x)$ and let K be such that $\bar{G}(K) = 1 - G(K) < 1/2$. Since $\log(1-x) \geq -2x$ if $x \in (0, 1/2)$, we see that

$$\log P(F_n) = \sum_{k=1}^n \log(1 - \bar{G}(\delta k + K)) \geq -2\sum_{k=1}^\infty \bar{G}(\delta k + K) = -2E[\lfloor (M_1 - K)^+/\delta \rfloor],$$

with $y^+ = \max\{y, 0\}$. Since $E[M_1] < \infty$, the last expression converges to 0 if $K \to \infty$ for any $\delta > 0$. Combining this fact with the weak law of large numbers for $S_n, n \geq 1$, we arrive at the following conclusion: for every $\epsilon > 0$ there exists a K such that $P(E_n \cap F_n) \geq 1 - \epsilon$.

For $n \geq 1$, we define the event

$$G_n = F_{n+1} \cap E_{n+1} \cap \{M^*_{n+1} > x + n(\delta + a) + 2K\} \cap \{M_{n+2} > -K\} \quad (82)$$

Observe that the events $G_n, n \geq 1$, are disjoint and that G_n implies $M > x$ for every $n \geq 1$. Consequently,

$$\begin{aligned} P(M > x) &\geq P(\cup_{n=1}^{\infty} G_n) = \sum_{n=1}^{\infty} P(G_n) \\ &\geq (1-\epsilon) \sum_{n=1}^{\infty} P(M^*_{n+1} > x + n(\delta + a) + 2K) P(M_{n+2} > -K) \\ &\sim \frac{1-\epsilon}{\delta + a} P(M_1 > -K) \int_{x+K}^{\infty} P(M^*_1 > u) \mathrm{d}u \\ &\sim \frac{1-\epsilon}{\delta + a} P(M_1 > -K) \int_{x}^{\infty} P(M^*_1 > u) \mathrm{d}u, \end{aligned}$$

where in the last two steps, we have used the fact that the integrated tail of M^*_1 is long-tailed. This implies

$$\liminf_{x \to \infty} \frac{P(M > x)}{\int_x^{\infty} P(M^*_1 > u) \mathrm{d}u} \geq \frac{1-\epsilon}{\delta + a} P(M_1 > -K).$$

The proof of the lower bound follows by letting $K \to \infty$ and $\delta, \epsilon \downarrow 0$.

To obtain an asymptotic upper bound, let $y > 0$ be given and construct the random walk $S^y_n, n \geq 0$: set $S^y_0 = 0$. For $k \geq 1$, set $X^y_k = X_k$ if $M^*_k \leq y$ and $X^y_k = M^*_k$ if $M^*_k > y$. Set finally $S^y_n = X^y_1 + \ldots + X^y_n, n \geq 1$. Informally, the increments of the random walk $S^y_n, n \geq 0$, are the same as those of $S_n, n \geq 0$, except when a large value of the perturbation M^*_n occurs.

Obviously, we have that $S_n \leq S^y_n$ for any $y > 0$ and $n \geq 1$. Moreover, we have the following crucial bound:

$$\sup_{n \geq 1} [S_{n-1} + M_n] \leq \sup_{n \geq 0} S^y_n + y. \quad (83)$$

For $x > y$, we have $P(X^y_k > x) = P(M^*_k > x)$, which implies that the integrated tail of X^y_k is subexponential. Thus, we can apply Theorem 3, yielding

$$P(\sup_{n \geq 1} S^y_n > x) \sim \frac{1}{-E[X^y_1]} \int_x^{\infty} P(M^*_1 > u) \mathrm{d}u. \quad (84)$$

Putting everything together, we conclude that

$$\limsup_{x\to\infty} \frac{P(M>x)}{\int_x^\infty P(M_1^*>u)\mathrm{d}u} \le \limsup_{x\to\infty} \frac{P(\sup_{n\ge 1} S^y_{T_n} > x-y)}{\int_x^\infty P(M_1^*>u)\mathrm{d}u} \le \frac{1}{-E[X_1^y]}.$$

