Regular Variation, Subexponentiality and Their Applications in Probability Theory

T. Mikosch
University of Groningen

Contents

0

Heavy–Tailed Distributions

There is no unique definition of a *heavy–tailed distribution* and there cannot exist a universal notion of heavy-tailedness. Indeed, this notion only makes sense in the context of the model we consider. What we usually expect when we talk about "heavy–tailed phenomena" is some kind of different qualitative behaviour of the underlying model, i.e. some deviation from the "normal behaviour", which is caused by the extremes of the sample.

For example, the central limit theorem for iid sums with Gaussian limit distribution holds for an enormous variety of distribution: all we need is a finite variance of the summands. A deviation from the central limit theorem can occur only if the variance of the summands is infinite. For a long time, this kind of distribution has been considered as exceptionally strange although infinite variance stable distributions, as the only limit distributions for sums of iid random variables apart from the normal distribution, have been intensively studied for decades. Only in the last few years infinite variance variables have been accepted as realistic models for various phenomena: the magnitude of earthquake aftershocks, the lengths of transmitted files, on and off periods of computers, the claim sizes in catastrophe insurance, and many others.

In these notes, we are mainly interested in maxima and sums of iid random variables and in related models of insurance mathematics. I have chosen the latter field of application because of my personal interests; alternatively one could have taken renewal theory, branching, queuing where one often faces the same problems and, up to a change of notation, sometimes even uses the same models. In the context mentioned, heavy–tailed distributions are roughly those whose tails decay to zero slower than at an expontial rate. The exponential distribution is usually considered as the borderline between heavy and light tails. In the following two tables we contrast "light–" and "heavy–tailed" distributions which are important for applications.

Two classes of heavy–tailed distribution have been most successful: the distributions with regularly varying tails and the subexponential distributions. One of the aims of these notes is to explain why these classes are "natural" in the context of sums and extremes of iid and weakly dependent random variables. I did not have enough time to include a section about the weak convergence of point processes generated from heavy–tailed distribution although point process techniques are extremely valuable for heavy–tailed modelling, in particular, in the presence of dependence in the underlying point sequence. Point process techniques allow one to handle not only the extremes of dependent sequences but also functionals of sum–type of dependent stationary sequences, including the sample mean, the sample autocovariances and sample autocorrelalions. Fortunately, monographs like Leadbetter, Lindgren and Rootzén [66], Resnick [100] or Falk, Hüsler and Reiss [38] treat the convergence of point processes in the context of the extreme value theory for dependent sequences. The exciting theory for sum–type functionals such as the sample autocovariances and sample autocorrelations for depedent non-linear stationary sequences with regularly varying finite-dimensional distributions has been treated quite recently in various papers. To name a few: Davis and Hsing [23], Davis and Resnick [26], Davis and Mikosch [24], Davis et al. [25]; see also the recent survey by Resnick [101].

These notes were written for the Workshop "Heavy tails and queues" held at the EURANDOM Institute in Eindhoven in April 1999. Parts of the text were adapted from the corresponding chapters in Embrechts, Klüppelberg and Mikosch [34] and cannot replace the complexity of the latter monograph. Other parts of the notes were adapted from recent publications with various of my coauthors, in particular, with B. Basrak, M. Braverman, R.A. Davis, A.V. Nagaev, G. Samorodnitsky, A. Stegeman and C. Stărică; see the list of references.

Groningen, April 1999

Name	Tail $\overline{F} = 1 - F$ or density f	Parameters
Exponential	$\overline{F}(x) = e^{-\lambda x}$	$\lambda > 0$
Gamma	$f(x) = \dfrac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\beta x}$	$\alpha,\ \beta > 0$
Weibull	$\overline{F}(x) = e^{-cx^\tau}$	$c > 0,\ \tau \geq 1$
Truncated normal	$f(x) = \sqrt{\frac{2}{\pi}}\, e^{-x^2/2}$	—
Any distribution with bounded support		

Table 0.0.1 *"Light–tailed" distributions.*

Name	Tail $\overline{F}$ or density f	Parameters
Lognormal	$f(x) = \dfrac{1}{\sqrt{2\pi}\,\sigma x} e^{-(\ln x - \mu)^2/(2\sigma^2)}$	$\mu \in \mathbb{R},\ \sigma > 0$
Pareto	$\overline{F}(x) = \left(\dfrac{\kappa}{\kappa + x}\right)^\alpha$	$\alpha, \kappa > 0$
Burr	$\overline{F}(x) = \left(\dfrac{\kappa}{\kappa + x^\tau}\right)^\alpha$	$\alpha, \kappa, \tau > 0$
Benktander–type–I	$\overline{F}(x) = (1 + 2(\beta/\alpha)\ln x)$ $e^{-\beta(\ln x)^2 - (\alpha+1)\ln x}$	$\alpha, \beta > 0$
Benktander–type–II	$\overline{F}(x) = e^{\alpha/\beta} x^{-(1-\beta)} e^{-\alpha\, x^\beta/\beta}$	$\alpha > 0$ $0 < \beta < 1$
Weibull	$\overline{F}(x) = e^{-cx^\tau}$	$c > 0$ $0 < \tau < 1$
Loggamma	$f(x) = \dfrac{\alpha^\beta}{\Gamma(\beta)} (\ln x)^{\beta-1}\, x^{-\alpha-1}$	$\alpha, \beta > 0$
Truncated α–stable	$\overline{F}(x) = P(\lvert X\rvert > x)$ where X is α–stable	$1 < \alpha < 2$

Table 0.0.2 *"Heavy–tailed" distributions.*

1

Regular Variation

1.1 Definition

In various fields of applied mathematics we observe power law behaviour. Power laws usually occur in perturbed form. To describe the deviation from pure power laws, the notion of *regular variation* was introduced.

Definition 1.1.1 (Karamata [60]) *A positive measurable function f is called* regularly varying (at infinity) with index $\alpha \in \mathbb{R}$ *if*

- *It is defined on some neighbourhood $[x_0, \infty)$ of infinity.*
- $$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^\alpha \quad \textit{for all } t > 0. \tag{1.1}$$

If $\alpha = 0$, f is said to be slowly varying (at infinity).

Remark 1.1.2 The definition of regular variation can be relaxed in various ways. For example, it suffices to require that the limit in (1.1) exists, is positive and finite for all $t > 0$. Then the limiting function χ satisfies the relation $\chi(ts) = \chi(t)\chi(s)$ which implies that χ is a power function.

Remark 1.1.3 It is easy to see that every regularly varying function f of index α has representation

$$f(x) = x^\alpha L(x)\,,$$

where L is some slowly varying function.

Remark 1.1.4 Regular variation of a function f can also be defined at any point $x_0 \in \mathbb{R}$ by requiring that $f(x_0 - x^{-1})$ is regularly varying at infinity. In what follows we usually deal with regular variation at infinity.

Remark 1.1.5 An encyclopaedic treatment of regular variation can be found in Bingham, Goldie and Teugels [6]. A useful survey of regular variation is given in Seneta [105]. There exist various other books which contain surveys on regularly varying functions and their properties; see for example Feller [39], Ibragimov and Linnik [58], Resnick [100], Embrechts, Klüppelberg and Mikosch [34].

Example 1.1.6 Typical examples of slowly varying functions are positive constants or functions converging to a positive constant, logarithms and iterated logarithms. For instance for all real α the functions

$$x^\alpha\,, x^\alpha \ln(1+x)\,,\ (x\ln(1+x))^\alpha\,,\ x^\alpha \ln(\ln(\mathrm{e}+x))$$

are regularly varying at ∞ with index α. The following examples are not regularly varying

$$2+\sin x\,,\quad \mathrm{e}^{[\ln(1+x)]}\,,$$

where $[\cdot]$ stands for integer part. In Theorem 1.2.1 below we give a general representation of regularly varying functions. It is perhaps interesting to note that a slowly varying function L may exhibit infinite oscillation in that it can happen that

$$\liminf_{x\to\infty} L(x) = 0 \quad \text{and} \quad \limsup_{x\to\infty} L(x) = \infty\,.$$

An example is given by

$$L(x) = \exp\left\{(\ln(1+x))^{1/2}\cos\left((\ln(1+x))^{1/2}\right)\right\}\,.$$

□

1.2 Basic Properties

In this section we collect some of the most important properties of regularly varying functions.

Theorem 1.2.1 (Representation theorem)
A positive measurable function L on $[x_0,\infty)$ is slowly varying if and only if it can be written in the form

$$L(x) = c(x)\exp\left\{\int_{x_0}^{x}\frac{\varepsilon(y)}{y}\,dy\right\}\,, \tag{1.2}$$

where $c(\cdot)$ is a measurable non–negative function such that $\lim_{x\to\infty} c(x) = c_0 \in (0,\infty)$ and $\varepsilon(x)\to 0$ as $x\to\infty$.

Remark 1.2.2 From the representation theorem it is clear that a regularly varying function f with index α has representation

$$f(x) = x^\alpha c(x)\exp\left\{\int_{x_0}^{x}\frac{\varepsilon(y)}{y}\,dy\right\}\,,$$

where $c(\cdot)$ and $\varepsilon(\cdot)$ are as above.

Remark 1.2.3 From the representation theorem we may conclude that for regularly varying f with index $\alpha\neq 0$, as $x\to\infty$,

$$f(x)\to\begin{cases}\infty & \text{if} \quad \alpha>0\,,\\ 0 & \text{if} \quad \alpha<0\,.\end{cases}$$

Moreover, if L is slowly varying then for every $\epsilon>0$

$$x^{-\epsilon}L(x)\to 0 \quad \text{and} \quad x^{\epsilon}L(x)\to\infty \quad \text{as } x\to\infty.$$

The latter property gives some intuitive meaning to the notion of “slow variation”.

An important result is the fact that convergence in (1.1) is *uniform on each compact subset of* $(0,\infty)$.

Theorem 1.2.4 (Uniform convergence theorem for regularly varying functions) *If f is regularly varying with index α (in the case $\alpha > 0$, assuming f bounded on each interval $(0,x]$, $x > 0$), then for $0 < a \le b < \infty$,*

$$\lim_{x\to\infty} \frac{f(tx)}{f(x)} = t^\alpha\,, \quad \textit{uniformly in } t$$

(*a*) *on each* $[a,b]$ *if* $\alpha = 0$,

(*b*) *on each* $(0,b]$ *if* $\alpha > 0$,

(*c*) *on each* $[a,\infty)$ *if* $\alpha < 0$. □

In what follows, for any positive functions f and g,

$$f(x) \sim g(x) \quad \text{as } x \to x_1$$

means that

$$\lim_{x\to x_1} \frac{f(x)}{g(x)} = 1\,.$$

In applications the following question is of importance. Suppose f is regularly varying with index α. Can one find a smooth regularly varying function f_1 with the same index so that $f(x) \sim f_1(x)$ as $x \to \infty$? In the representation (1.2) we have a certain flexibility in constructing the functions c and ε. By taking the function c for instance constant, we already have a (partial) positive answer to the above question. Much more can however be obtained as can be seen from the following result by Adamovič; see Bingham et al. [6], Proposition 1.3.4.

Proposition 1.2.5 (Smooth versions of slow variation) *Suppose L is slowly varying, then there exists a slowly varying function $L_1 \in C^\infty$ (the space of infinitely differentiable functions) so that $L(x) \sim L_1(x)$ as $x \to \infty$. If L is eventually monotone, so is L_1.*

The following result of Karamata is often applicable. It essentially says that integrals of regularly varying functions are again regularly varying, or more precisely, one can take the slowly varying function out of the integral.

Theorem 1.2.6 (Karamata's theorem) *Let L be slowly varying and locally bounded in $[x_0,\infty)$ for some $x_0 \ge 0$. Then*

(*a*) *for* $\alpha > -1$,

$$\int_{x_0}^{x} t^\alpha\, L(t)\, dt \sim (\alpha+1)^{-1}\, x^{\alpha+1} L(x)\,, \quad x\to\infty\,,$$

(*b*) *for* $\alpha < -1$,

$$\int_{x}^{\infty} t^\alpha\, L(t)\, dt \sim -(\alpha+1)^{-1}\, x^{\alpha+1} L(x)\,, \quad x\to\infty\,.$$

Remark 1.2.7 The result remains true for $\alpha = -1$ in the sense that then

$$\frac{1}{L(x)} \int_{x_0}^{x} \frac{L(t)}{t} dt \to \infty\,, \quad x \to \infty\,,$$

and $\int_{x_0}^x (L(t)/t)dt$ is slowly varying. If $\int_{x_0}^\infty (L(t)/t)dt < \infty$ then

$$\frac{1}{L(x)}\int_x^\infty \frac{L(t)}{t}dt \to \infty\,, \quad x \to \infty\,,$$

and $\int_x^\infty (L(t)/t)dt$ is slowly varying.

Remark 1.2.8 The conclusions of Karamata's theorem can alternatively be formulated as follows. Supppose f is regularly varying with index α and f is locally bounded on $[x_0, \infty)$ for some $x_0 \geq 0$. Then

(*a'*) for $\alpha > -1$,

$$\lim_{x\to\infty} \frac{\int_{x_0}^x f(t)\,dt}{xf(x)} = \frac{1}{\alpha+1}\,,$$

(*b'*) for $\alpha < -1$,

$$\lim_{x\to\infty} \frac{\int_x^\infty f(t)\,dt}{xf(x)} = -\frac{1}{\alpha+1}\,.$$

Whenever $\alpha \neq -1$ and the limit relations in either (a') or (b') hold for some positive function f, locally bounded on some interval $[x_0, \infty)$, $x_0 \geq 0$, then f is regularly varying with index α.

The following result is crucial for the differentiation of regularly varying functions.

Theorem 1.2.9 (Monotone density theorem)
Let $U(x) = \int_0^x u(y)\,dy$ (*or* $\int_x^\infty u(y)\,dy$) *where* u *is ultimately monotone* (*i.e.* u *is monotone on* (z, ∞) *for some* $z > 0$). *If*

$$U(x) \sim cx^\alpha L(x)\,, \quad x \to \infty\,,$$

with $c \geq 0$, $\alpha \in \mathbb{R}$ *and* L *is slowly varying, then*

$$u(x) \sim c\alpha x^{\alpha-1} L(x)\,, \quad x \to \infty\,.$$

For $c = 0$ *the above relations are interpreted as* $U(x) = o(x^\alpha L(x))$ *and* $u(x) = o(x^{\alpha-1}L(x))$.

The applicability of regular variation is further enhanced by *Karamata's Tauberian theorem* for Laplace–Stieltjes transforms.

Theorem 1.2.10 (Karamata's Tauberian theorem)
Let U *be a non–decreasing, right–continuous function defined on* $[0, \infty)$. *If* L *is slowly varying,* $c \geq 0$, $\alpha \geq 0$, *then the following are equivalent:*

(*a*) $U(x) \sim cx^\alpha L(x)/\Gamma(1+\alpha)\,, \quad x \to \infty,$

(*b*) $\widehat{u}(s) = \int_0^\infty \mathrm{e}^{-sx}\,dU(x) \sim cs^{-\alpha} L(1/s)\,, \quad s \downarrow 0.$

When $c = 0$, (a) *is to be interpreted as* $U(x) = o(x^\alpha L(x))$ *as* $x \to \infty$; *similarly for* (b). □

This is a remarkable result in that not only the power coefficient α is preserved after taking Laplace–Stieltjes transforms but even the slowly varying function L. From either (a) or (b) in the case $c > 0$, it follows that

(*c*) $U(x) \sim \widehat{u}(1/x)/\Gamma(1+\alpha)\,, \quad x \to \infty\,.$

A surprising result is that the converse (i.e. (c) implies (a) and (b)) also holds. This so–called *Mercerian theorem* is discussed in Bingham et al. [6], p. 274. Various extensions of the above result exist; see for instance Bingham et al. [6], Theorems 1.7.6 and 8.1.6.

1.3 Regularly Varying Random Variables

The notion of regular variation appears in a natural way in various fields of applied probability, so in queuing theory, extreme value theory, renewal theory, theory of summation of random variables, point process theory. In those areas, random variables (and more generally, random vectors) with regularly varying tails have plenty of applications.

In what follows, we write

$$F_X(x) = P(X \le x), \quad x \in \mathbb{R},$$

for the *distribution function* of any random variable X and

$$\overline{F}_X(x) = 1 - F_X(x), \quad x \in \mathbb{R},$$

for its right tail.

Definition 1.3.1 (Regularly varying random variable/distribution)
A non–negative random variable X and its distribution are said to be regularly varying with index $\alpha \ge 0$ *if the right distribution tail $\overline{F}_X$ is regularly varying with index $-\alpha$.*

For convenience we recall here the basic properties of regularly varying functions in terms of regularly varying distributions. For proofs and further references see Bingham et al. [6].

Proposition 1.3.2 (Regularly varying distributions)
Suppose F is a distribution function with $F(x) < 1$ for all $x \ge 0$.

(*a*) *If the sequences (a_n) and (x_n) satisfy $a_n/a_{n+1} \to 1$, $x_n \to \infty$, and if for some real function g and all λ from a dense subset of $(0,\infty)$,*

$$\lim_{n\to\infty} a_n \overline{F}(\lambda x_n) = g(\lambda) \in (0,\infty),$$

then $g(\lambda) = \lambda^{-\alpha}$ for some $\alpha \ge 0$ and $\overline{F}$ is regularly varying.

(*b*) *Suppose F is absolutely continuous with density f such that for some $\alpha > 0$, $\lim_{x\to\infty} x f(x)/\overline{F}(x) = \alpha$. Then f is regularly varying with index $-(1+\alpha)$ and consequently $\overline{F}$ is regularly varying with index $-\alpha$.*

(*c*) *Suppose the density f of F is regularly varying with index $-(1+\alpha)$ for some $\alpha > 0$. Then $\lim_{x\to\infty} x f(x)/\overline{F}(x) = \alpha$. The latter statement also holds if $\overline{F}$ is regularly varying with index $-\alpha$ for some $\alpha > 0$ and the density f is ultimately monotone.*

(*d*) *Suppose X is a regularly varying non–negative random variable with index $\alpha > 0$. Then*

$$\begin{aligned} EX^\beta &< \infty \quad \text{if} \quad \beta < \alpha, \\ EX^\beta &= \infty \quad \text{if} \quad \beta > \alpha. \end{aligned}$$

(*e*) *Suppose $\overline{F}$ is regularly varying with index $-\alpha$ for some $\alpha > 0$, $\beta \ge \alpha$. Then*

$$\lim_{x\to\infty} \frac{x^\beta \overline{F}(x)}{\int_0^x y^\beta \, dF(y)} = \frac{\beta - \alpha}{\alpha}.$$

The converse also holds in the case that $\beta > \alpha$. If $\beta = \alpha$ one can only conclude that $\overline{F}(x) = o(x^{-\alpha} L(x))$ for some slowly varying L.

(*f*) *The following are equivalent:*

(1) $\int_0^x y^2\, dF(y)$ *is slowly varying,*

(2) $\overline{F}(x) = o\left(x^{-2} \int_0^x y^2\, dF(y)\right), \quad x \to \infty.$

Remark 1.3.3 The statements (b), (c), (e) and (f) above are special cases of the general version of Karamata's theorem and the monotone density theorem. For more general formulations of (e) and (f) see Bingham et al. [6], p. 331. Relations (e) and (f) are important in the analysis of the domain of attraction of stable laws; see for instance Section 1.4.1.

