A Short Course on Heavy Tails

Søren Asmussen*
Aarhus University, Denmark

Incomplete draft
Universidad de Chile, March 12, 2018

Contents

Course Plan

1°, March 7
Classification of tails
Regular variation

2°, March 8
Lognormal and Weibull
General heavy-tailed theory

3°, March 14
Extreme values
Heavy-tailed random walks

4°, March 21
Dependence

1 General set-up

$X, X_1, X_2, \ldots$ i.i.d. $\sim F$ (sometimes $\sim B$)
Tail $\overline{F}(x) = 1 - F(x) = \mathbb{P}(X > x)$
If density exists: f or b
Key focus: tails of

$$S_n = X_1 + \cdots + X_n\,, \qquad M_n = \max(X_1, \ldots, X_n)$$

Asymptotics of either
$\mathbb{P}(S_n > x)$, $\mathbb{P}(M_n > x)$ as $x \to \infty$ with n fixed; or
$\mathbb{P}(S_n > x_n)$, $\mathbb{P}(M_n > x_n)$ as $n \to \infty$ with suitable choice of x_n
Also tail of $M = \max(0, S_1, S_2 \ldots)$

Easiest case $\mathbb{P}(M_n > x)$ (completely general):

Proposition 1.1 $\mathbb{P}(M_n > x) \;\sim\; n\mathbb{P}(X > x) \;=\; n\overline{F}(x)$ *as* $x \to \infty$.

Proof: By inclusion-exclusion

$$\begin{aligned}
\mathbb{P}(M_2 > x) &= \mathbb{P}(X_1 > x \text{ or } X_2 > x) \\
&= \mathbb{P}(X_1 > x) + \mathbb{P}(X_2 > x) - \mathbb{P}(X_1 > x \text{ and } X_2 > x) \\
&= 2\mathbb{P}(X > x) - \mathbb{P}(X > x)^2 \;=\; 2\mathbb{P}(X > x) + \mathrm{o}(\mathbb{P}(X > x)) \\
&\sim 2\mathbb{P}(X > x)
\end{aligned}$$

□

Notes and references Most of the material in this set of notes is essentially a cut-and-paste from my forthcoming book [5] (co-author M. Steffensen). However, the random walk material in Section 5 is from [2].

2 Classification of tails

The term 'heavy tails' is used in a different meaning in different application areas. In finance, it means heavy compared to the Black-Scholes model, that is, heavier than the normal. According to this definition, the Weibull tail $\overline{F}(x) = \mathrm{e}^{-x^\beta}$ is heavy-tailed for $\beta < 2$, say for the exponential distribution that has $\beta = 1$. In contrast in most other areas, in particular OR and insurance, the benchmark is the exponential distribution rather than the normal so that the Weibull is heavy-tailed only when $\beta < 1$. We follow here this tradition which roughly leads to calling a distribution F with tail $\overline{F}(x) = 1 - F(x)$ light-tailed if the m.g.f. $\widehat{F}[s] = \int e^{sx}\, F(dx)$ is finite for some $s > 0$ and heavy-tailed if it is infinite for all $s > 0$.

Here is then a rough classification of well-behaved distributions F in terms of tails, arranged in order of decreasing heavy-tailedness (increasing light-tailedness), and with the line marking the cut between light and heavy tails:

Heavier-than-regularly varying: $\overline{F}(x) \to 0$ slower than $\dfrac{1}{x^\alpha}$ for all $\alpha > 0$

Regularly varying or *power-like:* $\overline{F}(x)$ close to $\dfrac{c}{x^\alpha}$ for some $\alpha > 0$

Lognormal-like: $\overline{F}(x)$ close to $\exp\{-c(\log x)^\beta\}$ with $\beta > 1$

Heavy-tailed Weibull-like: $\overline{F}(x)$ close to $\exp\{-cx^\beta\}$ with $\beta < 1$

Exponential-like: $\overline{F}(x)$ close to $\exp\{-\lambda x\}$

Light-tailed Weibull-like: $\overline{F}(x)$ close to $\exp\{-cx^\beta\}$ with $\beta > 1$

Lighter-than-Weibull: $\overline{F}(x) \to 0$ faster than $\exp\{-cx^\beta\}$ for all β.