By dominated convergence, we have that $-E[X_1^y] \to a$ as $y\to\infty$. □

5.2. **Application to a fluid model.** To illustrate the general theory developed in the previous section, we now investigate a simple example. Let $J(t), t \ge 0$, be an alternating renewal (0-1) process with generic on-period T_{on} and generic off-period T_{off}, i.e. T_{on} is the period where $J(s) = 1$ and T_{off} is the period where $J(s) = 0$. Let $Y(t) = r\int_0^t J(s)\mathrm{d}s, t \ge 0$, be the associated integrated on-off process. The constant $r > 0$ is called the *on rate*. Assume that $J(t)$ is such that an on-period starts at time 0. Let the sequence $(T_{on,i}, T_{off,i}), i \ge 1$ representing on-times and off-times be i.i.d. with $(T_{on,1}, T_{off,1}) \stackrel{d}{=} (T_{on}, T_{off})$. We assume T_{on} and T_{off} to be independent and assume that $T_{on} + T_{off}$ has finite mean. Assume further that $E[J(t)] \to \rho \in (0, c)$ for some constant $c > 0$ which is called the *drain rate*. Under these conditions, the process $S(t) = Y(t) - ct, t \ge 0$, is converging a.s. to $-\infty$. The renewal epochs for the process $S(t), t \ge 0$, are given by $T_i = \sum_{k=1}^i (T_{on,k} + T_{off,k}), i \ge 0$. In this setting, the distribution of M can be viewed as the distribution of the amount of fluid in a buffer fed by an on-off source, given that an on-period begins. This is a simple and well-known model (see e.g. Heath *et al.* (1998)) [20], and as such it provides a concise illustration of the theory developed in the previous section. In the setting of the previous section, we have $X_1 = (r-c)T_{on} - cT_{off}$ and $M_1 = (r-c)T_{on}$.

Assume that $r > c$. We have that $M_1 = (r-c)T_{on}$, $X_1 = (r-c)T_{on} - cT_{off}$, and $M_1^* = M_1$. Moreover $r, c, E[T_{on}], E[T_{off}]$ are such that $E[X_1]$ is strictly negative. Assume that $\int_x^\infty P(T_{on} > u)\mathrm{d}u$ is subexponential.

In that case, the condition of Theorem 10 is satisfied, and we obtain

$$\begin{aligned}
P(M>x) &\sim \frac{1}{a}\int_{u=x}^\infty P((r-c)T_{on} > u)du \\
&= \frac{r-c}{a}\int_{v=x/(r-c)}^\infty P(T_{on} > v)dv \\
&= \frac{r-c}{a}E[T_{on}]P(T_{on}^r > x/(r-c)) \\
&= p\frac{r-c}{c-\rho}P(T_{on}^r > x/(r-c))
\end{aligned}$$

In the last step, use that $p = E[T_{on}]/(E[T_{on}] + E[T_{off}])$, and $\rho = rp$.

5.3. **Stationary on-off source.** The on-off source in the previous section was constructed in such a way that an on-period was about to begin at time 0. We now consider a $0-1$ alternating renewal process $I(t), t \geq 0$ that is stationary, i.e. $P(I(0)=1)=1$ is p. If the source is on at time 0, its residual on-time is distributed as T_{on}^r, similarly, if the source is off at time t, the residual off time is T_{off}^r. Let T_0 be the time until a first regular on-period occurs.

Note that $T_0 = T_{on}^r + T_{off,0}$ with probability p and note that $T_0 = T_{off}^r$ with probability $1-p$.