1.3.1 Closure Properties

One of the reasons for the popularity of regularly varying functions in probability theory is the following elementary property of regularly varying random variables.

Lemma 1.3.4 (Convolution closure of regularly varying distributions)
Let X and Y be two independent, regularly varying, non–negative random variables with index $\alpha \geq 0$. Then $X+Y$ is regularly varying with index α and

$$P(X+Y>x) \sim P(X>x) + P(Y>x) \quad \text{as } x \to \infty\,.$$

Proof. Using $\{X+Y>x\} \supset \{X>x\} \cup \{Y>x\}$ one easily checks that

$$P(X+Y>x) \geq [P(X>x)+P(Y>x)]\,(1-o(1))\ .$$

If $0<\delta<1/2$, then from

$$\{X+Y>x\} \subset \{X>(1-\delta)x\} \cup \{Y>(1-\delta)x\} \cup \{X>\delta x, Y>\delta x\}\ ,$$

it follows that

$$\begin{aligned} &P(X+Y>x) \\ &\leq \; P(X>(1-\delta)x) + P(Y>(1-\delta)x) + P(X>\delta x)\,P(Y>\delta x) \\ &= \; [P(X>(1-\delta)x) + P(Y>(1-\delta)x)]\,(1+o(1))\,. \end{aligned}$$

Hence

$$\begin{aligned} 1 &\leq \liminf_{x\to\infty} \frac{P(X+Y>x)}{P(X>x)+P(Y>x)} \\ &\leq \limsup_{x\to\infty} \frac{P(X+Y>x)}{P(X>x)+P(Y>x)} \\ &\leq (1-\delta)^{-\alpha}\,, \end{aligned}$$

which proves the result upon letting $\delta \downarrow 0$.

Remark 1.3.5 If X and Y are non–negative, not necessarily independent random variables such that $P(Y>x) = o(P(X>x))$ and X is regularly varying with index α, then $P(X+Y>x) \sim P(X>x)$ as $x\to\infty$. This follows along the lines of the proof of Lemma 1.3.4. In particular, if X and Y are regularly varying with index α_X and α_Y, respectively, and if $\alpha_X < \alpha_Y$, then $X+Y$ is regularly varying with index α_X.

An immediate consequence is the following result.

Corollary 1.3.6 *Let $X, X_1, \ldots, X_n$ be iid non–negative regularly varying random variables and*

$$S_n = X_1 + \cdots + X_n\,, \quad n \geq 1\,.$$

Then

$$P(X_1 + \cdots + X_n > x) \sim n\, P(X > x) \quad \text{as } x \to \infty\,. \tag{1.3}$$

Remark 1.3.7 Relation (1.3) has an intuitive interpretation. Let

$$M_n = \max_{i=1,\ldots,n} X_i\,, \quad n \geq 1\,.$$

Then it is easily seen that for every $n \geq 1$,

$$P(S_n > x) \sim n\, P(X > x) \sim P(M_n > x)\,.$$

This means that, for large x, the event $\{S_n > x\}$ is essentially due to the event $\{M_n > x\}$. The relation $P(M_n > x) \sim nP(X > x)$ as $x \to \infty$ also implies that M_n is regularly varying with the same index as X. This is another closure property of the class of regularly varying random variables.

Another immediate consequence of Lemma 1.3.4 is the following

Corollary 1.3.8 *Let $X, X_1, \ldots, X_n$ be iid non–negative regularly varying random variables with index α and $\psi_1, \ldots, \psi_n$ be non–negative constants. Then*

$$P\left(\psi_1 X_1 + \cdots + \psi_n X_n > x\right) \sim P(X > x)\ \left(\psi_1^\alpha + \cdots + \psi_n^\alpha\right)\,.$$

Under additional conditions, this result can be extended for infinite series

$$Y = \sum_{j=0}^{\infty} \psi_j X_j\,,$$

where (X_n) is an iid sequence of regularly varying random variables and (ψ_j) is a sequence of non–negative numbers. The ψ_js then have to satisfy a certain summability condition in order to ensure the almost sure convergence of Y. Various conditions on (ψ_j) can be found in the literature; see for example Embrechts et al. [34], Lemma A3.26. The weakest conditions on (ψ_j) can be found in recent work by Mikosch and Samorodnitsky [79].

In applications one often has to deal with products of independent random variables where one of them is regularly varying.

Proposition 1.3.9 *Let ξ and η be independent non–negative random variables.*

(a) *Assume that ξ and η are both regularly varying with index $\alpha > 0$. Then $\xi\eta$ is regularly varying with index $\alpha > 0$.*

(b) *Assume that ξ is regularly varying with index α and $E\eta^{\alpha+\epsilon} < \infty$ for some $\epsilon > 0$. Then $\xi\eta$ is regularly varying with index $\alpha > 0$. Moreover,*

$$P(\xi\eta > x) \sim E\eta^\alpha\ P(\xi > x) \quad \text{as } x \to \infty.$$

Remark 1.3.10 Part 2 was proved by Breiman [13]. Part 1 follows from Cline [18]. He also proved that if X_1 and X_2 are iid satisfying a tail balance condition for some $\alpha > 0$:

$$P(X_1 > x) \sim p\, x^{-\alpha}\, L(x) \quad \text{and} \quad P(X_1 \leq -x) \sim q\, x^{-\alpha}\, L(x)\,, \quad \text{as } x \to \infty,$$

where L is slowly varying, $1 - q = p \in [0, 1]$ and if $E|X_1|^\alpha = \infty$, then X_1X_2 is regularly varying with index α and

$$\lim_{x\to\infty} \frac{P(X_1X_2 > x)}{P(|X_1X_2| > x)} = p^2 + q^2\,.$$

1.4 Applications

Regular variation of the tails of a distribution appears as a natural condition in various theoretical results of probability theory. To name a few: domain of attraction conditions for sums of independent random variables, maximum domains of attraction, stationary solutions to stochastic recurrence equations are characterised via regular variation. In what follows, we consider some of these results.

1.4.1 Stable Distributions and Their Domains of Attraction

Stable Distributions

Consider a sequence of independent random variables $X, X_1, X_2, \ldots$ with common distribution F. Consider the random walk

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

with step size distribution F. It is a natural question to ask:

What are the possible limit laws of the standardised random walk

$$a_n^{-1}(S_n - b_n)\,, \quad n \geq 1\,, \tag{1.4}$$

for properly chosen deterministic $a_n > 0$ and $b_n \in \mathbb{R}$?

The well–known answer is that the only limit laws must be α–stable.

Definition 1.4.1 (Stable random variable/distribution)
A random variable Y and its distribution are said to be stable *if for iid copies Y_1, Y_2 of Y and all choices of non-negative constants c_1, c_2 there exist numbers $a = a(c_1, c_2) > 0$ and $b = b(c_1, c_2) \in \mathbb{R}$ such that the following identity in law holds:*

$$c_1 Y_1 + c_2 Y_2 \stackrel{d}{=} a\, Y + b\,.$$

Remark 1.4.2 If X is stable, for every $n \geq 1$ we can find constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that $S_n \stackrel{d}{=} a_n X + b_n$.

It is convenient to describe the stable distributions by their characteristic functions.

Theorem 1.4.3 (Spectral representation of a stable law)
A stable random variable X has characteristic function

$$\phi_X(t) = E\exp\{iXt\} = \exp\left\{i\gamma t - c|t|^\alpha (1 - i\beta\, \mathrm{sign}\,(t)\, z(t,\alpha))\right\}\,, \quad t \in \mathbb{R}\,,$$

where γ is a real constant, $c > 0$, $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, and

$$z(t,\alpha) = \begin{cases} \tan\left(\dfrac{\pi\alpha}{2}\right) & \text{if} \quad \alpha \neq 1\,, \\ -\dfrac{2}{\pi}\ln|t| & \text{if} \quad \alpha = 1\,. \end{cases}$$

Remark 1.4.4 Since α is the parameter which determines the essential properties of a stable distribution it is common to refer to α–stable distributions and random variables. From the characteristic function we see that 2–stable distributions are the Gaussian ones. Although the α–stable distributions with $\alpha < 2$ have densities, in general they cannot be expressed in terms of elementary functions. One of the few exceptions is the symmetric 1-stable distribution. It is the Cauchy law.

Remark 1.4.5 As for Laplace–Stieltjes transforms there exist Tauberian and Mercerian theorems (see Theorem 1.2.10) for characteristic functions as well. The behaviour of the characteristic function $\phi_X(t)$ in some neighbourhood of the origin can be translated into the tail behaviour of the distribution F of X. In particular, for $\alpha < 2$ power law behaviour of the characteristic function at zero implies power law behaviour of the tails, i.e. for any α–stable random variable X there exist non-negative constants p, q with $p+q>0$ and $c_\alpha > 0$ such that

$$P(X > x) \sim p\, c_\alpha\, x^{-\alpha} \quad \text{and} \quad P(X \le -x) \sim q\, c_\alpha\, x^{-\alpha} \quad \text{as } x \to \infty.$$

In particular, α–stable random variables with $\alpha < 2$ have infinite variance.

Domains of Attraction

It is also natural to ask:

Which conditions on F ensure that the standardised random walk (1.4) converges in distribution to a given α–stable random variable?

Before we answer this question we introduce some further notion:

Definition 1.4.6 (Domain of attraction)
We say that the random variable X and its distribution F belong to the domain of attraction of the α–stable distribution G_α *if there exist constants $a_n > 0$, $b_n \in \mathbb{R}$ such that*

$$a_n^{-1}\,(S_n - b_n) \xrightarrow{d} G_\alpha\,, \quad n \to \infty\,,$$

holds.

If we are interested only in the fact that X (or F) is attracted by some α–stable law whose concrete form is not of interest we will simply write $X \in \mathrm{DA}(\alpha)$ (or $F \in \mathrm{DA}(\alpha)$).

The following result characterises the domain of attraction of a stable law completely in terms of regular variation.

Theorem 1.4.7 (Characterisation of domain of attraction)

(*a*) *The distribution F belongs to the domain of attraction of a normal law if and only if*

$$\int_{|y|\le x} y^2\, dF(y)$$

is slowly varying.

(*b*) *The distribution F belongs to the domain of attraction of an α–stable law for some $\alpha < 2$ if and only if*

$$F(-x) = \frac{q+o(1)}{x^\alpha}\, L(x)\,, \quad \overline{F}(x) = \frac{p+o(1)}{x^\alpha}\, L(x)\,, \quad x \to \infty\,,$$

where L is slowly varying and p, q are non–negative constants such that $p+q>0$.

First we study the case $\alpha = 2$ more in detail. If $EX^2 < \infty$ then

$$\int_{|y|\le x} y^2\, dF(y) \to EX^2\,, \quad x \to \infty\,,$$

hence $X \in \mathrm{DA}(2)$. Moreover, by Proposition 1.3.2(f) we conclude that slow variation of $\int_{|y|\le x} y^2\,dF(y)$ is equivalent to the tail condition

$$G(x) = P(|X| > x) = o\left(x^{-2}\int_{|y|\le x} y^2\,dF(y)\right), \quad x\to\infty\,. \tag{1.5}$$

Thus we derived

Corollary 1.4.8 (Domain of attraction of a normal distribution)
A random variable X is in the domain of attraction of a normal law if and only if one of the following conditions holds:

(*a*) $EX^2 < \infty$,

(*b*) $EX^2 = \infty$ *and* (1.5).

The situation is completely different for $\alpha < 2$: $X \in \mathrm{DA}(\alpha)$ implies that

$$G(x) = x^{-\alpha}L(x)\,, \quad x > 0\,, \tag{1.6}$$

for a slowly varying function L and

$$x^2G(x)\Big/\int_{|y|\le x} y^2\,dF(y) \to \frac{2-\alpha}{\alpha}, \quad x\to\infty\,. \tag{1.7}$$

The latter follows from Proposition 1.3.2(e). Hence the second moment of X is infinite. The regular variation of $P(|X| > x)$ is closely related to regular variation of the tails of the limiting α–stable distribution; see Remark 1.4.5.

Relation (1.6) and Corollary 1.4.8 show that the domain of attraction of the normal distribution is much more general than the domain of attraction of an α–stable law with exponent $\alpha < 2$. We see that DA(2) contains at least all distributions that have a second finite moment.

From Corollary 1.4.8 and from (1.6) we conclude the following about the moments of distributions in $\mathrm{DA}(\alpha)$:

Corollary 1.4.9 (Moments of distributions in $\mathrm{DA}(\alpha)$)
If $X \in \mathrm{DA}(\alpha)$ then

$$\begin{aligned} E|X|^\delta &< \infty \quad \textit{for } \delta < \alpha\,, \\ E|X|^\delta &= \infty \quad \textit{for } \delta > \alpha \textit{ and } \alpha < 2\,. \end{aligned}$$

In particular,

$$\begin{aligned} \mathrm{var}(X) &= \infty \quad \textit{for } \alpha < 2\,, \\ E|X| &< \infty \quad \textit{for } \alpha > 1\,, \\ E|X| &= \infty \quad \textit{for } \alpha < 1\,. \end{aligned}$$

Note that $E|X|^\alpha = \int_0^\infty P(|X|^\alpha > x)dx < \infty$ is possible for certain $X \in \mathrm{DA}(\alpha)$, but $E|X|^\alpha = \infty$ for an α–stable X for $\alpha < 2$.

Notes and Comments

The theory above is classical and can be found in detail in Araujo and Giné [1], Bingham et al. [6], Feller [39], Gnedenko and Kolmogorov [44], Ibragimov and Linnik [58], Loève [71] and many other textbooks.

There exists some more specialised literature on stable distributions and stable processes. Mijnheer [77] is one of the first monographs on the topic. Zolotarev

[113] covers a wide range of interesting properties of stable distributions, including asymptotic expansions of the stable densities and many useful representations and transformation formulae. Some limit theory for distributions in the domain of attraction of a stable law is given in Christoph and Wolf [17]. An encyclopaedic treatment of stable laws, multivariate stable distributions and stable processes can be found in Samorodnitsky and Taqqu [104]; see also Kwapień and Woyczyński [65] and Janicki and Weron [59]. The latter book also deals with numerical aspects, in particular the simulation of stable random variables and processes.

Recently there have been some efforts to obtain efficient methods for the numerical calculation of stable densities. This has been a problem for many years and was one of the reasons that practitioners expressed doubts about the applicability of stable distributions for modelling purposes. McCulloch and Panton [74] and Nolan [90, 91] provided tables and software for calculating stable densities for a large variety of parameters α and β. Their methods allow one to determine those densities for small and moderate arguments with high accuracy; the determination of the densities in their tails needs further investigation.

1.4.2 Extreme Value Distributions and Their Domains of Attraction

Extreme Value Distributions

As in Section 1.4.1, consider a sequence of independent random variables X,X_1,X_2, $X_3, \ldots$ with common distribution F. Consider the sequence of partial maxima

$$M_1 = X_1\,, \quad M_n = \max_{i=1,\ldots,n} X_i\,, \quad n \geq 2\,.$$

As for sums of iid random variables, it is a natural question to ask:

What are the possible limit laws of the standardised maxima

$$c_n^{-1}(M_n - d_n)\,, \quad n \geq 1\,, \tag{1.8}$$

for properly chosen deterministic $c_n > 0$ and $d_n \in \mathbb{R}$?

The well–known answer is that the only limit laws must be max–stable.

Definition 1.4.10 (Max–stable distribution)
A non–degenerate random variable X and its distribution are called max–stable *if they satisfy the relation*

$$M_n \stackrel{d}{=} c_n\, X + d_n\,, \quad n \geq 2\,, \tag{1.9}$$

for iid $X, X_1, X_2, \ldots$, appropriate constants $c_n > 0$, $d_n \in \mathbb{R}$.

Assume for the moment that (X_n) is a sequence of iid max–stable random variables. Then (1.9) may be rewritten as follows

$$c_n^{-1}\ (M_n - d_n) \stackrel{d}{=} X\,. \tag{1.10}$$

We conclude that every max–stable distribution is a limit distribution for maxima of iid random variables. Moreover, max–stable distributions are the only limit laws for normalised maxima.

Theorem 1.4.11 (Limit property of max–stable laws)
The class of max–stable distributions coincides with the class of all possible (non–degenerate) limit laws for (properly normalised) maxima of iid random variables.

For a proof via the convergence to types theorem see for example Resnick [100]; cf. Embrechts et al. [34], Theorem 3.2.2.

The following result is *the* basis of classical extreme value theory.

Theorem 1.4.12 (Fisher–Tippett theorem, limit laws for maxima)
Let (X_n) be a sequence of iid random variables. If there exist constants $c_n > 0$, $d_n \in \mathbb{R}$ and some non–degenerate distribution H such that

$$c_n^{-1}(M_n - d_n) \xrightarrow{d} H\,, \tag{1.11}$$

then H belongs to the type of one of the following three distribution functions:

$$\textit{Fréchet:} \qquad \Phi_\alpha(x) = \begin{cases} 0\,, & x \le 0 \\ \exp\{-x^{-\alpha}\}\,, & x > 0 \end{cases} \qquad \alpha > 0.$$

$$\textit{Weibull:} \qquad \Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^{\alpha}\}\,, & x \le 0 \\ 1\,, & x > 0 \end{cases} \qquad \alpha > 0.$$

$$\textit{Gumbel:} \qquad \Lambda(x) = \exp\{-\mathrm{e}^{-x}\}\,, \quad x \in \mathbb{R}.$$

Sketch of the proof. Though a full proof is rather technical, we would like to show how the three limit–types appear; the main ingredient is the convergence to types theorem. Indeed, (1.11) implies that for all $t > 0$,

$$F^{[nt]}\left(c_{[nt]}x + d_{[nt]}\right) \to H(x)\,, \quad x \in \mathbb{R}\,,$$

where $[\cdot]$ denotes the integer part. However,

$$F^{[nt]}(c_n x + d_n) = \left(F^n(c_n x + d_n)\right)^{[nt]/n} \to H^t(x)\,,$$

so that by the convergence to types theorem there exist functions $\gamma(t) > 0$, $\delta(t) \in \mathbb{R}$ satisfying

$$\lim_{n\to\infty} \frac{c_n}{c_{[nt]}} = \gamma(t)\,, \quad \lim_{n\to\infty} \frac{d_n - d_{[nt]}}{c_{[nt]}} = \delta(t)\,, \quad t > 0\,,$$

and

$$H^t(x) = H(\gamma(t)x + \delta(t))\,. \tag{1.12}$$

It is not difficult to deduce from (1.12) that for $s, t > 0$

$$\gamma(st) = \gamma(s)\,\gamma(t)\,, \quad \delta(st) = \gamma(t)\,\delta(s) + \delta(t)\,. \tag{1.13}$$

The solution of the functional equations (1.12) and (1.13) leads to the three types Λ, Φ_α, Ψ_α. Details of the proof are for instance to be found in Resnick [100], Proposition 0.3. □

Remark 1.4.13 The limit law in (1.11) is unique only up to affine transformations. If the limit appears as $H(cx + d)$, i.e.

$$\lim_{n\to\infty} P\left(c_n^{-1}(M_n - d_n) \le x\right) = H(cx + d)\,,$$

then $H(x)$ is also a limit under a simple change of the constants c_n and d_n:

$$\lim_{n\to\infty} P\left(\tilde{c}_n^{-1}\left(M_n - \tilde{d}_n\right) \le x\right) = H(x)$$

with $\tilde{c}_n = c_n/c$ and $\tilde{d}_n = d_n - dc_n/c$. The convergence to types theorem shows precisely how affine transformations, weak convergence and types are related.