In these notes, we treat the asymptotics of $\mathbb{P}(S_n > x)$ for the heavy-tailed RV, lognormal-like and Weibull-like classes in Section 3, whereas for comparison a short summary of the same problem for light tails is in Section 7.

3 Heavy Tails. Subexponential Distributions

Definition of subexponentiality and sufficient conditions

We are only concerned with positive r.v.s X, i.e. with distributions F concentrated on $(0, \infty)$. The main cases of heavy tails are:

(a) distributions with a regularly varying tail, $\overline{F}(x) = L(x)/x^\alpha$ where $\alpha > 0$ and $L(x)$ is slowly varying, $L(tx)/L(x) \to 1$, $x \to \infty$, for all $t > 0$.[1] A main example is the Pareto, i.e. the distribution of $X = aY + b$ where $\mathbb{P}(Y > y) = 1/(1+y)^\alpha$;

(b) the lognormal distribution (the distribution of e^V where $V \sim N(\mu, \sigma^2)$). The

[1] Some main examples of slowly varying functions are: i) functions with a limit in $(0, \infty)$, $L(x) \to \ell$ where $0 < \ell < \infty$; ii) $L(x) \sim (\log x)^\beta$ with $-\infty < \beta < \infty$. But see Exercise 3.1 for an example with weird properties.

density and asymptotic tail (cf. Mill's ratio) are given by

$$f(x) \;=\; \frac{1}{x\sqrt{2\pi\sigma^2}}\mathrm{e}^{-(\log x-\mu)^2/2\sigma^2}\,, \qquad (3.1)$$

$$\overline{F}(x) \;=\; \overline{\Phi}\big((\log x-\mu)/\sigma\big) \;\sim\; \frac{1}{\log x\sqrt{2\pi}}\mathrm{e}^{-(\log x-\mu)^2/2\sigma^2}\,; \qquad (3.2)$$

(c) the Weibull distribution with decreasing failure rate, $\overline{F}(x)=\mathrm{e}^{-x^\beta}$ with $0<\beta<1$.

Here the regularly varying tails are the heaviest (and the heavier the smaller α is) and the Weibull the lightest (and the lighter the larger β is). The lognormal case is intermediate (here the heaviness increases with σ^2). This ordering is illustrated by the behavior of the pth moments $\mathbb{E}X^p$ which for regular variation are finite only for $p<\alpha$ and for the lognormal and Weibull for all p; however, for the lognormal $\mathbb{E}X^p=\exp\{p\mu+p^2\sigma^2/2\}$ grows so rapidly in p that the lognormal is not determined by its moments (cf. the celebrated *Stieltjes moment problem*).

The definition $\widehat{F}[s]=\infty$ for all $s>0$ of heavy tails is too general to allow for many general non-trivial results, and instead we shall work within the class $\mathcal{S}$ of *subexponential* distributions; good general reference are Foss, Korshunov & Zachary [15] and Embrechts, Klüppelberg & Mikosch [10]. For the definition, we require that F is concentrated on $(0,\infty)$ and say then that F is subexponential ($F\in\mathcal{S}$) if

$$\frac{\overline{F^{*2}}(x)}{\overline{F}(x)} \;\to\; 2, \quad x\to\infty. \qquad (3.3)$$

Here F^{*2} is the convolution square, that is, the distribution of independent r.v.s X_1,X_2 with distribution F. In terms of r.v.'s, (3.3) then means $\mathbb{P}(X_1+X_2>x)\sim 2\mathbb{P}(X_1>x)$.