Define $S(t) = r\int_0^t I(t) - ct$, and set $V_c = \sup_{t>0} S(t)$. Assume that T_{on}^r is subexponential. We want to obtain the tail behavior of V_c using the result from the previous subsection, from which we know that

$$P(\sup_{t>T_0}[S(t)-S(T_0)] > x) \sim p\frac{r-c}{c-\rho}P(T_{on}^r > x/(r-c)). \tag{85}$$

Since $\sup_{t<T_0} S(t) = 0$ if $I(0)=0$ and$\sup_{t<T_0} S(t) = (r-c)T_{on}^r$ if $I(0)=1$, we find that

$$P(\sup_{t<T_0} S(t) > x) \sim pP(T_{on}^r > x/(r-c)). \tag{86}$$

Theorem 11. *In the setting of this subsection, we have*

$$\begin{aligned} P(V_c > x) &\sim \left(p + p\frac{r-c}{c-\rho}\right) P(T_{on}^r > x/(r-c)) \\ &= (1-p)\frac{\rho}{c-\rho}P(T_{on}^r > x/(r-c)). \end{aligned}$$

Proof. The second line follows from the first one by a straightforward computation, so we focus on the first line. To obtain an asymptotic upper bound, note that

$$\begin{aligned} P(V_c > x) &= P(\sup_{t<T_0} S(t) > x \text{ or } S(T_0) + \sup_{t>T_0}[S(t)-S(T_0)] > x) \\ &\leq P(\sup_{t<T_0} S(t) + \sup_{t>T_0}[S(t)-S(T_0)] > x) \\ &\sim pP((r-c)T_{on} > x) + p\frac{r-c}{c-\rho}P((r-c)T_{on}^r > x) \end{aligned}$$

The third line follows from the second since the random variables are independent and have, up to a constant, the same tail behavior. This settles the proof of the upper bound. For the lower bound, take $y > 0$

and write

$$\begin{aligned} P(V_c > x) &= P(\sup_{t<T_0} S(t) > x \text{ or } S(T_0) + \sup_{t>T_0}[S(t) - S(T_0)] > x) \\ &\geq P(\{\sup_{t<T_0} S(t) > x; S(T_0) \geq -y\} \text{ or } \sup_{t>T_0}[S(t) - S(T_0)] > x + y) \\ &= P(\sup_{t<T_0} S(t) > x; S(T_0) \geq -y) + P(\sup_{t>T_0}[S(t) - S(T_0)] > x + y) \\ &\quad -P(\sup_{t<T_0} S(t) > x; S(T_0) \geq -y)P(\sup_{t>T_0}[S(t) - S(T_0)] > x + y), \end{aligned}$$

using independence where appropriate. Now observe that

$$\begin{aligned} P(\sup_{t<T_0} S(t) > x; S(T_0) \geq -y) &= pP((r-c)T_{on}^r > x; (r-c)T_{on}^r - T_{off} > -y) \\ &\sim pP((r-c)T_{on}^r > x), \end{aligned}$$

since $T_{on}^r \in \mathcal{L}$. Combining the last two displays, and using long-tailedness of $\sup_{t>T_0}[S(t) - S(T_0)]$ completes the proof. □

This result, which is due to [21] as well as Theorem 3, will be building blocks when considering more general models, where the number of input processes may be more than one. This is the subject of the next topic.

6. Fluid models with multiple input processes

In the previous section we examined the tail behavior of $\sup_{t>0} S(t)$ where $S(t)$ was the net input process of an on-off source. Theorem 3 deals with the case where $S(t)$ is a discrete random walk, or equivalently, a compound renewal process. In the present section, we examine what happened when we superpose such input processes.

6.1. **Perturbed risk models.** We first examine the case where the main input processes is a compound renewal process: Define

$X(t) = \sum_{i=1}^{N(t)} B_i$, with $B_i, i \geq 1$ an i.i.d. sequence, and $N(t)$ a (possibly delayed) renewal process). Let A be a generic inter-renewal time. Then $X(t)/t \to \rho = E[B]/E[A]$.

It follows from Theorem 3 (even if the process $N(t)$ is not a pure but a delayed renewal process) that for $c > \rho$,

$$P(\sup_{t>0}[X(t) - ct] > x) \sim \frac{\rho}{c-\rho} P(B^r > x), \tag{87}$$

compare this also with Theorem 6, where we implicitly assumed $c = 1$.