Remark 1.4.14 Though, for modelling purposes, the types of Λ, Φ_α and Ψ_α are very different, from a mathematical point of view they are closely linked. Indeed, one immediately verifies the following properties. Suppose $X > 0$, then the following are equivalent:

- X has distribution Φ_α.
- $\ln X^\alpha$ has distribution Λ.
- $-X^{-1}$ has distribution Ψ_α.

Definition 1.4.15 (Extreme value distribution and extremal random variable)
The distributions Φ_α, Ψ_α and Λ as presented in Theorem 1.4.12 *are called* standard extreme value distributions*, random variables with these distributions are* standard extremal random variables*. Distributions of the types of Φ_α, Ψ_α and Λ are* extreme value distributions*; the random variables with these distributions are* extremal random variables*.*

By Theorem 1.4.11, the extreme value distributions are precisely the max–stable distributions. Hence if X is an extremal random variable it satisfies (1.10). In particular, the three cases in Theorem 1.4.12 correspond to

Fréchet:	$M_n \stackrel{d}{=} n^{1/\alpha} X$
Weibull:	$M_n \stackrel{d}{=} n^{-1/\alpha} X$
Gumbel:	$M_n \stackrel{d}{=} X + \ln n\,.$

Remark 1.4.16 It is not difficult to see that the Fréchet distribution Φ_α is regularly varying with index α. Moreover, the Weibull distribution Ψ_α has regularly varying right tail at zero with index $-\alpha$.

Maximum Domains of Attraction

We learnt from Remark 1.4.16 that two kinds of extreme value distributions have regularly varying right tails: the Fréchet distribution Φ_α and the Weibull distribution Ψ_α. This suggests that domains of attraction of these distribution have a close connection with regular variation as well.

As before, we ask the question:

Which conditions on F ensure that the standardised partial maxima (1.8) *converge in distribution to a given extremal random variable?*

As for partial sums and stable distributions, we can introduce domains of attraction.

Definition 1.4.17 (Maximum domain of attraction)
We say that the random variable X and its distribution F belong to the maximum domain of attraction of the extreme value distribution H *if there exist constants $c_n > 0$, $d_n \in \mathbb{R}$ such that*

$$c_n^{-1}\,(M_n - d_n) \xrightarrow{d} H \quad \textit{as } n \to \infty$$

holds. We write $X \in \mathrm{MDA}(H)$ ($F \in \mathrm{MDA}(H)$).

Remark 1.4.18 Notice that the extreme value distribution functions are continuous on $\mathbb{R}$, hence $c_n^{-1}(M_n - d_n) \xrightarrow{d} H$ is equivalent to

$$\lim_{n\to\infty} P\left(M_n \le c_n x + d_n\right) = \lim_{n\to\infty} F^n\left(c_n x + d_n\right) = H(x)\,, \quad x \in \mathbb{R}\,.$$

Having the last remark in mind, the following result is not difficult to derive. It will be quite useful in what follows.

Proposition 1.4.19 (Characterisation of MDA(H))
The distribution F belongs to the maximum domain of attraction of the extreme value distribution H with constants $c_n > 0$, $d_n \in \mathbb{R}$ if and only if

$$\lim_{n\to\infty} n\overline{F}\left(c_n x + d_n\right) = -\ln H(x)\,, \quad x \in \mathbb{R}\,.$$

When $H(x) = 0$ the limit is interpreted as ∞.

For every standard extreme value distribution one can characterise its maximum domain of attraction. Using the concept of regular variation this is not too difficult for the Fréchet distribution Φ_α and the Weibull distribution Ψ_α. The maximum domain of attraction of the Gumbel distribution Λ is not so easily characterised; it consists of distribution functions whose right tail decreases to zero faster than any power function.

MDA of the Fréchet Distribution Φ_α

Recall the definition of the Fréchet distribution Φ_α. By Taylor expansion,

$$1 - \Phi_\alpha(x) = 1 - \exp\left\{-x^{-\alpha}\right\} \sim x^{-\alpha}\,, \quad x \to \infty\,,$$

hence the tail of Φ_α decreases like a power law. We ask:

How far away can we move from a power tail and still remain in MDA(Φ_α)*?*

We show that the maximum domain of attraction of Φ_α consists of distribution functions F whose right tail is regularly varying with index $-\alpha$.

Theorem 1.4.20 (Maximum domain of attraction of Φ_α)
The distribution F belongs to the maximum domain of attraction of Φ_α, $\alpha > 0$, if and only if $\overline{F}(x) = x^{-\alpha}L(x)$ for some slowly varying function L.
If $F \in$ MDA(Φ_α), then

$$c_n^{-1}\, M_n \xrightarrow{d} \Phi_\alpha\,, \tag{1.14}$$

where the constants c_n can be chosen as the $(1-n^{-1})$–quantile of F:

$$\begin{aligned} c_n = F^{\leftarrow}(1-n^{-1}) &= \inf\left\{x\in\mathbb{R} : F(x) \geq 1 - n^{-1}\right\} \\ &= \inf\left\{x\in\mathbb{R} : \left(1/\overline{F}\right)(x) \geq n\right\} \\ &= \left(1/\overline{F}\right)^{\leftarrow}(n)\,. \end{aligned}$$

Remark 1.4.21 Notice that this result implies in particular that for every F in MDA(Φ_α), the right distribution endpoint $x_F = \infty$. Furthermore, the constants c_n form a regularly varying sequence, more precisely, $c_n = n^{1/\alpha}L_1(n)$ for some slowly varying function L_1.

Proof. Let $\overline{F}$ be regularly varying with index $-\alpha$ for some $\alpha > 0$. By the choice of c_n and regular variation,

$$\overline{F}\left(c_n\right) \sim n^{-1}\,, \qquad n \to \infty\,, \tag{1.15}$$

and hence $\overline{F}(c_n) \to 0$ giving $c_n \to \infty$. For $x > 0$,

$$n\overline{F}\left(c_n x\right) \sim \frac{\overline{F}\left(c_n x\right)}{\overline{F}\left(c_n\right)} \to x^{-\alpha}\,, \quad n \to \infty\,.$$

For $x < 0$, immediately $F^n(c_n x) \le F^n(0) \to 0$, since regular variation requires that $F(0) < 1$. By Proposition 1.4.19, $F \in \mathrm{MDA}(\Phi_\alpha)$.

Conversely, assume that $\lim_{n\to\infty} F^n(c_n x + d_n) = \Phi_\alpha(x)$ for all $x > 0$ and appropriate $c_n > 0$, $d_n \in \mathbb{R}$. This leads to

$$\lim_{n\to\infty} F^n(c_{[ns]}x + d_{[ns]}) = \Phi_\alpha^{1/s}(x) = \Phi_\alpha(s^{1/\alpha}x), \quad s > 0, x > 0.$$

By the convergence to types theorem (see Resnick [100]),

$$c_{[ns]}/c_n \to s^{1/\alpha} \quad \text{and} \quad \left(d_{[ns]} - d_n\right)/c_n \to 0.$$

Hence (c_n) is a regularly varying sequence in the sense mentioned above, in particular $c_n \to \infty$. Assume first that $d_n = 0$, then $n\overline{F}(c_n x) \to x^{-\alpha}$ so that $\overline{F}$ is regularly varying with index $-\alpha$ because of Proposition 1.3.2(a). The case $d_n \ne 0$ is more involved, indeed one has to show that $d_n/c_n \to 0$. If the latter holds one can repeat the above argument by replacing d_n by 0. For details on this, see Bingham et al. [6], Theorem 8.13.2, or de Haan [52], Theorem 2.3.1. Resnick [100], Proposition 1.11, contains an alternative argument. □

The domain $\mathrm{MDA}(\Phi_\alpha)$ contains "very heavy–tailed distributions" in the sense that $E(X^+)^\delta = \infty$ for $\delta > \alpha$. Notice that $X \in \mathrm{DA}(G_\alpha)$ for some α–stable distribution G_α with $\alpha < 2$ and $P(X > x) \sim c\, P(|X| > x)$, $c > 0$, as $x \to \infty$ (i.e. this distribution is not totally skewed to the left) imply that $X \in \mathrm{MDA}(\Phi_\alpha)$. If $F \in \mathrm{MDA}(\Phi_\alpha)$ for some $\alpha > 2$ and $EX^2 < \infty$, then F is in the domain of attraction of the normal distribution, i.e. (X_n) satisfies the CLT.

We conclude with some examples.

Example 1.4.22 (Pareto–like distributions)

- Pareto
- Cauchy
- Burr
- Stable with exponent $\alpha < 2$.

All these distributions are Pareto–like in the sense that their right tails are of the form

$$\overline{F}(x) \sim Kx^{-\alpha}, \quad x \to \infty,$$

for some K, $\alpha > 0$. Obviously $\overline{F}$ is regularly varying with index α which implies that $F \in \mathrm{MDA}(\Phi_\alpha)$. Then

$$(Kn)^{-1/\alpha} M_n \xrightarrow{d} \Phi_\alpha.$$

Example 1.4.23 (Loggamma distribution)

The loggamma distribution has tail

$$\overline{F}(x) \sim \frac{\alpha^{\beta-1}}{\Gamma(\beta)}(\ln x)^{\beta-1}x^{-\alpha}, \quad x \to \infty, \quad \alpha, \beta > 0. \tag{1.16}$$

Hence $\overline{F}$ is regularly varying with index $-\alpha$ which is equivalent to $F \in \mathrm{MDA}(\Phi_\alpha)$. The constants c_n can be chosen as

$$c_n \sim \left((\Gamma(\beta))^{-1}(\ln n)^{\beta-1}n\right)^{1/\alpha}.$$

Hence

$$\left((\Gamma(\beta))^{-1}(\ln n)^{\beta-1}n\right)^{-1/\alpha} M_n \quad \xrightarrow{d} \quad \Phi_\alpha.$$

MDA of the Weibull Distribution Ψ_α

An important, though not at all obvious fact is that all distribution functions F in $\mathrm{MDA}(\Psi_\alpha)$ have finite right endpoint x_F. As was already indicated in Remark 1.4.14, Ψ_α and Φ_α are closely related, indeed

$$\Psi_\alpha\left(-x^{-1}\right) = \Phi_\alpha(x)\,, \quad x > 0\,.$$

Therefore we may expect that also $\mathrm{MDA}(\Psi_\alpha)$ and $\mathrm{MDA}(\Phi_\alpha)$ will be closely related. The following theorem confirms this.

Theorem 1.4.24 (Maximum domain of attraction of Ψ_α)
The distribution F belongs to the maximum domain of attraction of Ψ_α, $\alpha > 0$, if and only if $x_F < \infty$ and $\overline{F}(x_F - x^{-1}) = x^{-\alpha}L(x)$ for some slowly varying function L.
If $F \in \mathrm{MDA}(\Psi_\alpha)$, then

$$c_n^{-1}\left(M_n - x_F\right) \xrightarrow{d} \Psi_\alpha\,,$$

where the c_ns can be chosen as $c_n = x_F - F^{\leftarrow}(1-n^{-1})$ and $d_n = x_F$.

We conclude this section with some examples of prominent $\mathrm{MDA}(\Psi_\alpha)$–members.

Example 1.4.25 (Uniform distribution on $(0,1)$)
Obviously, $x_F = 1$ and $\overline{F}(1-x^{-1}) = x^{-1}$. Then by Theorem 1.4.24 we obtain $F \in \mathrm{MDA}(\Psi_1)$. Since $\overline{F}(1-n^{-1}) = n^{-1}$, we choose $c_n = n^{-1}$. This implies in particular

$$n\left(M_n - 1\right) \xrightarrow{d} \Psi_1\,.$$

Example 1.4.26 (Power law behaviour at the finite right endpoint)
Let F be a distribution function with finite right endpoint x_F and distribution tail

$$\overline{F}(x) = K\left(x_F - x\right)^\alpha\,, \quad x_F - K^{-1/\alpha} \le x \le x_F\,, \quad K, \alpha > 0\,.$$

By Theorem 1.4.24 this ensures that $F \in \mathrm{MDA}(\Psi_\alpha)$. The constants c_n can be chosen such that $\overline{F}(x_F - c_n) = n^{-1}$, i.e. $c_n = (nK)^{-1/\alpha}$ and, in particular,

$$(nK)^{1/\alpha}\left(M_n - x_F\right) \xrightarrow{d} \Psi_\alpha\,.$$

Example 1.4.27 (Beta distribution)
The beta distribution is absolutely continuous with density

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)}\,x^{a-1}(1-x)^{b-1}\,, \quad 0 < x < 1\,, \quad a\,, b > 0\,.$$

Notice that $f(1-x^{-1})$ is regularly varying with index $-(b-1)$ and hence, by Karamata's theorem (Theorem 1.2.6),

$$\overline{F}(1-x^{-1}) = \int_{1-x^{-1}}^{1} f(y)\,dy = \int_x^\infty f(1-y^{-1})y^{-2}\,dy \sim x^{-1}f(1-x^{-1})\,.$$

Hence $\overline{F}(1-x^{-1})$ is regularly varying with index $-b$ and

$$\overline{F}(x) \sim \frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b+1)}\,(1-x)^b\,, \quad x \uparrow 1\,.$$

Thus the beta distribution is tail–equivalent to a distribution with power law behaviour at $x_F = 1$.

MDA of the Gumbel Distribution Λ

The maximum domain of attraction of the Gumbel distribution $\Lambda(x) = \exp\{e^{-x}\}$, $x \in \mathbb{R}$, contains a large variety of distributions with finite or infinite right endpoint. This maximum domain of attraction cannot be characterised by a simple regular variation condition. However, the necessary and sufficient conditions given below to some extent remind us the representation of regularly varying functions; see Theorem 1.2.1.

For a proof of the following result we refer to Resnick [100], Corollary 1.7 and Proposition 1.9.

Theorem 1.4.28 (Characterisation I of MDA(Λ))
The distribution F with right endpoint $x_F \le \infty$ belongs to the maximum domain of attraction of Λ if and only if there exists some $z < x_F$ such that F has representation

$$\overline{F}(x) = c(x) \exp\left\{ -\int_z^x \frac{g(t)}{a(t)}\, dt \right\}, \quad z < x < x_F,$$

where c and g are measurable functions satisfying $c(x) \to c > 0$, $g(x) \to 1$ as $x \uparrow x_F$, and $a(x)$ is a positive, absolutely continuous function (with respect to Lebesgue measure) with density $a'(x)$ having $\lim_{x \uparrow x_F} a'(x) = 0$.

A possible choice for the function a is

$$a(x) = \int_x^{x_F} \frac{\overline{F}(t)}{\overline{F}(x)}\, dt, \qquad x < x_F, \tag{1.17}$$

Remark 1.4.29 For a random variable X the function $a(x)$ as defined in (1.17) is nothing but the *mean excess function*

$$a(x) = E(X - x \mid X > x), \quad x < x_F.$$

Another characterisation of MDA(Λ) is the following.

Theorem 1.4.30 (Characterisation II of MDA(Λ))
The distribution F belongs to the maximum domain of attraction of Λ if and only if there exists some positive function $\widetilde{a}$ such that

$$\lim_{x \uparrow x_F} \frac{\overline{F}(x + t\widetilde{a}(x))}{\overline{F}(x)} = e^{-t}, \quad t \in \mathbb{R}.$$

holds. A possible choice is $\widetilde{a} = a$ as given in (1.17)

The proof of this result is for instance to be found in de Haan [52], Theorem 2.5.1.

Although MDA(Λ) cannnot be characterised by standard regular variation conditions it is closely related to *rapid variation.* Recall that a positive measurable function h on $[0, \infty)$ is *rapidly varying* with index $-\infty$ if

$$\lim_{x \to \infty} \frac{h(tx)}{h(x)} = \begin{cases} 0 & \text{if } t > 1, \\ \infty & \text{if } 0 < t < 1. \end{cases}$$

An example is given by $h(x) = e^{-x}$. We mention that various results for regularly varying functions can be extended to rapidly varying functions, for example the representation theorem and Karamata's theorem. See de Haan [52]. It is not difficult to see that all power moments of a distribution with rapidly varying tails exist and are finite,

Corollary 1.4.31 (Existence of moments)
Assume that the random variable X has distribution $F \in$ MDA(Λ) with infinite right endpoint. Then $\overline{F}$ is rapidly varying In particular, $E(X^+)^\alpha < \infty$ for every $\alpha > 0$, where $X^+ = \max(0, X)$.

Notes and Comments

Extreme value theory is a classical topic in probability theory and mathematical statistics. Its origins go back to Fisher and Tippett [42]. Since then a large number of books and articles on extreme value theory has appeared. The interested reader may, for instance, consult the following textbooks: Falk, Hüsler and Reiss [38], Gumbel [51], Leadbetter, Lindgren and Rootzén [66], Reiss [99] and Resnick [100].

Theorem 1.4.12 marked the beginning of extreme value theory as one of the central topics in probability theory and statistics. The limit laws for maxima were derived by Fisher and Tippett [42]. A first rigorous proof is due to Gnedenko [43]. De Haan [52] subsequently applied regular variation as an analytical tool. His work has been of great importance for the development of modern extreme value theory.

1.4.3 Stochastic Recurrence Equations

Probabilistic Properties

We consider a sequence (X_t) of random variables satisfying the *stochastic recurrence equation* (SRE)

$$X_t = A_t X_{t-1} + B_t\,, \quad t \in \mathbb{Z}\,, \tag{1.18}$$

where the sequence $((A_t, B_t))$ is supposed to be iid. We assume that (X_t) is a causal strictly stationary solution of (1.18).

There exist various results about the existence of a strictly stationary solution to (1.18); see for example Kesten [61], Vervaat [111], Grincevičius [49], Brandt [11], Bougerol and Picard [9]. Below we give a sufficient condition which remains valid for ergodic sequences $((A_n, B_n))$ (see Brandt [11]) and which is close to necessity (see Babillot et al. [4]).

Theorem 1.4.32 (Existence of stationary solution)
Assume $-\infty \le E\ln|A| < 0$ *and* $E\ln^+|B| < \infty$*. Then the series*

$$X_n = \sum_{k=0}^{\infty} A_n \cdots A_{n-k+1}\, B_{n-k} \tag{1.19}$$

converges a.s., and the so–defined process (X_n) *is the unique causal strictly stationary solution of* (1.18).