To capture the intuition behind this definition, note first the following fact:

Proposition 3.1 *Let F be any distribution on $(0,\infty)$. Then:*
(a) $\mathbb{P}(\max(X_1,X_2)>x) \;\sim\; 2\overline{F}(x), \quad x\to\infty.$
(b) $\liminf_{x\to\infty}\dfrac{\overline{F^{*2}}(x)}{\overline{F}(x)} \;\geq\; 2.$

Proof: By the inclusion-exclusion formula, $\mathbb{P}(\max(X_1,X_2)>x)$ equals

$$\mathbb{P}(X_1>x)+\mathbb{P}(X_2>x)-\mathbb{P}(X_1>x,X_2>x) \;=\; 2\overline{F}(x)-\overline{F}(x)^2 \;\sim\; 2\overline{F}(x),$$

proving (a). Since F is concentrated on $(0,\infty)$, we have $\{\max(X_1,X_2)>x\}\subseteq\{X_1+X_2>x\}$, and thus the lim inf in (b) is at least

$$\liminf \mathbb{P}(\max(X_1,X_2)>x)/\overline{F}(x) \;=\; 2\,. \qquad \square$$

□

The proof shows that the condition for $F \in \mathcal{S}$ is that the probability of the set $\{X_1 + X_2 > x\}$ is asymptotically the same as the probability of its subset $\{\max(X_1, X_2) > x\}$. That is, in the subexponential case *the only way* $X_1 + X_2$ *can get large is by one of the* X_i *becoming large.*

Example 3.2 As an example of what happens for a light-tailed distribution, consider the standard exponential distribution, $\overline{F}(x) = \mathrm{e}^{-x}$. Then $X_1 + X_2$ has an Erlang(2) distribution with density $x\mathrm{e}^{-x}$ so that $\overline{F^{*2}}(x) \sim x\mathrm{e}^{-x}$. Thus the lim inf in Proposition 3.1(b) is ∞. One can further check that

$$\left(\frac{X_1}{x}, \frac{X_2}{x}\right)\bigg| X_1 + X_2 > x \ \xrightarrow{\mathcal{D}}\ (U, 1-U) \tag{3.4}$$

where U is uniform on $(0,1)$. Thus, if X_1+X_2 is large, then (with high probability) so are both of X_1, X_2 but none of them exceeds x. ◇

Remark 3.3 As a generalization of (a), it is not difficult to show that if X_1, X_2 are independent with distribution F_i for X_i, then

$$\mathbb{P}\big(\max(X_1, X_2) > x\big) \ \sim\ \overline{F}_1(x) + \overline{F}_2(x)\,.$$ ◇

Regular variation is the simplest example of subexponentiality:

Proposition 3.4 *Any* F *with a regularly varying tail is subexponential. More generally, assume* F_1, F_2 *satisfy* $\overline{F}_1(x) \sim c_1\overline{F}_0(x)$, $\overline{F}_2(x) \sim c_2\overline{F}_0(x)$ *where* $\overline{F}_0(x) = L(x)/x^\alpha$ *for some slowly varying* $L(x)$, *some* $\alpha > 0$ *and some* $c_1, c_2 \geq 0$ *with* $c = c_1 + c_2 > 0$. *Then* $F = F_1 * F_2$ *satisfies* $\overline{F}(x) \sim cL(x)/x^\alpha$

Proof: It suffices to prove the second assertion, since the first is just the special case $c_1 = c_2 = 1$. Let $0 < \delta < 1/2$. Then the event $\{X_1 + X_2 > x\}$ is a subset of

$$\big\{X_1 > (1-\delta)x\big\} \cup \big\{X_2 > (1-\delta)x\big\} \cup \big\{X_1 > \delta x, X_2 > \delta x\big\}$$