Consider now an independent "noise" process $Y(t), t \geq 0$ which has stationary increments and zero expectation. We call $S(t) = X(t) + Y(t) - ct, t \geq 0$ a perturbed risk process, i.e., a the risk process $X(t) - ct$ is perturbed by $Y(t)$.

Assume that B^r is subexponential and assume that $Y(\cdot)$ is, in a certain sence, "nicer" than $X(\cdot)$ when it comes to its large deviation behavior. In particular, assume that

$$P(\sup_{t>0}[Y(t) - \epsilon t] > x) = o(P(B^r > x)), \tag{88}$$

as $x \to \infty$, for every $\epsilon > 0$. In this case, it is natural to suspect that

$$\sup_{t>0}[X(t) - ct] \approx \sup_{t>0}[X(t) + Y(t) - ct], \tag{89}$$

if one of the quantities is large. This is indeed the case, as is stated by the following theorem, which has been derived by Schlegel (1998), see Schmidli (1999) for a survey on perturbed risk models.

Theorem 12. *If (88) holds, if $Y(t)/t \to 0$ a.s., and if $B^r \in \mathcal{S}$, then*

$$P(\sup_{t>0}[X(t) + Y(t) - ct] > x) \sim P(\sup_{t>0}[X(t) - ct] > x), \tag{90}$$

Proof. Lets begin with the upper bound, which is instructive. Think of the fluid model interpretation. The idea is to make the system less efficient, by dividing the total "capacity" c between the two input processes X and Y. In particular, we take $\epsilon \in (0, c - \rho)$, and allocate $c - \epsilon$ capacity to X, and ϵ to Y. This gives two separate fluid models, and it is intuitively clear that the system as a whole is less efficient, i.e. the total amount of fluid in these two systems together is more than the total amount of fluid in the original system.

Formally, we use the subadditivity property of the sup operator, i.e. for two functions $f(t)$ and $g(t)$ we have the well-known property

$$\sup_{t>0}(f(t) + g(t)) \leq \sup_{t>0} f(t) + \sup_{t>0} g(t).$$

In particular,

$$\begin{aligned} P(\sup_{t>0}[X(t) + Y(t) - ct] > x) &= P(\sup_{t>0}[X(t) - (c-\epsilon)t + Y(t) - \epsilon t] > x) \\ &\leq P(\sup_{t>0}[X(t) - (c-\epsilon)t] + \sup_{t>0}[Y(t) - \epsilon t] > x). \end{aligned}$$

Since $\sup_{t>0}[X(t) - (c-\epsilon)t]$ and $\sup_{t>0}[Y(t) - \epsilon t]$ are independent, since

$$P(\sup_{t>0}[X(t) - (c-\epsilon)t] > x) \sim \frac{\rho}{c - \epsilon - \rho} P(B^r > x), \tag{91}$$

and since (88) holds, we see that

$$P(\sup_{t>0}[X(t) - (c-\epsilon)t] + \sup_{t>0}[Y(t) - \epsilon t] > x) \sim \frac{\rho}{c - \epsilon - \rho} P(B^r > x). \tag{92}$$

We conclude that

$$\limsup_{x\to\infty} \frac{P(\sup_{t>0}[X(t)+Y(t)-ct]>x)}{P(\sup_{t>0}[X(t)-ct]>x)} \leq \limsup_{x\to\infty} \frac{\frac{\rho}{c-\epsilon-\rho}P(B^r>x)}{\frac{\rho}{c-\rho}P(B^r>x)} = \frac{c-\rho}{c-\rho-\epsilon}. \tag{93}$$

The proof of the upper bound is now completed by letting $\epsilon \downarrow 0$.