Remark 1.4.33 A glance at formula (1.19) convinces one that products of A_ts and B_ts are the main ingredients to the stationary solution of (1.18). Assuming the A_ts and B_ts positive for the moment, one can see that the distribution of X is determined by exponentials of sums of the form $\sum_{t=n-k+1}^{n} \ln A_t$. The condition $E\ln A < 0$ ensures that these sums constitute a random walk with negative drift, and therefore the products $\Pi_k = A_n \cdots A_{n-k+1}$ in (1.19) decay to zero at an exponential rate. This implies the a.s. summability of the infinite series. The right tail of X is essentially determined by the products Π_k as well. Indeed, if $P(A > 1) > 0$, Π_k may exceed 1 finitely often with positive probability. It turns out that the tail of X is then basically determined by the distribution of $\exp\{\max_k \ln \Pi_k\}$, and so a renewal argument for determining the distribution of a random walk with negative drift can be applied. Below (Theorem 1.4.35) we will see that this naive argument can be made precise.

Remark 1.4.34 The Markov chain (X_n) satisfies a mixing condition under quite general conditions as for example provided in Meyn and Tweedie [76]. In particular, for the SRE in (1.18), suppose there exists an $\epsilon \in (0,1]$ such that $E|A|^\epsilon < 1$ and $E|B|^\epsilon < \infty$. Then there exists a unique stationary solution to (1.18) and the Markov chain (X_n) is geometrically ergodic.

Under general conditions, the stationary solution to the stochastic recurrence equation (1.18) satisfies a regular variation condition. This follows from work by Kesten [61]; see also Goldie [45] and Grey [48]. We state a modification of Kesten's fundamental result (Theorems 3 and 4 in [61]).

Theorem 1.4.35 (Kesten's theorem)
Let (A_n) be an iid sequence of non–negative random variables satisfying:

- *For some $\epsilon > 0$, $EA^\epsilon < 1$.*
- *A has a density with support $[0, \infty)$.*
- *There exists a $\kappa_0 > 0$ such that*

$$1 \le EA^{\kappa_0} \quad \textit{and} \quad E\left(A_0^{\kappa} \ln^+ A\right) < \infty\,.$$

Then there exists a unique solution $\kappa_1 \in (0, \kappa_0]$ to the equation $1 = EA^{\kappa_1}$.
If (X_n) is the stationary solution to the SRE in (1.18) with coefficients (A_n) satisfying the above conditions and B is non-negative with $EB^{\kappa_1} < \infty$, then X is regularly varying with index κ_1.

Remark 1.4.36 There are extensions of Theorem 1.4.35 to general A and B and to the multivariate case. Without the positivity constraints, the required conditions can be quite cumbersome. See Kesten [61] and Le Page [67].

Remark 1.4.37 Resnick and Willekens [102] considered SREs under slightly different conditions than those imposed in Theorem 1.4.35. They assume that (X_t) satisfies (1.18) with the additional condition that A_t and B_t are independent for every t, that B is regularly varying with index α and $E|A|^{\alpha+\epsilon} < \infty$ for some $\epsilon > 0$. Then X is also regularly varying with index α. Notice that the moment condition for B in Theorem 1.4.35 is not satisfied for this model.

GARCH models

Log–returns $X_t = \ln P_t - \ln P_{t-1}$ of foreign exchange rates, stock indices and share prices P_t, $t = 1, 2, \ldots$, typically share the following features:

- The frequency of large and small values (relative to the range of the data) is rather high, suggesting that *the data do not come from a normal, but from a heavy–tailed distribution.*
- Exceedances of high thresholds occur in clusters, which indicates that *there is dependence in the tails.*

Various models have been proposed in order to describe these empiricially observed features. Among them, models of the type

$$X_t = \sigma_t\, Z_t\,, \quad t \in \mathbb{Z}\,,$$

have become particularly popular. Here (Z_t) is a sequence of iid symmetric random variables with $EZ_1^2 = 1$. One often assumes the Z_ts to be standard normal. Moreover, the sequence (σ_t) consists of non–negative random variables such that Z_t and σ_t are independent for every fixed t. Models of this type include the ARCH (autoregressive conditionally heteroscedasticity) and GARCH (generalised ARCH) family; see for example Engle [36] for their definitions and properties. In what follows, we often write σ for a generic random variable with the distribution of σ_1, X for a generic random variable with the distribution of X_1, etc.

We restrict ourselves to one particular model which has very often been used in applications: the GARCH$(1,1)$ process. It is defined by specifying σ_t as follows:

$$\sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 X_{t-1}^2 = \alpha_0 + \sigma_{t-1}^2(\beta_1 + \alpha_1 Z_{t-1}^2)\,, \quad t \in \mathbb{Z}\,.$$

The parameters α_0, α_1 and β_1 are non–negative.

Despite its simplicity, the stationary GARCH$(1,1)$ process is believed to capture various of the empirically observed properties of log–returns. For example, the stationary GARCH$(1,1)$ processes can exhibit heavy–tailed marginal distributions of power law type and hence they could be appropriate tools to model the heavier-than-normal tails of the financial data. This follows from Kesten's Theorem 1.4.35.

The GARCH$(1,1)$ can be considered in the much wider context of SREs of type (1.18). Observe that σ_t^2 satisfies the recurrence equation

$$\sigma_t^2 = \alpha_0 + \sigma_{t-1}^2(\alpha_1 Z_{t-1}^2 + \beta_1)\,, \quad t \in \mathbb{Z}\,, \tag{1.20}$$

which is of the same type as (1.18), with $X_t = \sigma_t^2$, $A_t = \alpha_1 Z_{t-1}^2 + \beta_1$ and $B_t = \alpha_0$.

From Theorem 1.4.32, in combination with Babillot et al. [4] or Bougerol and Picard [10] one can derive the following result:

Corollary 1.4.38 *The conditions*

$$\alpha_0 > 0 \quad \textit{and} \quad E\ln(\alpha_1 Z^2 + \beta_1) < 0 \tag{1.21}$$

are necessary and sufficient for stationarity of (σ_t^2).

Remark 1.4.39 Notice that stationarity of σ_t^2 implies stationarity of the sequence

$$(X_t^2, \sigma_t^2) = \sigma_t^2\;(Z_t^2, 1)\,, \quad t \in \mathbb{Z}\,.$$

By construction of the sequence (X_t), stationarity of the sequence $((X_t, \sigma_t))$ follows.

In what follows, we always assume that condition (1.21) *is satisfied.* Then a stationary version of $((X_t, \sigma_t))$ exists.

Kesten's Theorem 1.4.35 immediately yields the following result for the tails of X and σ.

Corollary 1.4.40 *Assume Z has a density with unbounded support,* (1.21) *holds,*

$$E|\alpha_1 Z^2 + \beta_1|^{\kappa_0/2} \geq 1 \quad \textit{and} \quad E|Z|^{\kappa_0} \ln^+|Z| < \infty \quad \textit{for some } \kappa_0 > 0.$$

A) There exists a number $\kappa_1 \in (0, \kappa_0]$ which is the unique solution of the equation

$$E(\alpha_1 Z^2 + \beta_1)^{\kappa_1/2} = 1\,, \tag{1.22}$$

and there exists a positive constant $c_0 = c_0(\alpha_0, \alpha_1, \beta_1)$ such that

$$P(\sigma > x) \sim c_0\, x^{-\kappa_1} \quad \textit{as } x \to \infty.$$

B) If $E|Z|^{\kappa_1+\epsilon} < \infty$ for some $\epsilon > 0$, then

$$P(|X| > x) \sim E|Z|^{\kappa_1}\; P(\sigma > x)\,. \tag{1.23}$$

Proof. Part A follows from an application of Theorem 1.4.35 to the SRE (1.20). Equation (1.23) is a consequence of Breiman's result; see Proposition 1.3.9. □

Remark 1.4.41 The exact value of the constant c_0 is given in Goldie [45].

Remark 1.4.42 Under the assumptions of Theorem 1.4.40, $\alpha_1 = 0$ is not a possible parameter choice.

Remark 1.4.43 Assume in addition to the conditions of Theorem 1.4.40 that $EZ^2 = 1$, $E|Z|^{2+\epsilon} < \infty$ for some $\epsilon > 0$ and $\alpha_1 + \beta_1 = 1$. Then (1.22) has the unique solution $\kappa_1 = 2$. This implies that $P(|X| > x) \sim c\, x^{-2}$ for some $c > 0$ and, in turn, that $EX^2 = \infty$. GARCH$(1,1)$ models fitted to real log–returns frequently have parameters α_1 and β_1 such that $\alpha_1 + \beta_1$ is close to 1. This indicates that one deals with time series models with extremely heavy tails. This fact led Engle and Bollerslev [37] to the introduction of the IGARCH (integrated GARCH) model. Mikosch and Stărică [80, 81] give a critical analysis of IGARCH. In particular, they explain in [81] that the $\alpha_1 + \beta_1 \approx 1$ effect can be due GARCH$(1,1)$ misspecification in presence of non–stationarity in real–life data.

Notes and Comments

In addition to the literature mentioned, an introduction to one–dimensional SREs can be found in Embrechts et al. [34], Section 8.4. There Kesten's theorem is proved in the particular case of an ARCH process. The general one-dimensional case is masterly treated in Goldie [45]. Kesten's theorem is remarkable insofar that light–tailed input (the sequence $((A_t, B_t))$) causes heavy–tailed output (the stationary solution (X_t) of (1.18)). This is in contrast to linear processes where only heavy–tailed input (innovations) can cause heavy–tailed output.

The case of general GARCH processes is treated in Davis, Mikosch and Basrak [25]. An approach similar to GARCH$(1,1)$ is possible for bilinear processes; they can also exhibit power law tails for light–tailed innovations; see Basrak, Davis and Mikosch [5] and Turkman and Turkman [109].

1.4.4 Estimation of a Regularly Varying Tail

Among the statistical estimators for the parameter α of a regularly varying tail $\overline{F}(x) = x^{-\alpha}L(x)$, *Hill's estimator* has become particularly popular. In what follows we explain the rationale behind it.

Suppose $X_1, \ldots, X_n$ are iid with distribution F satisfying $\overline{F}(x) = x^{-\alpha}L(x)$, $x > 0$, for a slowly varying function L and some $\alpha > 0$. For many applications the knowledge of the index α is of major importance. If for instance $\alpha < 2$ then $EX_1^2 = \infty$. This case is often observed in the modelling of insurance data; see for instance Hogg and Klugman [57]. Empirical studies on the tails of daily log–returns in finance have indicated that one frequently encounters values α between 3 and 4; see for instance Guillaume et al. [50], Longin [72] or Loretan and Phillips [73]. Information of the latter type implies that, whereas covariances of such data would be well defined, the construction of confidence intervals for the sample autocovariances and autocorrelations on the basis of asymptotic (central limit) theory may be questionable as typically a finite fourth moment condition is asked for.

The Hill estimator of α essentially takes on the following form:

$$\widehat{\alpha}^{(H)} = \widehat{\alpha}^{(H)}_{k,n} = \left(\frac{1}{k} \sum_{j=1}^{k} \ln X_{j,n} - \ln X_{k,n} \right)^{-1}, \qquad (1.24)$$

where $k = k(n) \to \infty$ in an appropriate way and

$$X_{n,n} \le \cdots \le X_{1,n} = M_n$$

denote the order statistics of the sample $X_1, \ldots, X_n$. This means that an increasing sequence of upper order statistics is used. One of the interesting facts concerning

(1.24) is that various asymptotically equivalent versions of $\widehat{\alpha}^{(H)}$ can be derived through essentially different methods, showing that the Hill estimator is very natural. Below we discuss some derivations.

The MLE Approach (Hill [56])

Assume for the moment that X is a random variable with distribution F so that for $\alpha > 0$

$$P(X > x) = \overline{F}(x) = x^{-\alpha}\,, \quad x \geq 1\,.$$

Then it immediately follows that $Y = \ln X$ has distribution

$$P(Y > y) = \mathrm{e}^{-\alpha y}\,, \quad y \geq 0\,,$$

i.e. Y is $Exp(\alpha)$ and hence the MLE of α is given by

$$\widehat{\alpha}_n = \overline{Y}_n^{\,-1} = \left(\frac{1}{n}\sum_{j=1}^{n} \ln X_j\right)^{-1} = \left(\frac{1}{n}\sum_{j=1}^{n} \ln X_{j,n}\right)^{-1}\,.$$

A trivial generalisation concerns

$$\overline{F}(x) = C x^{-\alpha}\,, \quad x \geq u > 0\,, \tag{1.25}$$

with u known. If we interpret (1.25) as fully specified, i.e. $C = u^{\alpha}$, then we immediately obtain as MLE of α

$$\widehat{\alpha}_n = \left(\frac{1}{n}\sum_{j=1}^{n} \ln\left(\frac{X_{j,n}}{u}\right)\right)^{-1} = \left(\frac{1}{n}\sum_{j=1}^{n} \ln X_{j,n} - \ln u\right)^{-1}\,. \tag{1.26}$$

Now we often do not have the precise parametric information of these examples, but in the spirit of MDA(Φ_α) we assume that $\overline{F}$ behaves like a Pareto distribution function above a certain known threshold u say. Let

$$K = \mathrm{card}\,\{i : X_{i,n} > u\,, i = 1, \ldots, n\}\,. \tag{1.27}$$

Conditionally on the event $\{K = k\}$, maximum likelihood estimation of α and C in (1.25) reduces to maximising the joint density of $(X_{k,n}, \ldots, X_{1,n})$. One can show that

$$\begin{aligned} & f_{X_{k,n},\ldots,X_{1,n}}(x_k, \ldots, x_1) \\ = \;& \frac{n!}{(n-k)!}\left(1 - C x_k^{-\alpha}\right)^{n-k} C^k \alpha^k \prod_{i=1}^{k} x_i^{-(\alpha+1)}\,, \quad u < x_k < \cdots < x_1\,. \end{aligned}$$

A straightforward calculation yields the conditional MLEs

$$\begin{aligned} \widehat{\alpha}_{k,n}^{(H)} &= \left(\frac{1}{k}\sum_{j=1}^{k} \ln\left(\frac{X_{j,n}}{X_{k,n}}\right)\right)^{-1} = \left(\frac{1}{k}\sum_{j=1}^{k} \ln X_{j,n} - \ln X_{k,n}\right)^{-1} \\ \widehat{C}_{k,n} &= \frac{k}{n} X_{k,n}^{\widehat{\alpha}_{k,n}^{(H)}}\,. \end{aligned}$$

So Hill's estimator has the same form as the MLE in the exact model underlying (1.26) but now having *the deterministic* u *replaced by the random threshold* $X_{k,n}$, where k is defined through (1.27). We also immediately obtain an estimate for the tail $\overline{F}(x)$

$$\left(\overline{F}(x)\right)\widehat{\ } = \frac{k}{n}\left(\frac{x}{X_{k,n}}\right)^{-\widehat{\alpha}_{k,n}^{(H)}} \tag{1.28}$$

and for the p–quantile

$$\widehat{x}_p = \left(\frac{n}{k}(1-p)\right)^{-1/\widehat{\alpha}_{k,n}^{(H)}} X_{k,n}\,. \tag{1.29}$$

From (1.28) we obtain an estimator of the excess distribution function $F_u(x-u)$, $x \geq u$, by using $F_u(x-u) = 1 - \overline{F}(x)/\overline{F}(u)$.

The Regular Variation Approach (de Haan [53])

This approach is based on a a suitable reformulation of the regular variation condition on the right tail. Indeed $\overline{F}$ is regularly varying with index $-\alpha$ if and only if

$$\lim_{t\to\infty} \frac{\overline{F}(tx)}{\overline{F}(t)} = x^{-\alpha}\,, \quad x > 0\,.$$

Using partial integration, we obtain

$$\int_t^\infty (\ln x - \ln t)\, dF(x) = \int_t^\infty \frac{\overline{F}(x)}{x}\, dx\,,$$

so that by Karamata's Theorem 1.2.6,

$$\frac{1}{\overline{F}(t)}\int_t^\infty (\ln x - \ln t)\, dF(x) \to \frac{1}{\alpha}\,, \quad t \to \infty\,. \tag{1.30}$$

How do we find an estimator from this result? Two choices have to be made:

(a) replace F by an estimator, the obvious candidate here is the empirical distribution function

$$F_n(x) = \frac{1}{n}\sum_{i=1}^n I_{\{X_i \leq x\}} = \frac{1}{n}\sum_{i=1}^n I_{\{X_{i,n} \leq x\}}\,,$$

(b) replace t by an appropriate high, data dependent level (recall $t \to \infty$); we take $t = X_{k,n}$ for some $k = k(n)$.

The choice of t is motivated by the fact that $X_{k,n} \xrightarrow{\text{a.s.}} \infty$ provided $k = k(n) \to \infty$ and $k/n \to 0$. From (1.30) the following estimator results

$$\frac{1}{\overline{F}\left(X_{k,n}\right)}\int_{X_{k,n}}^\infty \left(\ln x - \ln X_{k,n}\right)\, dF_n(x) = \frac{1}{k-1}\sum_{j=1}^{k-1} \ln X_{j,n} - \ln X_{k,n}$$

which, modulo the factor $k-1$, is again of the form $(\widehat{\alpha}^{(H)})^{-1}$ in (1.24). Notice that the change from k to $k-1$ is asymptotically negligible.

We summarise as follows.

Suppose $X_1, \dots, X_n$ are iid with distribution such that $\overline{F}$ is regularly varying with index $-\alpha$ for some $\alpha > 0$. Then a natural estimator for α is provided by *Hill's estimator*

$$\widehat{\alpha}_{k,n}^{(H)} = \left(\frac{1}{k} \sum_{j=1}^{k} \ln X_{j,n} - \ln X_{k,n} \right)^{-1}, \tag{1.31}$$

where $k = k(n)$ satisfies $k \to \infty$ and $k = o(n)$.

Below we summarise the main properties of the Hill estimator.

Theorem 1.4.44 (Properties of the Hill estimator)
Suppose (X_n) is strictly stationary with marginal distribution F satisfying for some $\alpha > 0$ and slowly varying L,

$$\overline{F}(x) = P(X > x) = x^{-\alpha} L(x), \quad x > 0.$$

Let $\widehat{\alpha}^{(H)} = \widehat{\alpha}_{k,n}^{(H)}$ be the Hill estimator (1.31).

(*a*) (Weak consistency) *Assume that one of the following conditions is satisfied:*

- *(X_n) is iid,*
- *(X_n) is weakly dependent,*
- *(X_n) is a linear process.*

If $k \to \infty$, $k/n \to 0$ for $n \to \infty$, then

$$\widehat{\alpha}^{(H)} \xrightarrow{P} \alpha.$$

(*b*) (Strong consistency) *If $k/n \to 0$, $k/\ln\ln n \to \infty$ for $n \to \infty$ and (X_n) is an iid sequence, then*

$$\widehat{\alpha}^{(H)} \xrightarrow{\text{a.s.}} \alpha.$$

(*c*) (Asymptotic normality) *If further conditions on k and F are satisfied and (X_n) is an iid sequence, then*

$$\sqrt{k}\left(\widehat{\alpha}^{(H)} - \alpha\right) \xrightarrow{d} N\left(0, \alpha^2\right). \qquad \square$$

Notes and Comments

There exists a whole industry on tail and quantile estimation. In particular, there exist many different estimators of the parameter α. Next to Hill's estimator, the Pickands and the Dekkers–Einmahl–de Haan estimators are most popular. As Hill's estimator, they are constructed from the upper order statistics in the sample. Their derivation is based on similar ideas as above, i.e. on reformulations of the regular variation condition. Moreover, the latter estimators can also be used in the more general context of estimating the extreme value index.