(if say $X_2 \leq \delta x$, one must have $X_1 > (1-\delta)x$ for $X_1 + X_2 > x$ to occur). Hence

$$\begin{aligned}
\limsup_{x\to\infty} \frac{\overline{F}(x)}{\overline{F}_0(x)} \ &= \ \limsup_{x\to\infty} \frac{\mathbb{P}(X_1 + X_2 > x)}{\overline{F}_0(x)} \\
&\leq\ \limsup_{x\to\infty} \frac{\overline{F}_1\big((1-\delta)x\big) + \overline{F}_2\big((1-\delta)x\big)\big) + \overline{F}_1(\delta x)\overline{F}_2(\delta x)}{\overline{F}_0(x)} \\
&=\ \limsup_{x\to\infty} \frac{cL\big(1-\delta)x\big)}{\big[(1-\delta)x\big]^\alpha L(x)/x^\alpha} \ +\ 0 \ =\ \frac{c}{(1-\delta)^\alpha}
\end{aligned}$$

(to get $\overline{F}_0(\delta x)^2/\overline{F}_0(x) \to 0$, note first that $\overline{F}_0(\delta x)^2/\overline{F}_0(\delta x) \to 0$ and next that $\overline{F}_0(\delta x)/\overline{F}_0(x) \to 0$ has limit $\delta^{-\alpha}$). Letting $\delta \downarrow 0$, we get $\limsup \overline{F}(x)/\overline{F}_0(x) \le c$, and combining with Remark 3.3 we get $\overline{F}(x)/\overline{F}_0(x) \to c$. □

For other types of distributions, a classical sufficient (and close to necessary) condition for subexponentiality is due to Pitman [22]. Recall that the failure rate $\lambda(x)$ of a distribution F with density f is $\lambda(x) = f(x)/\overline{F}(x)$.

Proposition 3.5 *Let F have density f and failure rate $\lambda(x)$ such that $\lambda(x)$ is decreasing for $x \ge x_0$ with limit 0 at ∞. Then $F \in \mathcal{S}$ provided*

$$\int_0^\infty \mathrm{e}^{x\lambda(x)} f(x)\,\mathrm{d}x \;<\; \infty\,.$$

Proof: We may assume that $\lambda(x)$ is everywhere decreasing (otherwise, replace F by a tail equivalent distribution with a failure rate which is everywhere decreasing, cf. Corollary 3.14 below). Define $\Lambda(x) = \int_0^x \lambda(y)\,\mathrm{d}y$. Then $\overline{F}(x) = \mathrm{e}^{-\Lambda(x)}$ (standard but the proof is immediate by observing that $\lambda(x) = -\mathrm{d}/\mathrm{d}x \log \overline{F}(x)$), and we get

$$\begin{aligned}
\frac{\overline{F^{*2}}(x)}{\overline{F}(x)} - 1 \;&=\; \frac{\overline{F^{*2}}(x) - \overline{F}(x)}{\overline{F}(x)} \;=\; \frac{F(x) - F^{*2}(x)}{\overline{F}(x)} \\
&= \int_0^x \frac{1 - F(x-y)}{\overline{F}(x)} f(y)\,\mathrm{d}y \;=\; \int_0^x \frac{\overline{F}(x-y)}{\overline{F}(x)} f(y)\,\mathrm{d}y \\
&= \int_0^x \mathrm{e}^{\Lambda(x)-\Lambda(x-y)-\Lambda(y)}\lambda(y)\,\mathrm{d}y \\
&= \int_0^{x/2} \mathrm{e}^{\Lambda(x)-\Lambda(x-y)-\Lambda(y)}\lambda(y)\,\mathrm{d}y \qquad (3.5) \\
&\quad + \int_0^{x/2} \mathrm{e}^{\Lambda(x)-\Lambda(x-y)-\Lambda(y)}\lambda(x-y)\,\mathrm{d}y\,. \qquad (3.6)
\end{aligned}$$