For the lower bound, we note that the assumption $Y(t)/t \to 0$ a.s. implies that $\inf_{t>0}[Y(t)+\epsilon t]$ is an a.s. bounded random variable. Observe that

$$\begin{aligned}\sup_{t>0}[X(t)+Y(t)-ct] &= \sup_{t>0}[X(t)-(c+\epsilon)t+Y(t)+\epsilon t] \\ &\geq \sup_{t>0}[X(t)-(c+\epsilon)t]+\inf_{t>0}[Y(t)+\epsilon t].\end{aligned}$$

Since

$$P(\sup_{t>0}[X(t)-(c+\epsilon)t]>x) \sim \frac{\rho}{c+\epsilon-\rho}P(B^r>x),$$

it is in particular long-tailed, and therefore

$$P(\sup_{t>0}[X(t)-(c+\epsilon)t]>x-\inf_{t>0}[Y(t)+\epsilon t) \sim \frac{\rho}{c+\epsilon-\rho}P(B^r>x).$$

We conclude that

$$\liminf_{x\to\infty} \frac{P(\sup_{t>0}[X(t)+Y(t)-ct]>x)}{P(\sup_{t>0}[X(t)-ct]>x)} \geq \liminf_{x\to\infty} \frac{\frac{\rho}{c+\epsilon-\rho}P(B^r>x)}{\frac{\rho}{c-\rho}P(B^r>x)} = \frac{c-\rho}{c-\rho+\epsilon}. \tag{94}$$

The proof of the lower bound is now completed by letting $\epsilon \downarrow 0$. □

An example of $Y(t)$ can be Brownian motion, or a (normalized) claim process where the claim sizes have a lighter tail than B. Similarly, the theorem can be applied to the $\sum GI/GI/1$ queue, where the input process is a superposition of compound renewal processes. In this case, $X(\cdot)$ will be the process with the service times having the heaviest tails, and $Y(\cdot)$ will be the normalized process containing the other processes.

Note that the process $X(\cdot)+Y(\cdot)$ may not be regenerative, so it is not obvious to apply the methods of the previous section directly.

6.2. **Perturbing an on-off source.** A natural question to ask is if we can adapt the above proof to the case where $X(t)$ is an on-off process, i.e. $X(t) = r\int_0^t I(u)du$.

The following results provides a partial answer to this question and is a theorem due to Jelenkovic and Lazar (1999).

Theorem 13. *Let $X(\cdot)$ be an on-off process, with $r > c$, let $Y(\cdot)$ be independent of $X(\cdot)$. Assume that T^r_{on} is of (intermediate) regular variation. Assume in addition that $Y(t)/t \to 0$ a.s., and*

$$P(\sup_{t>0}[Y(t) - \epsilon t] > x) = o(P(T^r_{on} > x)), \tag{95}$$

$$P(\sup_{t>0}[X(t) + Y(t) - ct] > x) \sim P(\sup_{t>0}[X(t) - ct] > x), \tag{96}$$

Proof. The proof technique we follow is due to Jelenkovic & Lazar as well, and is similar to the previous proof: write

$$\begin{aligned} P(\sup_{t>0}[X(t) + Y(t) - ct] > x) &= P(\sup_{t>0}[X(t) - (c-\epsilon)t + Y(t) - \epsilon t] > x) \\ &\leq P(\sup_{t>0}[X(t) - (c-\epsilon)t] + \sup_{t>0}[Y(t) - \epsilon t] > x). \end{aligned}$$

Since $\sup_{t>0}[X(t)-(c-\epsilon)t]$ and $\sup_{t>0}[Y(t)-\epsilon t]$ are independent, since

$$P(\sup_{t>0}[X(t) - (c-\epsilon)t] > x) \sim (1-p)\frac{\rho}{c-\epsilon-\rho}P(T^r_{on} > x(r-c-\epsilon)), \tag{97}$$

and since (95) holds, we see that

$$P(\sup_{t>0}[X(t)-(c-\epsilon)t]+\sup_{t>0}[Y(t)-\epsilon t] > x) \sim (1-p)\frac{\rho}{c-\epsilon-\rho}P(T^r_{on} > x/(r-c+\epsilon)). \tag{98}$$

We conclude that

$$\begin{aligned} \limsup_{x\to\infty} \frac{P(\sup_{t>0}[X(t) + Y(t) - ct] > x)}{P(\sup_{t>0}[X(t) - ct] > x)} &\leq \limsup_{x\to\infty} \frac{\frac{\rho}{c-\epsilon-\rho}P(T^r_{on} > x/(r-c+\epsilon))}{\frac{\rho}{c-\rho}P(T^r_{on} > x/(r-c))} \\ &= \frac{c-\rho}{c-\rho-\epsilon}\left(\frac{r-c}{r-c+\epsilon}\right)^{-\alpha} \end{aligned} \tag{99}$$

The proof of the upper bound is now completed by letting $\epsilon \downarrow 0$.