Although the tail estimators seem to be fine at the first sight their practical implementation requires a lot of experience and skill.

- One has to *choose the right number $k(n)$ of upper order statistics.* This is an art. Or some kind of educated guessing.

- One needs *large sample sizes.* A couple of hundred data is usually not enough. Indeed, since α is a tail parameter only the very large values in the sample can tell us about the size of α, and in small samples there are not sufficiently many large values.
- *Deviations from pure power laws* make these estimators extremely unreliable. This can be shown by simulations and also theoretically. Although, in principle, the theory allows one to adjust the estimators by using the concrete form of the slowly varying function, in practice we never know what these functions are.
- The mentioned estimators behave poorly *when the data are dependent.* Although for all these cases theoretical solutions have been found in the sense that the asymptotic estimation theory (consistency and asymptotic normality) works even for weakly dependent stationary data, simulation studies show that the estimation of α is becoming even more a problem for dependent data.

An honest discussion of tail estimation procedures can be found in Embrechts et al. [34], Chapter 6. There one can also find a large number of references.

1.5 Multivariate Regular Variation

This section is based on Davis et al. [25].

If one wants to deal with *dependence and regular variation* it is necessary to introduce the notion of regular variation of random vectors or multivariate regular variation.

Recall that a one-dimensional non-negative random variable X and its distribution are said to be regularly varying if the tail distribution can be written as

$$P(X > x) = x^{-\alpha} L(x)\,, \quad x > 0\,. \tag{1.32}$$

This relation can be extended to the multivariate setting in very different ways. We consider some of the possible definitions.

In what follows, we write $\mathbf{a} \le \mathbf{b}, \mathbf{a} < \mathbf{b}, \mathbf{a} > \mathbf{b}$, etc., where $\le, <, >, \ldots$ refer to the natural (componentwise) partial ordering in $\mathbb{R}^d$. We say that *the vector* $\mathbf{x}$ *is positive* ($\mathbf{x} > \mathbf{0}$) if each of its components is positive. We also use the following notation:

$$\mathbf{e} = (1,\ldots,1)\,, \quad \mathbf{0} = (0,\ldots,0)\,, \quad [\mathbf{a},\mathbf{b}] = \{\mathbf{x} : \mathbf{a} \le \mathbf{x} \le \mathbf{b}\}\,.$$

Unless stated otherwise, we assume in what follows that $\mathbf{X}$ *is positive with probability* 1. *However, in such cases we will choose the state space* $D = [0,\infty]^d \backslash \{\mathbf{0}\}$.

Condition R1. The following limit exists and is finite for all $\mathbf{x} > \mathbf{0}$:

$$\lim_{t\to\infty} \frac{P(\mathbf{X} \in t\,[\mathbf{0},\mathbf{x}]^c)}{P(\mathbf{X} \in t\,[\mathbf{0},\mathbf{e}]^c)}\,,$$

where the complements are taken in the state space D.

Condition R2. There exist $\alpha \ge 0$ and a random vector $\boldsymbol{\Theta}$ with values in

$$\mathbb{S}_+^{d-1} = \mathbb{S}^{d-1} \cap [0,\infty)^d$$

such that the following limit exists for all $x > 0$:

$$\frac{P(|\mathbf{X}| > t\,x\,,\, \mathbf{X}/|\mathbf{X}| \in \cdot\,)}{P(|\mathbf{X}| > t)} \stackrel{v}{\longrightarrow} x^{-\alpha}\, P_{\boldsymbol{\Theta}}(\cdot)\,, \quad t \to \infty\,, \tag{1.33}$$

where $\xrightarrow{v}$ denotes vague convergence on the Borel σ-field of $\mathbb{S}_+^{d-1}$ and $P_{\mathbf{\Theta}}$ is the distribution of $\mathbf{\Theta}$.

Condition R3. There exist a non-zero Radon measure μ on D and a relatively compact Borel set $E \subset D$ such that

$$\mu_t(\,\cdot\,) := \frac{P(\mathbf{X} \in t\,\cdot\,)}{P(\mathbf{X} \in t\,E)} \xrightarrow{v} \mu(\,\cdot\,)\,, \quad t \to \infty\,,$$

where $\xrightarrow{v}$ denotes vague convergence on the Borel σ-field of D.

Condition R4. The following limit exists and is finite for all $\mathbf{x} > \mathbf{0}$:

$$\lim_{t\to\infty} \frac{P((\mathbf{x},\mathbf{X}) > t)}{P((\mathbf{e},\mathbf{X}) > t)}\,.$$

Remark 1.5.1 Conditions R1 and R2 are frequently used in extreme value theory in order to characterize maximum domains of attractions of extreme value distributions; see for example Resnick [100]. They are also used for the characterization of domains of attractions of stable laws in the theory of sums for independent random vectors; see for example Araujo and Giné [1]. Condition R3 has been introduced by Meerschaert [75] in the context of regularly varying measures. Condition R4 was used by Kesten [61] in order to characterize the tail behaviour of solutions to stochastic recurrence equations; see Section 1.4.3.

Remark 1.5.2 If R3 holds there exists an $\alpha \geq 0$ such that $\mu(vS) = v^{-\alpha}\mu(S)$ for every Borel set $S \subset D$ and $v > 0$. If one uses polar coordinates $(r,\theta) = (|\mathbf{x}|, \mathbf{x}/|\mathbf{x}|)$ then it is convenient to describe the measure μ as $\alpha r^{-\alpha-1}dr \times P_{\mathbf{\Theta}}(d\theta)$, where $P_{\mathbf{\Theta}}$ is a probability measure on the Borel σ-field of $\mathbb{S}_+^{d-1}$. The case $\alpha = 0$ corresponds to the measure $\varepsilon_{\{\infty\}}(dr) \times P_{\mathbf{\Theta}}(d\theta)$, where $\varepsilon_{\mathbf{x}}$ denotes Dirac measure at $\mathbf{x}$.

Remark 1.5.3 R1–R4 can be formally extended to the case $\alpha = \infty$. For example, R2 can be formulated as follows: For a given $P_{\mathbf{\Theta}}$-continuity set S, the left-hand side probability in (1.33) converges to 0 or ∞ according as $x \in (0,1)$ or $x > 1$. For $d = 1$, the case $\alpha = \infty$ corresponds to rapid variation; see Bingham et al. [6].

Remark 1.5.4 The limits in R1–R4 can be compared by choosing in R3 the particular sets:

- for R1: $E = [\mathbf{0},\mathbf{e}]^c$,
- for R2: $E = \{\mathbf{x} > \mathbf{0} : \mathbf{x}/|\mathbf{x}| \in \mathbb{S}_+^{\mathbf{d-1}}, |\mathbf{x}| > \mathbf{1}\}$,
- for R4: $E = \{\mathbf{y} : (\mathbf{y},\mathbf{e}) > 1\}$

Remark 1.5.5 If R3 holds for some set E, it holds for any bounded Borel set $E \subset D$ with $\mu(\partial E) = 0$. It is also straightforward to show that this condition is equivalent to the existence of a Radon measure ν on D and a sequence (a_n), $a_n \to \infty$ such that

$$n\,P(a_n^{-1}\mathbf{X} \in \,\cdot\,) \xrightarrow{v} \nu(\,\cdot\,)\,, \tag{1.34}$$

where ν is a measure with the property that $\nu(E) > 0$ for at least one relatively compact set $E \subset D$. The other conditions can be represented similarly.

Theorem 1.5.6 *Conditions R1–R4 are equivalent.*

Remark 1.5.7 One can show that R2 and R3 are equivalent on the enlarged state space $D = \overline{\mathbb{R}}^d \backslash \{\mathbf{0}\}$ provided one replaces $\mathbb{S}_+^{d-1}$ with $\mathbb{S}^{d-1}$ in R2.

In view of Theorem 1.5.6 and Remark 1.5.7 we adopt the following definition for regular variation:

Definition 1.5.8 *The random vector* $\mathbf{X}$ *with values in* $\mathbb{R}^d$ *is said to be* regularly varying with index α and spectral measure $P_{\mathbf{\Theta}}$ *if condition R2 holds.*

1.5.1 Functions of regularly varying vectors

In what follows we consider a regularly varying random vector $\mathbf{X} > \mathbf{0}$ with index $\alpha \geq 0$ and spectral measure $P_{\mathbf{\Theta}}$. We consider suitable transformations of $\mathbf{X}$ such that the transformed vector is again regularly varying.

Our first result tells us that regular variation is preserved under power transformations. Define for any vector $\mathbf{x} > \mathbf{0}$ and $p > 0$,

$$\mathbf{x}^p = (x_1^p, \dots, x_d^p) .$$

Proposition 1.5.9 *Assume* $\mathbf{X}$ *is regularly varying in the sense of R2 where* $|\cdot|$ *denotes the max-norm. For every* $p > 0$, $\mathbf{X}^p$ *is regularly varying with index* α/p *and spectral measure* $P_{\mathbf{\Theta}^p}$.

Our next theorem extends Breiman's [13] result for products of independent random variables to the case $d > 1$. Recall from Proposition 1.3.9 that for any independent non-negative random variables ξ and η such that ξ is regularly varying with index α and $E\eta^{\alpha+\epsilon} < \infty$ for some $\epsilon > 0$,

$$P(\xi\, \eta > x) \sim E\eta^{\alpha} P(\xi > x) . \tag{1.35}$$

In view of (1.34) and Remark 1.5.5 the regular variation condition can be formulated as follows: there exists a sequence (a_n) and a measure μ on D such that

$$\frac{P\left(a_n^{-1}\, \mathbf{X} \in \cdot\right)}{P(|\mathbf{X}| > a_n)} \xrightarrow{v} \mu(\cdot) ,$$

where $\xrightarrow{v}$ denotes vague convergence on the Borel σ-field of D.

The multivariate version of Breiman's result reads as follows.

Proposition 1.5.10 *Let* $\mathbf{A}$ *be a random* $q \times d$ *matrix, positive with probability* 1 *and independent of* $\mathbf{X}$ *which vector satisfies* (1.34). *Also assume that* $E\|\mathbf{A}\|^{\gamma} < \infty$ *for some* $\gamma > \alpha$. *Then*

$$n\, P\left(a_n^{-1}\mathbf{A}\mathbf{X} \in \cdot\right) \xrightarrow{v} \widetilde{\mu}(\cdot) := E[\mu \circ \mathbf{A}^{-1}(\cdot)] ,$$

where $\xrightarrow{v}$ *denotes vague convergence on the Borel* σ*-field of* D *and* $\mathbf{A}^{-1}$ *is the inverse image of* $\mathbf{A}$.

Remark 1.5.11 Assume that $\mathbf{B}$ is a d-dimensional random vector such that

$$P(|\mathbf{B}| > x) = o(P(|\mathbf{X}| > x)) \quad \text{as } x \to \infty$$

and the conditions of Proposition 1.5.10 hold. Then $\mathbf{AX} + \mathbf{B}$ is regularly varying with the same limit measure $\widetilde{\mu}$. Moreover, if $\mathbf{B}$ itself is regularly varying with index α and independent of $\mathbf{AX}$, then $\mathbf{AX} + \mathbf{B}$ is regularly varying with index α. This follows from the fact that $(\mathbf{AX}, \mathbf{B})$ is regularly varying.

In what follows, $\mathbf{X} = (X_1, \dots, X_d)$ may assume values in $\mathbb{R}^d$.

Proposition 1.5.12 *Let* $\mathbf{X}$ *be regularly varying with index* α *and* $Y_1, \dots, Y_d$ *be independent random variables such that* $E|Y_i|^{\alpha+\epsilon} < \infty$ *for some* $\epsilon > 0$, $i = 1, \dots, d$. *Then* $(Y_1 X_1, \dots, Y_d X_d)$ *is regularly varying with index* α.

A particular consequence is the following.

Proposition 1.5.13 *Let* $\varepsilon = (\varepsilon_1, \dots, \varepsilon_d)$ *be a vector of iid Bernoulli random variables such that* $P(\varepsilon_1 = \pm 1) = 0.5$. *Also assume that* ε *and* $\mathbf{X}$ *are independent. Then* $\mathbf{Y} = (\varepsilon_1 X_1, \dots, \varepsilon_d X_d)$ *is regularly varying with index* α *and spectral measure* $P_{(\varepsilon_1 \theta_1, \dots, \varepsilon_d \theta_d)}$.

Here is a result about the regular variation of products of dependent random variables.

Proposition 1.5.14 *If* $\mathbf{X}$ *is regularly varying, For every* $h = 1, \ldots, d$*, the random vector* $X_h \mathbf{X}$ *is regularly varying with index* $\alpha/2$*.*

Remark 1.5.15 Notice that in the dependent case the product of regularly varying random variables is regularly varying with index $\alpha/2$. This is in contrast to Proposition 1.3.9 for products of independent random variables which says that this product is regularly varying with index α. In the independent case, the vector Θ from definition (1.33) is concentrated on the axes. This implies that the limit in (1.33) is the null measure.

2

Subexponential Distributions

2.1 Definition

Let $X, X_1, X_2, \ldots$ be iid non–negative random variables. From Corollary 1.3.6 and Remark 1.3.7 we learnt that, for regularly varying X,

$$P(S_n > x) \sim n\, P(X > x) \sim P(M_n > x) \quad \text{as } x \to \infty \text{ for } n = 2, 3, \ldots, \tag{2.1}$$

where

$$S_n = X_1 + \cdots + X_n \quad \text{and} \quad M_n = \max_{i=1,\ldots,n} X_i\,.$$

The intuitive interpretation of (2.1) is that the maximum M_n of $X_1, \ldots, X_n$ makes a major contribution to its sum: exceedances of high thresholds by the sum S_n are due to the exceedance of this threshold by the largest value in the sample. This interpretation suggests one way of defining a "heavy–tailed" distribution: the tail of the sum is essentialy determined by the tail of the maximum. This intuitive approach leads to the definition of a sufficiently large class of "heavy–tailed" distributions.

Definition 2.1.1 (Subexponential distribution)
A non–negative random variable X and its distribution is said to be subexponential *if iid copies X_i of X satisfy relation* (2.1). *The class of subexpenential distributions is denoted by $\mathcal{S}$.*

Remark 2.1.2 Chistyakov [15] proved that (2.1) holds for all $n \geq 2$ if and only if it holds for $n = 2$. Embrechts and Goldie [30] showed that (2.1) holds for all $n \geq 2$ if it holds for some $n \geq 2$. Moreover, it suffices to require that the relation

$$\limsup_{x \to \infty} \frac{P(S_n > x)}{n\, P(X > x)} \leq 1$$

holds for some $n \geq 2$; see Lemmas 1.3.4 and A3.14 in Embrechts et al. [34].

Notes and Comments

The class of subexponential distributions was independently introduced by Chistyakov [15] and Chover, Ney and Wainger [16] mainly in the context of branching processes. An early textbook treatment is given in Athreya and Ney [3]. An independent introduction of $\mathcal{S}$ through questions in queuing theory is to be found in Borovkov [7, 8]; see also Pakes [92]. The importance of $\mathcal{S}$ as a useful class of heavy–tailed distribution functions in the context of applied probability in general, and insurance mathemactics in particular, was realised early on by Teugels [107]. A recent survey paper is Goldie and Klüppelberg [46]. A textbook treatment of subexponential distributions is given in Embrechts et al. [34].

2.2 Basic Properties

In what follows, we give some of the elementary properties of subexponential distributions. The following is Lemma 1.3.5 of Embrechts et al. [34].

Lemma 2.2.1 (Basic properties of subexponential distributions)

(a) *If $F \in \mathcal{S}$, then uniformly on compact y–sets of $(0,\infty)$,*

$$\lim_{x\to\infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = 1\,. \tag{2.2}$$

(b) *If (2.2) holds then, for all $\varepsilon > 0$,*

$$\mathrm{e}^{\varepsilon x}\overline{F}(x) \to \infty\,, \quad x\to\infty\,.$$

(c) *If $F \in \mathcal{S}$ then, given $\varepsilon > 0$, there exists a finite constant K so that for all $n \geq 2$,*

$$\frac{\overline{F^{n*}}(x)}{\overline{F}(x)} \leq K\,(1+\varepsilon)^n\,, \quad x \geq 0\,. \tag{2.3}$$

Proof. (a) For $x \geq y > 0$, by straightforward calculation,

$$\begin{aligned}\frac{\overline{F^{2*}}(x)}{\overline{F}(x)} &= 1 + \int_0^y \frac{\overline{F}(x-t)}{\overline{F}(x)}\,dF(t) + \int_y^x \frac{\overline{F}(x-t)}{\overline{F}(x)}\,dF(t)\\ &\geq 1 + F(y) + \frac{\overline{F}(x-y)}{\overline{F}(x)}\big(F(x)-F(y)\big)\,.\end{aligned}$$

Thus, for x large enough so that $F(x)-F(y) \neq 0$,

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

In the latter estimate, the right-hand side tends to 1 as $x\to\infty$. The property (2.2) is equivalent to saying that $\overline{F}\circ\ln$ is slowly varying so that uniform convergence follows from the uniform convergence theorem for regularly varying functions; see Theorem 1.2.4.

(b) By (a), $\overline{F}\circ\ln$ is slowly varying. But then the conclusion that $x^{\varepsilon}\overline{F}(\ln x)\to\infty$ as $x\to\infty$ follows immediately from the representation theorem for slowly varying functions; see Theorem 1.2.1.