For $y < x/2$,

$$\Lambda(x) - \Lambda(x-y) \;\le\; y\lambda(x-y) \;\le y\lambda(y).$$

The rightmost bound shows that the integrand in (3.5) is bounded by $\mathrm{e}^{y\lambda(y)-\Lambda(y)}\lambda(y) = \mathrm{e}^{y\lambda(y)} f(y)$, an integrable function by assumption. The middle bound shows that it converges to $f(y)$ for any fixed y since $\lambda(x-y) \to 0$. Thus by dominated convergence, (3.5) has limit 1. Since $\lambda(x-y) \le \lambda(y)$ for $y < x/2$, we can use the same domination for (3.6) but now the integrand has limit 0. Thus $\overline{F^{*2}}(x)/\overline{F}(x) - 1$ has limit $1 + 0$, proving $F \in \mathcal{S}$. □

Example 3.6 Consider the DFR (decreasing failure rate) Weibull case $\overline{F}(x) = \mathrm{e}^{-x^\beta}$ with $0 < \beta < 1$. Then $f(x) = \beta x^{\beta-1}\mathrm{e}^{-x^\beta}$, $\lambda(x) = \beta x^{\beta-1}$. Thus $\lambda(x)$ is everywhere decreasing, and $\mathrm{e}^{x\lambda(x)}f(x) = \beta x^{\beta-1}\mathrm{e}^{-(1-\beta)x^\beta}$ is integrable. Thus, the DFR Weibull distribution is subexponential. ◇

Example 3.7 In the lognormal distribution, we have by (3.1), (3.2) that $\lambda(x) \sim \log x/\sigma x$. This yields easily that $\mathrm{e}^{x\lambda(x)}f(x)$ is integrable. Further, elementary but tedious calculations (which we omit) show that $\lambda(x)$ is ultimately decreasing. Thus, the lognormal distribution is subexponential. ◇

Further mathematical properties

Proposition 3.8 *A function $L > 0$ is slowly varying if and only if it has the form*

$$L(x) \;=\; c(x)\exp\Bigl\{\int_a^x \frac{\epsilon(u)}{u}\,\mathrm{d}u\Bigr\}$$

where $c(x) \to c \in (0,\infty)$ and $\epsilon(x) \to 0$.

Proof: Omitted. □

Proposition 3.9 *If $F \in \mathcal{S}$, then $\dfrac{\overline{F}(x-y)}{\overline{F}(x)} \to 1$ uniformly in $y \in [0, y_0]$ as $x \to \infty$.*

In terms of r.v's: if $X \sim F \in \mathcal{S}$, then the overshoot $X - x | X > x$ converges in distribution to ∞. This follows since the probability of the overshoot to exceed y is $\overline{F}(x+y)/\overline{F}(x)$ which has limit 1. A distribution with the property $\overline{F}(x-y)/\overline{F}(x) \to 1$ is often called *long-tailed*.

Proof: Consider first a fixed y. Here and in the following, we shall several times use the identity

$$\begin{aligned}\frac{\overline{F^{*(n+1)}}(x)}{\overline{F}(x)} \;&=\; 1 + \frac{F(x) - F^{*(n+1)}(x)}{\overline{F}(x)} \\ &=\; 1 + \int_0^x \frac{1 - F^{*n}(x-z)}{\overline{F}(x)}\,F(\mathrm{d}z) \;=\; 1 + \int_0^x \frac{\overline{F^{*(n)}}(x-z)}{\overline{F}(x)}\,F(\mathrm{d}z)\,. \qquad (3.7)\end{aligned}$$

Taking $n = 1$ and splitting the integral into two corresponding to the intervals $[0, y]$ and $(y, x]$, we get

$$\frac{\overline{F^{*2}}(x)}{\overline{F}(x)} \;\ge\; 1 + F(y) + \frac{\overline{F}(x-y)}{\overline{F}(x)}\bigl(F(x) - F(y)\bigr)\,.$$

If $\limsup \overline{F}(x-y)/\overline{F}(x) > 1$, we therefore get

$$\limsup \frac{\overline{F^{*2}}(x)}{\overline{F}(x)} \; > \; 1 + F(y) + 1 - F(y) \; = 2,$$

a contradiction. Finally $\liminf \overline{F}(x-y)/\overline{F}(x) \geq 1$ since $y \geq 0$.