The proof of the lower bound can be adapted in a similar way. □

We see that the proof technique ("ϵ-splitting) does not work if one has on-times that are not of (intermediate) regular variation. A different way to deal with this, has been proposed by Jelenkovic, Momcilovic & Zwart (2004) [22]. Informally, the idea is as follows. Take a large value $M > 0$:

$$\begin{aligned} P(\sup_{t>0}[X(t)+Y(t)-ct] > x) &\approx P(\sup_{t<Mx}[X(t)+Y(t)-ct] > x) \\ &\leq P(\sup_{t<Mx}[X(t)-ct] + \sup_{t<Mx} Y(t) > x) \\ &\leq P(\sup_{t>0}[X(t)-ct] + \sup_{t<Mx} Y(t) > x). \end{aligned}$$

The first inequality needs to be made rigorous, by bounding $P(\sup_{t>Mx}[X(t)+Y(t)-ct] > x)$. This is possible, but non-trivial. The idea is now to consider a setting where $Y(\cdot)$ satisfies the assumptions of Theorem 9. This essentially allows one to replace $\sup_{t<Mx} Y(t)$ with $MO(\sqrt{x})$, and square root insensitivity does the rest. If square root insensitivity does not hold, then the asymptotics are fundamentally different. In the case that $Y(\cdot)$ is a Brownian motion, this has been analyzed by [**?**].

As an application, consider the case where $S(t) = X_1(t)+X_2(t)-ct$, and $X_i(t)$ is an on-off process with peak rate r_i and mean rate ρ_i. Assume that $r_1 + \rho_2 > c$, assume that the on-periods of source 1 are regularly varying with index, and assume that the on-periods of source 2 have a lighter tail than source 1. Then

$$P(\sup_{t>0}[X_1(t)+X_2(t)-ct] > x) \sim P(\sup_{t>0}[X_1(t)-(c-\rho_2)t] > x), \quad (100)$$

and the asymptotic behavior of the right hand side follows from Theorem 11. We see that the second source can essentially be replaced by its mean.

6.3. **Multiple on-off sources.** Assume now that we have multiple statistically identical stationary on-off sources $X_i(\cdot), i = 1, ..., N$. The mean rate of each source is taken to be ρ and the peak rate is taken to be r. Assume $N\rho < c < Nr$. Define

$$V_c^{\{1,\dots,N\}} = \sup_{t>0} \sum_{i=1}^{N} [X_i(t) - ct].$$

Question is: what is the tail behavior of $V_c^{\{1,\dots,N\}}$? We will only discuss intuition and results, we do not give proofs. First we need to identify the most likely way overflow occurs. Intuitively, this occurs if several very long on-periods occur simultaneously, and sources that do not generate these on periods behave in a typical manner.

What is the drift of the process in this case? Assume the number of long on-periods is k, then this drift which we define $d(k)$ is

$$d(k) = kr + (n-k)\rho - c.$$

Define now $k^* = \min\{k : kr + (n-k)\rho - c > 0\}$, and assume that $d(k^*-1) < 0$.

The intuition is now that a subset of k^* sources are responsible for the rare event to occur, and all other sources can be replaced by their mean. Since there are $\binom{N}{k^*}$ choices to choose this subset of k^* sources, it is natural to conjecture that, if $P(T_{on} > x) = L(x)x^{-\alpha}$,

$$P(V_c^{\{1,\dots,N\}} > x) \sim \binom{N}{k^*} P(V_{c-(N-k^*)\rho} > x)$$
$$P(V_{c-(N-k^*)\rho} > x) \sim K(L(x)x^{1-\alpha})^{k^*},$$

for some constant K. These conjectures are proven in Zwart, Borst & Mandjes [36]. The constant K is unfortunately rather complicated.