(c) Let $\alpha_n = \sup_{x\geq 0}\overline{F^{n*}}(x)/\overline{F}(x)$. Using the relation

$$\begin{aligned}\frac{\overline{F^{(n+1)*}}(x)}{\overline{F}(x)} &= 1 + \frac{F(x)-F^{(n+1)*}(x)}{\overline{F}(x)}\\ &= 1 + \int_0^x \frac{\overline{F^{n*}}(x-t)}{\overline{F}(x)}\,dF(t)\\ &= 1 + \left(\int_0^{x-y} + \int_{x-y}^x\right)\left(\frac{\overline{F^{n*}}(x-t)}{\overline{F}(x-t)}\frac{\overline{F}(x-t)}{\overline{F}(x)}\right)\,dF(t)\,,\end{aligned}$$

we obtain, for every $T < \infty$,

$$\begin{aligned}\alpha_{n+1} &\leq 1 + \sup_{0\leq x\leq T}\int_0^x \frac{\overline{F^{n*}}(x-y)}{\overline{F}(x)}\,dF(y)\\ &\quad + \sup_{x\geq T}\int_0^x \frac{\overline{F^{n*}}(x-y)}{\overline{F}(x-y)}\frac{\overline{F}(x-y)}{\overline{F}(x)}\,dF(y)\\ &\leq 1 + A_T + \alpha_n \sup_{x\geq T}\frac{F(x)-F^{2*}(x)}{\overline{F}(x)},\end{aligned}$$

where $A_T = (\overline{F}(T))^{-1} < \infty$. Now since $F \in \mathcal{S}$ we can, given any $\varepsilon > 0$, choose T such that

$$\alpha_{n+1} \leq 1 + A_T + \alpha_n(1+\varepsilon)\,.$$

Hence

$$\alpha_n \leq (1 + A_T)\,\varepsilon^{-1}\,(1+\varepsilon)^n\,,$$

implying (2.3). □

Remark 2.2.2 Lemma 2.2.1(b) justifies the name subexponential for $F \in \mathcal{S}$; indeed $\overline{F}(x)$ decays to 0 slower than any exponential $\mathrm{e}^{-\varepsilon x}$ for $\varepsilon > 0$. Furthermore, since for any $\varepsilon > 0$:

$$\int_y^\infty \mathrm{e}^{\varepsilon x}\,dF(x) \geq \mathrm{e}^{\varepsilon y}\,\overline{F}(y)\,, \quad y \geq 0\,,$$

it follows from Lemma 2.2.1(b) that for $F \in \mathcal{S}$, the moment generating function of F does not exist in any neighbourhood of zero.. Therefore the Laplace–Stieltjes transform of a subexponential distribution function has an essential singularity at 0. This result was first proved by Chistyakov [15], Theorem 2. As follows from the proof of Lemma 2.2.1(b) the latter property holds true for the larger class of distribution functions satisfying (2.2).

Remark 2.2.3 Condition (2.2) can be taken as another definition for a "heavy–tailed distribution". Notice that for some random variable X with distribution F relation (2.2) can be rewritten as

$$\lim_{x\to\infty} P\,(X > x+y \mid X > x) = \lim_{x\to\infty}\frac{\overline{F}(x+y)}{\overline{F}(x)} = 1\,, \quad y > 0\,.$$

Intuitively this means that if X ever exceeds a large value then it is likely to exceed any larger value as well. Notice that the "light–tailed" exponential distribution with $\overline{F}(x) = \exp\{-\lambda x\}$, $x > 0$, for some $\lambda > 0$, satisfies

$$\lim_{x\to\infty}\frac{\overline{F}(x+y)}{\overline{F}(x)} = \mathrm{e}^{-\lambda y}\,, \quad y > 0\,.$$

2.3 Examples

We have already established that the class of distributions with regularly varying right tail is a subset of $\mathcal{S}$; see Corollary 1.3.6, and that the exponential distribution does not belong to $\mathcal{S}$; see Remark 2.2.3.

What can be said about classes "in between", such as for example the important class of Weibull–type variables where $\overline{F}(x) \sim \exp\{-x^\tau\}$ with $0 < \tau < 1$?

Let F be absolutely continuous with density f. Recall the notions of the *hazard rate* $q = f/\overline{F}$ and *hazard function*

$$Q(x) = \int_0^x q(y)\, dy = -\ln \overline{F}(x)\ .$$

An interesting result yielding a complete answer to $\mathcal{S}$–membership for absolutely continuous F with hazard rate eventually decreasing to 0 is given in Pitman [97]; see Proposition A3.16 in Embrechts et al. [34].

Proposition 2.3.1 (A characterisation theorem for $\mathcal{S}$)
Suppose F is absolutely continuous with density f and hazard rate $q(x)$ eventually decreasing to 0. *Then*

(*a*) $F \in \mathcal{S}$ *if and only if*

$$\lim_{x\to\infty} \int_0^x e^{y\, q(x)} f(y)\, dy = 1\ . \tag{2.4}$$

(*b*) *If the function* $x \mapsto \exp\{x\, q(x)\}\, f(x)$ *is integrable on* $[0,\infty)$ *then* $F \in \mathcal{S}$.

Proof. (a) We restrict ourselves to the sufficiency part; a complete proof can be found in the references mentioned above. Suppose that the condition (2.4) holds. After splitting the integral below over $[0,x]$ into two integrals over $[0,x/2]$ and $(x/2,x]$ and making a substitution in the second integral, we obtain

$$\begin{aligned}
&\int_0^x e^{Q(x)-Q(x-y)-Q(y)}\, q(y)\, dy \\
&= \int_0^{x/2} e^{Q(x)-Q(x-y)-Q(y)}\, q(y)\, dy \\
&\quad + \int_0^{x/2} e^{Q(x)-Q(x-y)-Q(y)}\, q(x-y)\, dy \\
&= I_1(x) + I_2(x)\ .
\end{aligned}$$

It follows by a monotonicity argument that $I_1(x) \ge F(x/2)$. Moreover, for $y \le x/2$ and therefore $x - y \ge x/2$,

$$Q(x) - Q(x-y) \le y\, q(x-y) \le y\, q(x/2)\ .$$

Therefore

$$F(x/2) \le I_1(x) \le \int_0^{x/2} e^{y\, q(x/2) - Q(y)}\, q(y)\, dy\ ,$$

and (2.4) implies that

$$\lim_{x\to\infty} I_1(x) = 1\ . \tag{2.5}$$

The integrand in $I_1(x)$ converges pointwise to $f(y) = \exp\{-Q(y)\}q(y)$. Thus we can reformulate (2.5) as "the integrand of $I_1(x)$ converges in f–mean to 1". The integrand in $I_2(x)$ converges pointwise to 0, it is however everywhere bounded by the integrand of $I_1(x)$. From this and an application of Pratt's lemma (see Pratt [98]), it follows that $\lim_{x\to\infty} I_2(x) = 0$. Consequently,

$$\lim_{x\to\infty} \frac{\overline{F^{2*}}(x)}{\overline{F}(x)} - 1 = 1\ ,$$

i.e. $F \in \mathcal{S}$, proving sufficiency of (2.4) and hence assertion (a).

(b) The assertion follows immediately from Lebesgue's dominated convergence theorem, since $q(x) \le q(y)$ for $y \le x$. $\square$

Proposition 2.3.1 immediately yields the following examples in $\mathcal{S}$. Note that, using the above notation, $\overline{F}(x) = \exp\{-Q(x)\}$.

Example 2.3.2 (Examples of subexponential distributions)

(a) Take F a Weibull distribution with parameters $0 < \tau < 1$ and $c > 0$, i.e.

$$\overline{F}(x) = e^{-c\,x^\tau}\,, \quad x \geq 0\,.$$

Then $f(x) = c\tau x^{\tau-1} e^{-cx^\tau}$, $Q(x) = cx^\tau$ and $q(x) = c\tau x^{\tau-1}$ which decreases to 0 if $\tau < 1$. We can immediately apply Proposition 2.3.1(b) since

$$x \mapsto e^{x\,q(x)} f(x) = e^{c\,(\tau-1)x^\tau} c\,\tau\,x^{\tau-1}$$

is integrable on $(0, \infty)$ for $0 < \tau < 1$. Therefore $F \in \mathcal{S}$.

(b) Using Proposition 2.3.1, one can also prove for

$$\overline{F}(x) \sim e^{-x(\ln x)^{-\beta}}\,, \quad x \to \infty\,, \quad \beta > 0\,,$$

that $F \in \mathcal{S}$. This example shows that one can come fairly close to exponential tail behaviour while staying in $\mathcal{S}$.

(c) At this point one could hope that for

$$\overline{F}(x) \sim e^{-x^\tau L(x)}\,, \quad x \to \infty\,, \quad 0 \leq \tau < 1\,, \quad L \text{ slowly varying}\,,$$

F would belong to $\mathcal{S}$. Again, in this generality the answer to this question is *no*. One can construct examples of slowly varying L so that the corresponding F does not even satisfy condition (2.2). An example for $\tau = 0$ was communicated by Charles Goldie; see also Cline [18], where counterexamples for $0 < \tau < 1$ are given.

2.4 Further Properties

In this section we collect some further properties of subexponential distributions; their proofs can be found in Embrechts et al. [34].

The following amazing result is due to Goldie; see Embrechts, Goldie and Veraverbeke [32] or Embrechts et al. [34], Proposition A3.18.

Proposition 2.4.1 (Convolution root closure of $\mathcal{S}$)
If $F^{n*} \in \mathcal{S}$ for some positive integer n, then $F \in \mathcal{S}$.

Quite often it is of interest to consider transforms of subexponential distributions. One of them is a compound Poisson distribution of a distribution in $\mathcal{S}$. Assume that for some $\lambda > 0$,

$$G(x) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} F^{k*}(x)\,, \quad x \geq 0\,, \tag{2.6}$$

which is the distribution of $\sum_{i=1}^N X_i$, where the X_i are iid with distribution F, independent of the $Poi(\lambda)$ random variable N.

The proof of the following result can be found in [34], Theorem A3.19.

Theorem 2.4.2 (Subexponentiality and compound Poisson distributions)
Let G, F and λ be related by (2.6). *Then the following assertions are equivalent:*

(a) $G \in \mathcal{S}$,

(b) $F \in \mathcal{S}$,

(*c*) $\lim_{x\to\infty} \overline{G}(x)/\overline{F}(x) = \lambda$.

The same proof appplies to more general compound distributions as demonstrated in Theorem A3.20 of [34].

The compound Poisson model (2.6) leads to one of the key examples of the so–called *infinitely divisible distributions*. A distribution F is infinitely divisible if for all $n \in \mathbb{N}$, there exists a distribution H_n such that $F = H_n^{n*}$. The Laplace–Stieltjes transform of an infinitely divisible distribution F on $[0, \infty)$ can be expressed (see Feller [40], p. 450) as

$$\widehat{f}(s) = \exp\left\{-as - \int_0^\infty \left(1 - e^{-sx}\right) \, d\nu(x)\right\}, \qquad s \geq 0,$$

where $a \geq 0$ is constant and ν is a Borel measure on $(0, \infty)$ for which $\nu(1, \infty) < \infty$ and $\int_0^1 x \, d\nu(x) < \infty$. This is the so–called Lévy–Khinchin representation theorem. In Embrechts et al. [32] the following result is proved.

Theorem 2.4.3 (Infinite divisibility and $\mathcal{S}$)
For F infinitely divisible on $(0, \infty)$ with Lévy–Khinchin measure ν, the following are equivalent:

(*a*) $F \in \mathcal{S}$,

(*b*) $\nu(1, x]/\nu(1, \infty) \in \mathcal{S}$,

(*c*) $\lim_{x\to\infty} \overline{F}(x)/\nu(x, \infty) = 1$. □

A key step in the proof of the above result concerns the following question.

If F, $G \in \mathcal{S}$, does it always follow that the convolution $F * G \in \mathcal{S}$?

The rather surprising answer to this question is

In general, NO!

The latter was shown by Leslie [68]. Necessary and sufficient conditions for convolution closure are given in Embrechts and Goldie [30]. The main result needed in the proof of Theorem 2.4.3 is the first part of the following lemma (Embrechts et al. [32], Proposition 1). A proof of the second part can be found in Embrechts and Goldie [31].

Lemma 2.4.4 (Convolution in $\mathcal{S}$)

(*a*) *Let $H = F * G$ be the convolution of two distributions on $(0, \infty)$. If $G \in \mathcal{S}$ and $\overline{F}(x) = o(\overline{G}(x))$ as $x \to \infty$, then $H \in \mathcal{S}$.*

(*b*) *If $F \in \mathcal{S}$ and $\overline{G}_i(x) \sim c_i \overline{F}(x)$ for $c_i > 0$, then $\overline{G_1 * G_2}(x) \sim (c_1 + c_2)\, \overline{F}(x)$ as $x \to \infty$.*

(*c*) *If $F \in \mathcal{S}$ and $\overline{G}(x) \sim c\overline{F}(x)$ for some positive c, then $\overline{F * G}(x) \sim (1 + c)\overline{F}(x)$ as $x \to \infty$.*

(*d*) *If $F, G \in \mathcal{S}$, then $F * G \in \mathcal{S}$ if and only if $[p\, F + (1 - p)\, G] \in \mathcal{S}$ for some (all) $p \in (0, 1)$.*

With respect to asymptotic properties of convolution tails, the papers by Cline [18, 19] offer a useful source of information.

In the discrete case, the following classes of distributions have been found to yield various interesting results.

Definition 2.4.5 (Discrete subexponentiality)
Suppose (p_n) *defines a probability measure on* $\mathbb{N}_0$, *let* $p_n^{2*} = \sum_{k=0}^n p_{n-k}p_k$ *be the two–fold convolution of* (p_n). *Then* $(p_n) \in \mathrm{SD}$, *if*

(a) $\lim_{n\to\infty} p_n^{2*}/p_n = 2 < \infty$,

(b) $\lim_{n\to\infty} p_{n-1}/p_n = 1$.

The class SD *is the class of* discrete subexponential sequences. □

Suppose (p_n) is infinitely divisible and $\widehat{p}(r) = \sum_{k=0}^\infty p_k r^k$ its generating function. Then by the Lévy–Khinchin representation theorem

$$\widehat{p}(z) = \exp\left\{-\lambda\left(1-\sum_{j=1}^\infty \alpha_j\, z^j\right)\right\},$$

for some $\lambda > 0$ and a probability measure (α_j). The following result is the discrete analogue to Theorem 2.4.3, or for that matter Theorem 2.4.2 (Embrechts and Hawkes [33]).

Theorem 2.4.6 (Discrete infinite divisibility and subexponentiality)
The following three statements are equivalent:

(a) $(p_n) \in \mathrm{SD}$,

(b) $(\alpha_n) \in \mathrm{SD}$,

(c) $\lim_{n\to\infty} p_n/\alpha_n = \lambda$ *and* $\lim_{n\to\infty} \alpha_n/\alpha_{n+1} = 1$. □

See the above paper for applications and further references.

A recent survey paper on subexponentiality is Goldie and Klüppelberg [46].

2.5 Applications

2.5.1 Ruin Probabilities

One of the classical fields of applications for subexponential distributions is insurance mathematics. Such distributions are used as realistic models for describing the sizes of real–life insurance claims which can have distributions with very heavy tails. The class $\mathcal{S}$ of subexponential distributions is quite flexible for modelling a large variety of tails, including regularly varying, lognormal and heavy–tailed Weibull tails.

In what follows, we intend to consider one of the classical insurance models and want to see how the subexpontential distributions come into consideration in a very natural way. The *classical insurance risk process* is defined as

$$R(t) = u + ct - \sum_{i=1}^{N_t} X_i\,, \quad t \geq 0\,,$$

where

- $u \geq 0$ is the *initial capital* or risk reserve. It is usually assumed that u is very large.
- $c > 0$ is the *premium rate*. This means that there is a linear premium income as a function of time.

- $(N_t)_{t\geq 0}$ is the *number of claims that occurred until time t.* Usually, N is modelled as a homogeneous Poisson process or as a renewal counting process, i.e.
$$N_t = \#\{i : T_n = Y_1 + \cdots + Y_n \leq t\}\,, \quad t \geq 0\,,$$
where (Y_i) is an iid sequence of non-negative random variables. We will assume that $EY = \lambda^{-1}$ exists and is finite. Then λ is called the *claim time intensity.*

- $X_1, X_2, \ldots$ are the *claim sizes.* They are supposed to be iid with common *claim size distribution* F and usually also non–negative. Notice that the nth claim occurs at time T_n. Also write $\mu = EX$, where we assume that the latter expectation is finite.

- *Claim sizes and claim times are independent.* This means that (X_i) and (Y_i) are independent.

Now all ingredients of the risk process are defined. The event

$$\{R(t) < 0 \quad \text{for some } t\} \quad \text{for a given initial capital } u$$

is referred to as the *ruin*, and the corresponding probability

$$\psi(u) := P\left(R(t) < 0 \quad \text{for some } t\right)\,, \quad u \geq 0\,,$$

is said to be the *ruin probability.* In insurance mathematics rhe ruin probability, as a function of u, is considered as an important measure of risk. Therefore quite some work in probability theory has been done in order to evaluate the size of the quantity $\psi(u)$. Early on, starting with Cramér [21], it was realised that the ruin probability is closely related to the distribution of a random walk with negative drift. Indeed, in order to avoid ruin with probability 1, we have to assume that for large t,

$$ct > S(t) := \sum_{i=1}^{N_t} X_i \quad \text{a.s.}$$

An appeal to the strong law of large numbers ensures that this condition holds, provided

$$ct > E\left[\sum_{i=1}^{N_t} X_i\right] = EN_t\,\mu \sim \lambda\, t\, \mu\,.$$

This amounts to the following *net profit condition*:

$$\rho := \frac{c}{\lambda\mu} - 1 > 0\,. \tag{2.7}$$

In what follows, we assume this condition to be satisfied.

For the ease of representation assume from now on that $(N_t)_{t\geq 0}$ is a homogeneous Poisson process with intensity λ. Alternatively, the Y_i's consistitute a sequence of iid $Exp(\lambda)$ random variables. In the insurance context, the risk process is then called *Cramér–Lundberg model.* Since ruin can occur *only at the claim times T_n* we have

$$\begin{aligned}\psi(u) &= P\left(\inf_{t\geq 0}\,[u + ct - S(t)]\; < 0\right)\\ &= P\left(\inf_{n\geq 0}\left[u + cT_n - \sum_{i=1}^{n} X_i\right]\; < 0\right)\end{aligned}$$

$$= P\left(\inf_{n\geq 0} \sum_{i=1}^{n}[cY_i - X_i] < -u\right)$$

$$= P\left(\sup_{n\geq 0} \sum_{i=1}^{n}[X_i - cY_i] > u\right).$$

By virtue of the net profit condition (2.7), $E[X - cY] < 0$. Therefore and by the independence of (X_n) and (Y_n), $\psi(u)$ is nothing but the right tail of the distribution of the supremum of a random walk with negative drift. This probability can be determined for instance via Spitzer's identity (cf. Feller [40], p. 613). An application of the latter result allows one to express the non–ruin probability $1 - \psi(u)$ as a compound geometric distribution, i.e.

$$1 - \psi(u) := \frac{\rho}{1+\rho} \sum_{n=0}^{\infty}(1+\rho)^{-n} F_I^{n*}(u), \tag{2.8}$$

where ρ is defined in (2.7) and

$$F_I(x) := \frac{1}{\mu}\int_0^x \overline{F}(y)\, dy, \quad x \geq 0,$$

denotes the *integrated tail distribution.*

In his classical work, Cramér [21] gave asymptotic estimates for $\psi(u)$ as $u \to \infty$. In particular, he obtained the bound

$$\psi(u) \leq \mathrm{e}^{-\nu u}, \quad u \geq 0,$$

which holds under the assumption that the equation (*Cramér–Lundberg condition*)

$$\int_0^\infty \mathrm{e}^{\nu x}\, \overline{F}(x) = \frac{c}{\lambda}$$

has a solution. If the latter exists it is unique.