The uniformity now follows from what has been shown for $y = y_0$ and the obvious inequality

$$1 \;\leq\; \frac{\overline{F}(x-y)}{\overline{F}(x)} \;\leq\; \frac{\overline{F}(x-y_0)}{\overline{F}(x)}, \quad y \in [0, y_0].$$

□

Proposition 3.10 *If $F \in \mathcal{S}$, then for any n $\overline{F^{*n}}(x)/\overline{F}(x) \;\to\; n$, $x \to \infty$.*

This extension from the case $n = 2$ is often taken as definition of the class $\mathcal{S}$; its intuitive content is the same as discussed in the case $n = 2$ above. Note that in the special case of regular variation, we can take $F_0 = F_1 = F$ and $F_2 = F^{*2}$ in Proposition 3.4 to get $\overline{F}^{*3}(x) \sim 3\overline{F}(x)$. Continuing in this manner, Proposition 3.10 follows for regular variation.

Proof: We use induction. The case $n = 2$ is just the definition, so assume the proposition has been shown for n. Given $\epsilon > 0$, choose y such that $\left|\overline{F^{*n}}(x)/\overline{F}(x) - n\right| \leq \epsilon$ for $x \geq y$. Then by (3.7),

$$\frac{\overline{F^{*(n+1)}}(x)}{\overline{F}(x)} \;=\; 1 + \left(\int_0^{x-y} + \int_{x-y}^{x}\right) \frac{\overline{F^{*n}}(x-z)}{\overline{F}(x-z)} \frac{\overline{F}(x-z)}{\overline{F}(x)} \, F(\mathrm{d}z)\,.$$

Here the second integral can be bounded above by

$$\sup_{v \geq 0} \frac{\overline{F^{*n}}(v)}{\overline{F}(v)} \frac{F(x) - F(x-y)}{\overline{F}(x)}\,,$$

which converges to 0 by Proposition 3.9 and the induction hypothesis. The first integral is bounded by

$$\begin{aligned}
&(n+\epsilon) \int_0^{x-y} \frac{\overline{F}(x-z)}{\overline{F}(x)} \, F(\mathrm{d}z) \\
&\quad = \; (n+\epsilon) \left\{ \frac{F(x) - F^{*2}(x)}{\overline{F}(x)} - \int_{x-y}^{x} \frac{\overline{F}(x-z)}{\overline{F}(x)} \, F(\mathrm{d}z) \right\}.
\end{aligned}$$

Here the first term in $\{\cdot\}$ converges to 1 (by the definition of $F \in \mathcal{S}$) and the second to 0 since it is bounded by $\big(F(x) - F(x-y)\big)/\overline{F}(x)$. Combining these estimates and letting $\epsilon \downarrow 0$ completes the proof. □

Corollary 3.11 *If $F \in \mathcal{S}$, then* $\mathrm{e}^{\epsilon x}\overline{F}(x) \to \infty$, $\widehat{F}[\epsilon] = \infty$ *for all* $\epsilon > 0$.