If we have **heterogeneous** on-off sources, it is also possible to derive the tail behavior. Assume that sources i has mean rate ρ_i, peak rate r_i and on-times that are regularly varying of index $-\alpha_i$.

Let F be a subset of $\{1,...,N\}$. How likely is the event that all sources in F are simultaneously on for a long time, say larger than x? By a simple renewal argument it follows that the probability of this event is regularly varying with index $-\sum_{i\in F}(\nu_i - 1)$. The drift $d(F)$ of the system during such an event is

$$d(F) = \sum_{i\in F} r_i + \sum_{j\notin F} \rho_j.$$

What is now the most likely subset of badly behaving sources? The claim is that this is the subset F that minimizes $\sum_{i\in F}(\nu_i - 1)$ subject to the constraint $d(F) > 0$. In [36] it is indeed shown that the tail of $V_c^{\{1,\dots,N\}}$ is regularly varying with index $-\min_{F:d(F)>0}\sum_{i\in F}(\nu_i - 1)$.

References

[1] D. Anick, D.Mitra, and M.M. Sondhi. Stochastic theory of a data-handling system with multiple sources, Bell System Tech. J. 61 (1982), no. 8, 18711894.
[2] Araman, V. F. and Glynn, P. (2006). Tail asymptotics for the maximum of perturbed random walk. Ann. Appl. Prob. 16, 1411–1431.
[3] Asmussen, S., Schmidli, H. and Schmidt, V. (1999). Tail probabilities for non-standard risk and queueing processes with subexponential jumps. Adv. Appl. Prob. 31, 422–447.
[4] Asmussen, S. Applied Probability and Queues. Springer, 2003.
[5] Asmussen, S. Ruin Probabilities. World Scientific, 2000.
[6] Asmussen, S.; Klüppelberg, C., Sigman, K. Sampling at subexponential times, with queueing applications. Stochastic Process. Appl. 79 (1999), no. 2, 265–286.

[7] Bingham, N. H.; Goldie, C. M.; Teugels, J. L. Regular variation. Encyclopedia of Mathematics and its Applications, 27. Cambridge University Press, Cambridge, 1987.

[8] Borst, Sem; Núñez-Queija, Rudesindo; Zwart, Bert Sojourn time asymptotics in processor-sharing queues. Queueing Syst. 53 (2006), no. 1-2, 31–51.

[9] Borovkov, A. A. Stochastic processes in queueing theory. Translated from the 1972 Russian Edition by Kenneth Wickwire. Applications of Mathematics, No. 4. Springer-Verlag, New York-Berlin, 1976.

[10] Cohen, J. W. Some results on regular variation for distributions in queueing and fluctuation theory. J. Appl. Probability 10 (1973), 343–353.

[11] Pakes, A. G. On the tails of waiting-time distributions. J. Appl. Probability 12 (1975), no. 3, 555–564.

[12] Dembo, Amir; Zeitouni, Ofer Large deviations techniques and applications. Second edition. Applications of Mathematics (New York), 38. Springer-Verlag, New York, 1998. xvi+396 pp.

[13] Denisov, D., Dieker, T., Shneer, V. (2007). Large deviations for random walks under subexponentiality: the big-jump domain. `http://arxiv.org/abs/math.PR/0703265`

[14] Denisov, Denis; Foss, Serguei; Korshunov, Dima Tail asymptotics for the supremum of a random walk when the mean is not finite. Queueing Syst. 46 (2004), no. 1-2, 15–33.

[15] Embrechts, Paul; Klppelberg, Claudia; Mikosch, Thomas Modelling extremal events. For insurance and finance. Applications of Mathematics (New York), 33. Springer-Verlag, Berlin, 1997.