The latter equation only makes sense if $\overline{F}(x)$ decays to zero at an exponential rate. However, we learnt in Section 2.1 that subexponential distributions do not have finite exponential moments and therefore the classical Cramér theory does not apply. In this context the representation of $\psi(u)$ via the infinite series (2.8) is very useful. Indeed, consider

$$\frac{\psi(u)}{\overline{F}_I(u)} = \frac{\rho}{1+\rho} \sum_{n=0}^{\infty}(1+\rho)^{-n} \frac{\overline{F_I^{n*}}(u)}{\overline{F}_I(u)}. \tag{2.9}$$

If F_I is subexponential, we know that

$$\overline{F_I^{n*}}(u) \sim n\, \overline{F}_I(u) \text{ for every fixed } n \text{ as } u \to \infty. \tag{2.10}$$

This fact together with the basic property (2.3), i.e.

$$\frac{\overline{F^{n*}}(x)}{\overline{F}(x)} \leq K\, (1+\varepsilon)^n, \quad x \geq 0,$$

and a dominated convergence argument ensure that we may interchange the limit as $u \to \infty$ and the infinite series in (2.9) to obtain

$$\frac{\psi(u)}{\overline{F}_I(u)} \sim \frac{\rho}{1+\rho} \sum_{n=0}^{\infty} n\, (1+\rho)^{-n} = \rho^{-1}. \tag{2.11}$$

This elegant solution to the ruin problem for subexponential distributions was proposed by Embrechts and Veraverbeke [35] generalising earlier work by Embrechts et al. [32] and by Pakes [92], Smith [106] and Veraverbeke [110] in the queuing context.

Remark 2.5.1 Consider a GI/G/1 queue with renewal arrival stream and general service time distribution F with finite mean μ and corresponding integrated tail distribution F_I. Also assume that the queue is stable in the sense that the traffic intensity is smaller than 1. (This is the net profit condition (2.7).) Then the stationary waiting time distribution can be represented as the distribution of the supremum of a random walk with negative drift; cf. Feller [40], VI.9. Hence analogues of relation (2.11) (corresponding to an M/G/1 queue) are immediate. Another interpretation of (2.11) has been given in the context of branching theory; see Athreya and Ney [3], Chistyakov [15] and Chover et al. [16].

Remark 2.5.2 The asymptotic relation (2.11) is nothing but a qualitative result. Assuming second or even higher order subexponentiality, i.e. requiring conditions on the rate of convergence in (2.10), one can obtain an improvement on the rate of convergence in the asymptotic relation (2.11). This, however, would be of a rather restricted practical interest because we do not know the exact tail behaviour of a distribution in real–life situations. Mikosch and Nagaev [78] show that even for the simple case of a distribution F_I with $\overline{F}_I(x) \sim \mathit{const}\ x^{-\alpha}$, the error term $\psi(u)/\overline{F}_I(u) - \rho^{-1}$ can decay to zero arbitrarily slow.

In the context of the approximation (2.11) it is an important question as to whether F_I is subexponentai or not. In particular, one would like to have simple conditions on the distribution F which ensure that F_I is subexponential.

Example 2.5.3 (Regularly varying distribution)
It is clear from Karamata's Theorem 1.2.6 that regular variation of $\overline{F}(x)$ with index $-\alpha < -1$ implies that

$$\overline{F}_I(u) \sim \frac{u\,\overline{F}(u)}{\mu(\alpha-1)}\,, \quad x \to \infty\,.$$

Hence $F_I(x)$ is subexponential. □

In general, it is difficult to determine whether $F_I \in \mathcal{S}$. Klüppelberg [62] introduced the class $\mathcal{S}^*$ in order to handle this problem.

Definition 2.5.4 (The class $\mathcal{S}^*$)
For any distribution F with support $(0,\infty)$ and finite mean μ, $F \in \mathcal{S}^*$ if the following relation holds:

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

The class $\mathcal{S}^*$ makes it less difficult to decide whether F_I is subexponential:

Proposition 2.5.5 *If $F \in \mathcal{S}^*$, then $F \in \mathcal{S}$ and $F_I \in \mathcal{S}$.*

Conditions for $F \in \mathcal{S}$ and F_I are often formulated in terms of *hazard functions* $Q(x) = -\ln \overline{F}(x)$ or its density, the *hazard rates* $q(x)$. (It can be shown that every $F \in \mathcal{S}^*$ is asymptotically tail equivalent to an absolutely continuous distribution with hazard rate $q(x) \to 0$ as $x \to \infty$.) The following results are cited from Goldie and Klüppelberg [46].

Proposition 2.5.6 (Conditions for $F \in \mathcal{S}^*$)

(a) *If* $\lim_{x\to\infty} x\, q(x) < \infty$ *then* $F \in \mathcal{S}^*$.

(b) *If there exist* $\delta \in (0,1)$ *and* $v \geq 1$ *such that* $Q(xy) \leq y^\delta Q(x)$ *for all* $x \geq v$, $y \geq 1$ *and* $\liminf_{x\to\infty} x\, q(x) \geq (2-2^\delta)^{-1}$, *then* $F \in \mathcal{S}^*$.

(c) *If* q *is eventually decreasing to* 0, *then* $F \in \mathcal{S}^*$ *if and only if*

$$\lim_{x\to\infty} \int_0^x \mathrm{e}^{y\, q(x)} \overline{F}(y)\, dy = \mu\,.$$

(d) *If*

$$\lim_{x\to\infty} q(x) = 0 \quad \textit{and} \quad \limsup_{x\to\infty} x\, q(x) = \infty\,,$$

and $\limsup_{x\to\infty} xq(x)/Q(x) < 1$, *then* $F \in \mathcal{S}^*$.

Other conditions for $F \in \mathcal{S}^*$ are given in Klüppelberg [62, 63], Cline [18] and Goldie and Klüppelberg [46]; see also Embrechts et al. [34], Section 1.4.

Example 2.5.7 (Distributions $F_I \in \mathcal{S}$)
The heavy–tailed Weibull distribution F with parameter $\tau \in (0,1)$ has $F_I \in \mathcal{S}$. So have the log–normal distribution and the Benktander–type I and II distributions.

We finally mention that, in general, $F \in \mathcal{S}$ does not imply $F_I \in \mathcal{S}$ and vice versa.

Notes and Comments

Over the last few years various results on the asymptotic behaviour for ruin probabilities have been derived for various settings. The case of claim sizes with a distribution having a moment generating function in some neighbourhood of zero has attracted most attention, also for dependent claim size sequences. Grandell [47] gives an overview of techniques and results in this context. The heavy–tailed dependent case has been treated recently in various papers. Asmussen et al. [2] show that the Embrechts and Veraverbeke result (2.11) remains valid (in the queuing context) under a fairly general dependence structure of the interarrival times if the service times are still independent. Mikosch and Samorodnitsky [79] consider ruin probabilities for claims sizes which constitute a linear process with regularly varying tail distribution. They show that the Embrechts and Veraverbeke result does not remain valid in this case, although the asymptotic order of $\psi(u)$ (regular variation with index $(1-\alpha)$) remains the same. Braverman et al. [12] consider the case of a general α–stable strictly stationary ergodic sequence with $\alpha \in (1,2)$. They show that the ruin probability can be of the order $\psi(u) \sim u^{\gamma(1-\alpha)}L(u)$, where L is any slowly varying function and $\gamma \in (0,1)$. The value γ is the smaller the stronger the dependence in the sequence.

2.6 Large Deviations

This section is based on Mikosch and Nagaev [78].

Large deviation probabilities for heavy–tailed random variables can be considered as a natural extension of the relation

$$P(S_n > x) \sim P(M_n > x) \sim n\, \overline{F}(x) \tag{2.12}$$

as $x \to \infty$, where, as usual,

$$S_n = X_1 + \cdots + X_n\,, \quad M_n = \max(X_1, \ldots, X_n)\,, \quad n \geq 1\,,$$

for an iid sequence (X_n) with common distribution F. We know that (2.12) is a property of regularly varying distributions and can be taken as the definition for the class of subexponential distributions. A large deviation probability occurs when x and n in (2.12) are linked:

$$P(S_n > x_n) \sim P(M_n > x_n) \sim n\,\overline{F}(x_n)\,, \quad x_n \to \infty\,, \quad n \to \infty\,.$$

More generally, one is interested in x-regions B_n, say, where the latter relation holds uniformly in x:

$$\sup_{x \in B_n} \left| \frac{P(S_n > x)}{n\,\overline{F}(x)} - 1 \right| \to 0\,, \quad n \to \infty\,. \tag{2.13}$$

One usually chooses $B_n = [d_n, \infty)$ for some threshold sequence $d_n \to \infty$. Relation (2.13) again supports the idea that the event $\{S_n > x\}$ for large x is due to only one single event, namely to the extremal event $\{M_n > x\}$.

Clearly, if x is not large enough we may expect that, instead of an approximation by $P(M_n > x)$, the central limit theorem might be a good approximation to $P(S_n > x)$: write

$$\overline{\Phi}(x) := \frac{1}{\sqrt{2\pi}} \int_x^\infty \mathrm{e}^{-y^2/2}\,dy\,,$$

for the tail of the standard normal distribution. Assume that X has mean $\mu = 0$ and finite variance σ^2. Then one is interested in x–regions $(1, c_n]$ where the following normal approximation holds. such that

$$\sup_{1 \le x \le c_n} \left| \frac{P(S_n > x)}{\overline{\Phi}(x)/(\sigma\sqrt{n}))} - 1 \right| \to 0\,. \tag{2.14}$$

From the CLT we may conclude that there exists a sequence (d_n) with $d_n/\sqrt{n} \to \infty$ sufficiently slowly and such that (2.14), with $c_n = c\sqrt{n}$, holds. In his famous work, Cramér [22] proved that, given the existence of the moment generating function of X in a neighbourhood of the origin, (2.14) holds with $c_n = c\sqrt{n}$, for any $c > 0$ and $d_n = o(n^{2/3})$, while (2.14) fails in general for $d_n = O(n^{2/3})$; see for instance Petrov [95, 96]. The existence of the moment generating function is also referred to as *Cramér's condition*. It is crucial for the proof of Cramér's theorem. Petrov's refined version of Cramér's theorem (see Petrov [93], and also Petrov [96], Theorem 5.23) yields the following: if $x \ge 0$, $x = o(n)$ and Cramér's condition holds, then

$$\frac{P(S_n > x)}{\overline{\Phi}(x/(\sigma\sqrt{n}))} = \exp\left\{\frac{x^3}{n^2}\lambda\left(\frac{x}{n}\right)\right\}\left(1 + O\left(\frac{x+1}{n}\right)\right)\,. \tag{2.15}$$

Here $\lambda(z) = \sum_{k=0}^\infty c_k z^k$ is a power series with coefficients depending on the cumulants of the random variable X. The series $\lambda(z)$ converges for sufficiently small values of z. It is called *Cramér's series* and appears in many results related to large deviations.

Cramér's theorem has been extended and generalised in various directions, for example to sums of dependent random variables or to general sequences of random variables. Large deviation techniques are very useful tools in different areas of probability theory, statistics, statistical physics, insurance mathematics, renewal theory, and other applied fields. The main stream of research has been concentrated on the study of the quantities $\ln P(S_n > x)$ under Cramér's condition. We call them *rough large deviation probabilities*; see for instance the monographs by Bucklew [14], Dembo and Zeitouni [27], Deuschel and Stroock [28], and Ellis [29]. However, in many situations rough large deviation results do not suffice. Therefore *precise large* *deviation results* for $P(S_n > x)$ are needed; Cramér's asymptotic expansion (2.15) is a typical example.

Since the existence of the moment generating function is crucial for proving Cramér's theorem, the following question also arises: *what can be said about the asymptotic behaviour of $P(S_n > x)$ if Cramér's condition is violated?* This situation is typical for many distributions of interest in insurance, when one is interested in modelling large claims, or in queuing, when large interarrival times are described by distributions with heavy tails. In that case, distributions with exponentially decaying tails (which is a consequence of Cramér's condition) do not form realistic models. Typical distributions with heavy tails are the following:

Regularly varying tails RV(α)

$$\overline{F}(x) = x^{-\alpha} L(x) \,, \quad x > 0 \,,$$

where $\alpha > 0$, and L is a slowly varying function.

Lognormal–type tails LN(γ)

$$\overline{F}(x) \sim c x^{\beta} e^{-\lambda \ln^{\gamma} x} \,, \quad x \to \infty \,,$$

for some $\beta \in \mathbb{R}$, $\gamma > 1$, $\lambda > 0$ and appropriate $c = c(\beta, \gamma)$. In the abbreviation LN(γ) we suppress the dependence on β and λ.

Weibull–like tails WE(α)

$$\overline{F}(x) \sim c x^{\beta} e^{-\lambda x^{\alpha}} \,, \quad x \to \infty \,,$$

for some $\beta \in \mathbb{R}$, $\alpha \in (0,1)$, $\lambda > 0$ and appropriate $c = c(\beta, \alpha)$. In the abbreviation WE(α) we suppress the dependence on β and λ.

The lognormal distribution obviously belongs to LN(2). The heavy–tailed Weibull and the Benktander–type–II distributions are members of WE(α). All these distributions do not satisfy Cramér's condition, hence Cramér's theorem is not applicable. For a precise definition of these distributions we refer for instance to Embrechts et al. [34], Chapter 1.

Notation

Before we formulate the result about the large deviation probabilities for heavy–tailed distributions we introduce some notation. In what follows, $c_n \le d_n$ are two sequences of positive numbers. Write

$$x \ll c_n \quad \text{if} \quad x \in (0, c_n/h_n) \,,$$
$$x \gg d_n \quad \text{if} \quad x \in (d_n g_n, \infty) \,,$$

for *any* choice of sequences h_n, $g_n \to \infty$ as $n \to \infty$. The relation $c_n \ll x \ll d_n$ is defined analogously. The asymptotic relation

$$A_n(x) \sim B_n(x) \tag{2.16}$$

for $c_n \ll x \ll d_n$ means that

$$\sup_{c_n h_n \le x \le d_n / g_n} \left| \frac{A_n(x)}{B_n(x)} - 1 \right| = o(1)$$

for any choice of sequences h_n, $g_n \to \infty$ as $n \to \infty$. For $x \ll c_n$ and $x \gg d_n$, (2.16) is defined analogously. We also write $a_n \sim b_n$ for $a_n = b_n(1 + o(1))$.

A Review of Large Deviation Results for Heavy–Tailed Distributions

In this section we describe the asymptotic behaviour of large deviation probabilities $P(S_n > x)$ under a heavy–tail condition on F. If not mentioned otherwise, we assume that $\mu = EX = 0$, $\sigma^2 = \text{var}(X) = 1$ and $E|X|^{2+\delta} < \infty$ for some $\delta > 0$.

It is typical for large deviation probabilities $P(S_n > x)$ that there exist two threshold sequences $c_n \le d_n$ such that

$$\overline{P}(S_n > x) \sim \overline{\Phi}(x/\sqrt{n}) \,, \quad x \ll c_n \,, \tag{2.17}$$

and

$$\overline{P}(S_n > x) \sim n\overline{F}(x) \,, \quad x \gg d_n \,. \tag{2.18}$$

Proposition 2.6.1 *The following threshold sequences $(c_n), (d_n)$ can be chosen*

$F \in$	c_n	d_n
$\text{RV}(\alpha)\,, \alpha > 2$	$n^{1/2}\ln^{1/2} n$	$n^{1/2}\ln^{1/2} n$
$\text{LN}(\gamma)\,, 1 < \gamma \le 2$	$n^{1/2}\ln^{\gamma/2} n$	$n^{1/2}\ln^{\gamma/2} n$
$\text{LN}(\gamma)\,, \gamma > 2$	$n^{1/2}\ln^{\gamma/2} n$	$n^{1/2}\ln^{\gamma-1} n$
$\text{WE}(\alpha)\,, 0 < \alpha \le 0.5$	$n^{1/(2-\alpha)}$	$n^{1/(2-2\alpha)}$
$\text{WE}(\alpha)\,, 0.5 < \alpha < 1$	$n^{2/3}$	$n^{1/(2-2\alpha)}$

Remark 2.6.2 The sequences given above are, by definition, determined only up to rough asymptotic equivalence. For instance, c_n or d_n can be multiplied by any positive constant, and (2.17) and (2.18) remain valid.

Remark 2.6.3 For the heaviest tails $(\text{RV}(\alpha), \text{LN}(\gamma), 1 < \gamma \le 2)$ one can choose $c_n = d_n$, i.e. there exists one threshold sequence which separates the approximations (2.17) and (2.18). The case $\text{RV}(\alpha)$, $\alpha > 2$, was treated in A. Nagaev [83], where the case $x \ge n^{1/2}\ln n$ was studied. The separating sequence (c_n) given in the table appeared in A. Nagaev [85]. Rozovski [103] discovered that the parameter $\gamma = 2$ separates the classes $\text{LN}(\gamma)$, $\gamma > 2$, and $\text{LN}(\gamma)$, $1 < \gamma \le 2$, in the sense that, in the case $\gamma > 2$, (c_n) and (d_n) have to be chosen non-identically. The case $\text{WE}(\alpha)$, $0 < \alpha < 1$, was considered in A. Nagaev [84, 87]. Other reviews of this topic can be found in S. Nagaev [89] and Rozovski [103]; see also the monograph by Vinogradov [112].

As a historical remark, we mention that the validity of the normal approximation to $P(S_n > x)$ was already studied by Linnik [69, 70] and S. Nagaev [88]. They determined lower separating sequences (c_n) under the assumption that the tail $\overline{F}(x)$ is dominated by one of the regular subexponential tails used in Proposition 2.6.1.

Remark 2.6.4 The assumption $F \in \text{RV}(\alpha)$, $\alpha < 2$, in combination with a tail–balancing condition, implies that F is in the domain of attraction of an α–stable law; see Section 1.4.1. In particular, $\sigma^2 = \infty$. In the domain of attraction of an α–stable law G_α one can find constants $a_n \in \mathbb{R}$ and $b_n > 0$ such that $b_n^{-1}(S_n - a_n) \xrightarrow{d} G_\alpha$. The sequence (b_n) can be chosen such that $\overline{F}(b_n) \sim n^{-1}$. This implies that $b_n = n^{1/\alpha}L_1(n)$ for a slowly varying function L_1. Then the relation

$$P(S_n > x) \sim n\overline{F}(x) \,, \quad x \gg a_n + n^{1/\alpha}L_1(n) \,,$$

holds. The case $\text{RV}(\alpha)$, $\alpha < 2$, was treated by Heyde [54, 55]; he dealt with two–sided large deviation probabilities $P(|S_n - a_n| \ge x)$. Tkachuk [108] proved the corresponding results in the one–sided case, i.e. for $P(S_n > x)$. A unifying approach for $\text{RV}(\alpha)$, $\alpha > 0$, was considered by Cline and Hsing [20].

Remark 2.6.5 The exponential (WE(1)) and superexponential (WE(α), $\alpha > 1$) cases require tools which are quite different from those used for heavy–tailed distribution functions F. These cases can be partly treated by Cramér's theorem, see (2.15), provided that the left tail of F also decays to zero at an exponential rate. Then the moment generating function exists in a neighbourhood of the origin. For at least exponentially decreasing right tails $\overline{F}(x)$, further results in the spirit of Proposition 2.6.1 can be found in A. Nagaev [86].

Remark 2.6.6 Davis and Hsing [23] prove precise large deviation results for stationary mixing sequences with distribution function $F \in \mathrm{RV}(\alpha)$, $\alpha < 2$. Mikosch and Samorodnitsky [79] prove large deviation results for linear processes under very general conditions on the coefficients of the process.