[16] Embrechts, P.; Veraverbeke, N. Estimates for the probability of ruin with special emphasis on the possibility of large claims. Insurance Math. Econom. 1 (1982), no. 1, 55–72.

[17] Feller, William. An introduction to probability theory and its applications. Vol. II. Second edition John Wiley & Sons, Inc., New York-London-Sydney, 1971.

[18] Ganesh, Ayalvadi; O'Connell, Neil; Wischik, Damon Big queues. Lecture Notes in Mathematics, 1838. Springer-Verlag, Berlin, 2004.

[19] Guillemin, Fabrice; Robert, Philippe; Zwart, Bert Tail asymptotics for processor-sharing queues. Adv. in Appl. Probab. 36 (2004), no. 2, 525–543.

[20] Heath, David; Resnick, Sidney; Samorodnitsky, Gennady Heavy tails and long range dependence in ON/OFF processes and associated fluid models. Math. Oper. Res. 23 (1998), no. 1, 145–165.

[21] Jelenković, Predrag R.; Lazar, Aurel A. Asymptotic results for multiplexing subexponential on-off processes. Adv. in Appl. Probab. 31 (1999), no. 2, 394–421.

[22] Jelenković, Predrag; Momčilović, Petar; Zwart, Bert Reduced load equivalence under subexponentiality. Queueing Syst. 46 (2004), no. 1-2, 97–112.

[23] Klüppelberg, Claudia. Subexponential distributions and integrated tails. J. Appl. Probab. 25 (1988), no. 1, 132–141.

[24] Korshunov, D. On distribution tail of the maximum of a random walk. Stochastic Process. Appl. 72 (1997), no. 1, 97–103.

[25] Palmowski, Zbigniew; Zwart, Bert Tail asymptotics of the supremum of a regenerative process. J. Appl. Probab. 44 (2007), no. 2, 349–365.

[26] Pitman, E. J. G. Subexponential distribution functions. J. Austral. Math. Soc. Ser. A 29 (1980), no. 3, 337–347.

[27] Prohorov,Yu.V. [Prohorov, Ju.V.] An extremal problem in probability theory. (Russian. English summary) Teor. Veroyatnost. i Primenen 4 1959 211214

[28] Rolski, Tomasz; Schmidli, Hanspeter; Schmidt, Volker; Teugels, Jozef Stochastic processes for insurance and finance. (English summary) Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 1999.

[29] Samorodnitsky, G. (2002). Long-range dependence, Heavy tails, rare events. http://www.math.ku.dk/~mikosch/maphysto/gena1/gena4.pdf

[30] Samorodnitsky, G. (2007). Long-range dependence. http://people.orie.cornell.edu/~gennady/techreports/LRD-NOW.pdf

[31] Shwartz, Adam; Weiss, Alan Large deviations for performance analysis. Queues, communications, and computing. With an appendix by Robert J. Vanderbei. Stochastic Modeling Series. Chapman & Hall, London, 1995.

[32] Sigman, Karl. Appendix: a primer on heavy-tailed distributions. Queues with heavy-tailed distributions. Queueing Systems Theory Appl. 33 (1999), no. 1-3, 261–275.

[33] Veraverbeke, N. Asymptotic behaviour of Wiener-Hopf factors of a random walk. Stochastic Processes Appl. 5 (1977), no. 1, 27–37.

[34] Zachary, Stan A note on Veraverbeke's theorem. Queueing Syst. 46 (2004), no. 1-2, 9–14.

[35] Zwart, A.P. Queueing Systems with Heavy tails. Ph.D dissertation, Eindhoven University of Technology, 2001. http://alexandria.tue.nl/extra2/200112999.pdf

[36] Zwart, Bert; Borst, Sem; Mandjes, Michel Exact asymptotics for fluid queues fed by multiple heavy-tailed on-off flows. Ann. Appl. Probab. 14 (2004), no. 2, 903–957.

[37] Zwart, Bert; Borst, Sem; Dębicki, Krzystof Subexponential asymptotics of hybrid fluid and ruin models. Ann. Appl. Probab. 15 (2005), no. 1A, 500–517.