Remark 2.6.7 Large deviations for random sums of iid subexponential random variables were considered in Klüppelberg and Mikosch [64] and Mikosch and Nagaev [78]. The latter papers treat random sums where the summands are independent of the random index. Mikosch and Stegeman [82] consider a special case of dependence between the summands and the random index.

Example 2.6.8 (A quick estimate of the ruin probability)
Let $X, X_1, X_2, \ldots$ be iid with $EX = 0$, $E|X|^{2+\delta} < \infty$ for some $\delta > 0$ and such that $\overline{F}$ is regularly varying with index $-\alpha < 0$. (Clearly, $\alpha > 2$.) Then we know from Proposition 2.6.1 that

$$P(S_n > x) \sim n\overline{F}(x) \quad \text{for } x \gg (n \ln n)^{1/2}. \tag{2.19}$$

For the random walk with negative drift $(-nc + S_n)$, $c > 0$, we call

$$\psi(u) = P\left(\sup_{n\geq 0}(-nc + S_n) > u\right), \quad u \to \infty,$$

the *ruin probability with intial capital* u. See Section 2.5.1. We are interested in the asymptotic behaviour of $\psi(u)$ as $u \to \infty$. Notice that by (2.19),

$$\psi(u) \geq P(S_{[u]} > u(1+c) \sim const\, u^{1-\alpha}(1+c)^{-\alpha}.$$

On the other hand,

$$\psi(u) \leq \sum_{k=0}^{\infty} P\left(\sup_{2^k\leq n\leq 2^{k+1}} S_n > u + c2^k\right) \leq \sum_{k=0}^{\infty} P\left(\sup_{n\leq 2^{k+1}} S_n > u + c2^k\right).$$

Using the Lévy maximal inequality (for example Petrov [96]), one obtains for any $\epsilon \in (0,1)$ and large u,

$$\begin{aligned} P\left(\sup_{n\leq 2^{k+1}} S_n > u + c2^k\right) &\leq 2P\left(S_{2^{k+1}} > u + c2^k - \left(2^{k+2}\sigma^2\right)^{1/2}\right) \\ &\leq 2P\left(S_{2^{k+1}} > (u + c2^k)(1-\epsilon)\right). \end{aligned}$$

We conclude that

$$\psi(u) \leq const \sum_{k=0}^{\infty}(u + 2^k)^{-\alpha} \sim const\, u^{1-\alpha}, \quad u \to \infty.$$

Thus we showed by some simple means that

$$\psi(u) \asymp u^{1-\alpha} \quad \text{as } u \to \infty.$$

Bibliography

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

[2] Asmussen, S., Schmidli, H. and Schmidt, V. (1997): Tail probabilities for non–standard risk and queueing processes with subexponential jumps. *Advances in Applied Probability*, to appear.

[3] Athreya, K.B. and Ney, P.E. (1972) *Branching Processes.* Springer, Berlin.

[4] Babillot, M., Bougerol, P. and Elie, L. (1997) The random difference equation $X_n = A_n X_{n-1} + B_n$ in the critical case. *Ann. Probab.* **25** 478–493.

[5] Basrak, B., Davis, R.A. and Mikosch, T. (1999) The sample ACF of a simple bilinear process. *Stoch. Proc. Appl.*, to appear.

[6] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987) *Regular Variation.* Cambridge University Press, Cambridge.

[7] Borovkov, A.A. (1976) *Stochastic Processes in Queueing.* Springer, New York.

[8] Borovkov, A.A. (1984) *Asymptotic Methods in Queueing Theory.* Wiley, Chichester.

[9] Bougerol, P. and Picard, N. (1992a) Strict stationarity of generalized autoregressive processes. *Ann. Probab.* **20**, 1714–1730.

[10] Bougerol, P. and Picard, N. (1992) Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics* **52**, 115–127.

[11] Brandt, A. (1986) The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. Appl. Probab.* **18**, 211–220.

[12] Braverman, M., Mikosch, T. and Samorodnitsky, G. (1999) Ruin probability with claims modeled by a stationary ergodic process. In preparation.

[13] Breiman, L. (1965) On some limit theorems similar to the arc–sin law. *Theory Probab. Appl.* **10**, 323–331.

[14] Bucklew, J.A. (1991) *Large Deviation Techniques in Decision, Simulation, and Estimation.* Wiley, New York

[15] Chistyakov, V.P. (1964) A theorem on sums of independent positive random variables and its applications to branching random processes. *Theory Probab. Appl.* **9**, 640–648.

[16] Chover, J., Ney, P. and Wainger, S. (1972) Functions of probability measures. *J. Analyse Math.* **26**, 255–302.

[17] Christoph, G. and Wolf, W. (1992) *Convergence Theorems with a Stable Limit Law.* Akademie–Verlag, Berlin.

[18] Cline, D.B.H. (1986) Convolution tails, product tails and domains of attraction. *Probab. Theory Related Fields* **72**, 529–557.

[19] Cline, D.B.H. (1987) Convolutions of distributions with exponential and subexponential tails. *J. Austral. Math. Soc. Ser. A* **43**, 347–365.

[20] Cline, D.B.H. and Hsing, T. (1991) Large deviation probabilities for sums and maxima of random variables with heavy or subexponential tails. Texas A & M University. Preprint.

[21] Cramér, H. (1930) *On the Mathematical Theory of Risk.* Skandia Jubilee Volume, Stockholm. Reprinted in: Martin–Löf, A. (Ed.) Cramér, H. (1994) *Collected Works.* Springer, Berlin.

[22] Cramér, H. (1938) Sur un nouveau théorème–limite de la théorie des probabilités. *Actualités Sci. Indust.* **736**.

[23] Davis, R.A., and Hsing, T. (1995) Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Probab.* **23**, 879–917.

[24] Davis, R.A. and Mikosch, T. (1998) The sample autocorrelations of heavy–tailed processes with applications to ARCH. *Ann. Statist.* **26**, 2049–2080.

[25] Davis, R.A., Mikosch, T. and Basrak, B. (1999) Limit theory for the sample autocorrelations of solutions to stochastic recurrence equations with applications to GARCH processes. In preparation.

[26] Davis, R.A. and Resnick, S.I. (1996) Limit theory for bilinear processes with heavy tailed noise. *Ann. Appl. Probab.* **6**, 1191–1210.

[27] Dembo, A. and Zeitouni, O. (1993) *Large Deviations Techniques and Applications.* Jones and Bartlett, Boston.

[28] Deuschel, J.–D. and Stroock, D.W. (1989) *Large Deviations.* Academic Press, Duluth.

[29] Ellis, R.S. (1985) *Entropy, Large Deviations and Statistical Mechanics.* Springer, Berlin.

[30] Embrechts, P. and Goldie, C.M. (1980) On closure and factorization theorems for subexponential and related distributions. *J. Austral. Math. Soc. Ser. A* **29**, 243–256.

[31] Embrechts, P. and Goldie, C.M. (1982) On convolution tails. *Stoch. Proc. Appl.* **13**, 263–278.

[32] Embrechts, P., Goldie, C.M. and Veraverbeke, N. (1979) Subexponentiality and infinite divisibility. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **49**, 335–347.

[33] Embrechts, P. and Hawkes, J. (1982) A limit theorem for the tails of discrete infinitely divisible laws with applications to fluctuation theory. *J. Austral. Math. Soc. Ser. A* **32**, 412–422.

[34] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Finance and Insurance.* Springer, Berlin.

[35] Embrechts, P. and Veraverbeke, N. (1982) Estimates for the probability of ruin with special emphasis on the possibility of large claims. *Insurance: Math. Econom.* **1**, 55–72.

[36] Engle, R.F. (ed.) (1995) *ARCH Selected Readings.* Oxford University Press, Oxford (U.K.).

[37] Engle, R.F. and Bollerslev, T. (1986) Modelling the persistence of conditional variances. With comments and a reply by the authors. *Econometric Rev.* **5**, 1–87.

[38] Falk, M., Hüsler, J. and Reiss, R.–D. (1994) *Laws of Small Numbers: Extremes and Rare Events.* Birkhäuser, Basel.

[39] Feller, W. (1968) *An Introduction to Probability Theory and Its Applications I,II,* 3rd edition. Wiley, New York.

[40] Feller, W. (1971) *An Introduction to Probability Theory and Its Applications II.* Wiley, New York.

[41] Ferguson, T.S. (1996) *A Course in Large Sample Theory.* Chapman and Hall, London.

[42] Fisher, R.A. and Tippett, L.H.C. (1928) Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.* **24**, 180–190.

[43] Gnedenko, B.V. (1943) Sur la distribution limité du terme d'une série aléatoire. *Ann. Math.* **44**, 423–453.

[44] Gnedenko, B.V. and Kolmogorov, A.N. (1954) *Limit Theorems for Sums of Independent Random Variables.* Addison–Wesley, Cambridge, Mass.

[45] Goldie, C.M. (1991) Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1**, 126–166.

[46] Goldie, C.M. and Klüppelberg, C. (1998) Subexponential distributions. In: Adler, R., Feldman, R. and Taqqu, M.S. (Eds.) *A Practical Guide to Heavy Tails: Statistical Techniques and Applications.* Birkhäuser, Boston, pp. 435–460.

[47] Grandell, J. (1991) *Aspects of Risk Theory.* Springer, Berlin.

[48] Grey, D. R. (1994) Regular variation in the tail behaviour of solutions of random difference equations. *Ann. Appl. Probab.* **4**, 169–183.

[49] Grincevičius, A.K. (1981) A random difference equation. *Lith. Math. J.* **21**, 302–306.

[50] Guillaume, D.M., Dacarogna, M.M., Davé, R.R., Müller, U.A., Olsen, R.B. and Pictet, O.V. (1994) From the bird's eye to the microscope. A survey of new stylized facts of the intra–daily foreign exchange markets. Olsen & Associates, Research Institute for Applied Economics, Zürich.

[51] Gumbel, E.J. (1958) *Statistics of Extremes.* Columbia University Press, New York.

[52] Haan, L. de (1970) *On Regular Variation and Its Application to Weak Convergence of Sample Extremes.* CWI Tract **32**, Amsterdam.

[53] Haan, L. de (1994) Extreme value statistics. In: Galambos et al. (Eds.) *Extreme Value Theory and Applications*, pp. 93–122.

[54] Heyde, C.C. (1967a) A contribution to the theory of large deviations for sums of independent random variables. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **7**, 303–308.

[55] Heyde, C.C. (1967b) On large deviation problems for sums of random variables not attracted to the normal law. *Ann. Math. Statist.* **38**, 1575–1578.

[56] Hill, B.M. (1975) A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163–1174.

[57] Hogg, R.V. and Klugman, S.A. (1984) *Loss Distributions.* Wiley, New York.

[58] Ibragimov, I.A. and Linnik, Yu.V. (1971) *Independent and Stationary Sequences of Random Variables.* Wolters–Noordhoff, Groningen.

[59] Janicki, A. and Weron, A. (1993) *Simulation and Chaotic Behaviour of α–Stable Stochastic Processes.* Marcel Dekker, New York.

[60] Karamata, J. (1933) Sur un mode de croissance régulière. Théorèmes fondamentaux. *Bull. Soc. Math. France* **61**, 55–62.

[61] Kesten, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**, 207–248.

[62] Klüppelberg, C. (1988) Subexponential distributions and integrated tails. *J. Appl. Probab.* **25**, 132–141.

[63] Klüppelberg, C. (1989) Subexponential distributions and characterisations of related classes. *Probab. Theory Related Fields* **82**, 259–269.

[64] Klüppelberg, C. and Mikosch, T. (1997) Large deviations of heavy–tailed random sums with applications to insurance and finance. *J. Appl. Probab.* **34**, 293–308.

[65] Kwapień, S. and Woyczyński, W.A. (1992) *Random Series and Stochastic Integrals: Single and Multiple.* Birkhäuser, Boston.

[66] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Processes.* Springer, Berlin.

[67] Le Page, E. (1983). Théorèmes de renouvellement pour les produits d e matrices aléatoires. Équations aux différences aléatoires. Séminaires de probabilités Rennes 1983. *Publ. Sém. Math.* Univ. Rennes I, 116 pp.

[68] Leslie, J. (1989) On the non–closure under convolution of the subexponential family. *J. Appl. Probab.* **26**, 58–66.

[69] Linnik, Yu.V. (1961a) Limit theorems allowing large deviations for the sums of independent variables I. *Theory Probab. Appl.* **6**, 145–161.

[70] Linnik, Yu.V. (1961b) Limit theorems allowing large deviations for the sums of independent variables II. *Theory Probab. Appl.* **6**, 377–391.

[71] LOÉVE, M. (1978) *Probability Theory,* 4th edition. Springer, Berlin.

[72] LONGIN, F. (1996) The Asymptotic Distribution of Extreme Stock Market Returns. *J. Busin.* **63**, 383-408.

[73] LORETAN, M. AND PHILLIPS, C.B. (1994) Testing the covariance stationarity of heavy–tailed time series: An overview of the theory with applications to several financial instrument s. *J. Empirical Finance* **1**, 211–248.

[74] MCCULLOCH, J.H. AND PANTON, D.B. (1996) Precise tabulation of the maximally–skewed stable distributions and densities. *Comput. Statist. Data Anal.* **22**. To appear.

[75] MEERSCHAERT, M.M. (1986) Regular variation and domains of attraction in $\mathbb{R}^k$. *Statist. Probab. Lett.* **1**, 43–45.

[76] MEYN, S.P. AND TWEEDIE, R.L. (1993) *Markov Chains and Stochastic Stability.* Springer, London.

[77] MIJNHEER, J.L. (1975) *Sample Path Properties of Stable Processes.* CWI Tract **59**, Amsterdam.

[78] MIKOSCH, T. AND NAGAEV, A.V. (1998) Large deviations for heavy–tailed sums with applications to insurance. *Extremes* **1**, 81-110.

[79] MIKOSCH, T. AND SAMORODNITSKY, G. (1998) The supremum of a negative drift random walk with dependent heavy–tailed steps. Preprint. University of Groningen. (www.cs.rug.nl/~eke/iwi/preprints)

[80] MIKOSCH, T. AND STĂRICĂ, C. (1998) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. Preprint. University of Groningen. (www.cs.rug.nl/~eke/iwi/preprints)

[81] MIKOSCH, T. AND STĂRICĂ, C. (1999) Change of structure in financial data, long–range dependence and GARCH modelling. In preparation.

[82] MIKOSCH, T. AND A. STEGEMAN (1999) The interaction between tails and rates in self–similar teletraffic. In preparation.

[83] NAGAEV, A.V. (1969a) Limit theorems for large deviations where Cramér's conditions are violated (in Russian). *Izv. Akad. Nauk UzSSR Ser. Fiz.–Mat. Nauk* **6**, 17–22.

[84] NAGAEV, A.V. (1969b) Integral limit theorems for large deviations when Cramér's condition is not fulfilled I,II. *Theory Probab. Appl.* **14**, 51–64 and 193–208.

[85] NAGAEV, A.V. (1970) *On Large Deviation Probabilities for Sums of Independent Random Variables.* Doctor of Science Thesis. Mathematical Institute, Tashkent.

[86] NAGAEV, A.V. (1971) A limit distribution of extreme terms of variational series under conditions of the type of large deviations imposed on the sample mean. *Theory Probab. Appl.* **16**, 118–131.

[87] NAGAEV, A.V. (1977) A property of sums of independent random variables. *Theory Probab. Appl.* **22**, 335–346.

[88] NAGAEV, S.V. (1965) Limit theorems on large deviations. *Theory Probab. Appl.* **10**, 231–254.

[89] Nagaev, S.V. (1979) Large deviations of sums independent random variables. *Ann. Prob.* **7**, 745–789.

[90] Nolan, J. (1996) An algorithm for evaluating stable densities in Zolotarev's (M) parameterization. Preprint, American University Washington.

[91] Nolan, J. (1996) Numerical approximations of stable densities and distribution functions. Preprint, American University Washington.

[92] Pakes, A.G. (1975) On the tails of waiting time distributions. *J. Appl. Probab.* **12**, 555-564.

[93] Petrov, V.V. (1954) The generalisation of the Cramér limit theorem (in Russian). *Uspekhi Mat. Nauk* **9**, 195–202.

[94] Petrov, V.V. (1965) On the probabilities of large deviations of the sums of independent random variables. *Theory Probab. Appl.* **10**, 287–298.

[95] Petrov, V.V. (1975) *Sums of Independent Random Variables.* Springer, Berlin.

[96] Petrov, V.V. (1995) *Limit Theorems of Probability Theory.* Oxford University Press, Oxford.

[97] Pitman, E.J.G. (1980) Subexponential distribution functions. *J. Austral. Math. Soc. Ser. A* **29**, 337–347.

[98] Pratt, J.W. (1960) On interchanging limits and integrals. *Ann. Math. Statist.* **31**, 74–77.

[99] Reiss, R.D. (1989) *Approximate Distributions of Order Statistics.* Springer, New York.

[100] Resnick, S.I. (1987) *Extreme Values, Regular Variation, and Point Processes.* Springer, New York.

[101] Resnick, S.I. (1997) Heavy tail modeling and teletraffic data. With discussion and a rejoinder by the author. *Ann. Statist.* **25**, 1805–1869.

[102] Resnick, S. and Willekens, E. (1991) Moving averages with random coefficients and random coefficient autoregressive models. *Comm. Statist. Stochastic Models* **7** (1991), 511–525.

[103] Rozovski, L.V. (1993) Probabilities of large deviations on the whole axis. *Theory Probab. Appl.* **38**, 53–79.

[104] Samorodnitsky, G. and Taqqu, M.S. (1994) *Stable Non–Gaussian Random Processes. Stochastic Models with Infinite Variance.* Chapman and Hall, London.

[105] Seneta, E. (1976) *Functions of Regular Variation.* Lecture Notes in Mathematics **506**. Springer, New York.

[106] Smith, W.L. (1972) On the tails of queuing time distributions. Mimeo Series No. 830, Department of Statistics, University of North Carolina, Chapel Hill.

[107] Teugels, J.L. (1975) The class of subexponential distributions. *Ann. Probab.* **3**, 1001–1011.

[108] TKACHUK, S.G. (1977) *Limit Theorems for Sums of Independent Random Variables from the Domain of Attraction of a Stable Law.* PhD Thesis. University of Tashkent.

[109] TURKMAN, K.F. AND TURKMAN, M.A.A. (1997) Extremes of bilinear time series models. *J.Time Ser. Anal.* **18**, 305–319.

[110] VERAVERBEKE, N. (1977) Asymptotic behaviour of Wiener–Hopf factors of a random walk. *Stoch. Proc. Appl.* **5**, 27–37.

[111] VERVAAT, W. (1979) On a stochastic differential equation and a representation of a non–negative infinitely divisible random variables. *Adv. Appl. Prob.* **11**, 750–783.

[112] VINOGRADOV, V. (1994) *Refined Large Deviation Limit Theorems.* Longman, Harlow.

[113] ZOLOTAREV, V.M. (1986) *One–Dimensional Stable Distributions.* American Mathematical Society, Providence, R.I.