Proof: For $0 < \delta < \epsilon$, we have by Proposition 3.9 that $\overline{F}(n) \geq \mathrm{e}^{-\delta}\overline{F}(n-1)$ for all large n so that $\overline{F}(n) \geq c_1\mathrm{e}^{-\delta n}$ for all n. This implies $\overline{F}(x) \geq c_2\mathrm{e}^{-\delta x}$ for all x, and this immediately yields the desired conclusions. □

Proposition 3.12 *Let A_1, A_2 be distributions on $(0,\infty)$ such that $\overline{A}_i(x) \sim a_i\overline{F}(x)$ for some $F \in \mathcal{S}$ and some constants a_1, a_2 with $a_1 + a_2 > 0$. Then $\overline{A_1 * A_2}(x) \sim (a_1+a_2)\overline{F}(x)$.*

Proof: Let X_1, X_2 be independent r.v.s such that X_i has distribution A_i. Fix v. Then for $x > v/2$ the events

$$\{X_1 > x - v, X_2 > x - v\},\ \{X_1 \leq v\},\ \{X_2 \leq v\}, \tag{3.8}$$

$$\{X_1 \leq x - v, X_2 > v\},\ \{X_1 > v, X_2 \leq x - v\} \tag{3.9}$$

have disjoint intersections with $\{X_1 + X_2 > x\}$ so that

$$\overline{A_1 * A_2}(x) \ = \ \mathbb{P}(X_1 + X_2 > x) \ = C_0 + C_{12} + C_{21} + D_{12} + D_{21}$$

where $C_0, \ldots, D_{21}$ are the corresponding contributions to $\overline{A_1 * A_2}(x)$. For any fixed v, Proposition 3.9 easily yields

$$C_{12} \ = \ \int_0^v \overline{A}_2(x-y)A_1(\mathrm{d}y) \ \sim \ a_2\overline{F}(x)A_1(v) \ = \ a_2\overline{F}(x)\big(1 + o_v(1)\big)$$

and similarly $C_{21} = a_2\overline{F}(x)\big(1 + o_v(1)\big)$. Also

$$C_0 \ \leq \ \overline{A}_1(x-v)\overline{A}_2(x-v) \ \sim \ a_1a_2\overline{F}(x)^2 \ = \ \mathrm{o}_x\big(\overline{F}(x)\big)$$

so that

$$\overline{A_1 * A_2}(x) \ = \ \big(\big(a_1 + a_2 + \mathrm{o}_v(1)\big)\overline{F}(x) \ + D_{12} + D_{21}\,. \tag{3.10}$$

Taking $A_1 = A_2 = F$ for a moment and using $F \in \mathcal{S}$ gives $D_{12}^F + D_{21}^F = \mathrm{o}_x\big(\overline{F}(x)\big)$ in obvious notation. Finally

$$\begin{aligned} D_{12} \ &= \ \int_v^{x-v} \overline{A}_1(x-y)A_2(\mathrm{d}y) \ = \ \mathrm{O}(1)\int_v^{x-v} F(x-y)A_2(\mathrm{d}y) \\ &= \ \mathrm{O}(1)\Big[\overline{F}(x-v)\overline{A}_2(v) - \overline{A}_i(x-v)\overline{F}(v) + \int_v^{x-v}\overline{A}_2(x-y)B(\mathrm{d}y)\Big] \\ &\leq \ \overline{F}(x)\mathrm{o}_v(1) - \overline{F}(x)\mathrm{o}_v(1) + \mathrm{O}(1)D_{12}^F \ = \ \overline{F}(x)\mathrm{o}_v(1) \end{aligned}$$

where the third equality follows by integration by parts. Similarly, $C_{21} = \overline{F}(x)\mathrm{o}_v(1)$ so that (3.10) completes the proof. □

Corollary 3.13 *The class $\mathcal{S}$ is closed under tail–equivalence. That is, if $\overline{A}(x) \sim a\overline{F}(x)$ for some $F \in \mathcal{S}$ and some constant $a > 0$, then $A \in \mathcal{S}$.*

Proof: Taking $A_1 = A_2 = A$, $a_1 = a_2 = a$ gives $\overline{A^{*2}}(x) \sim 2a\overline{F}(x) = 2a\overline{A}(x)$. □

Corollary 3.14 *Let $F \in \mathcal{S}$ and let A be any distribution with a lighter tail, $\overline{A}(x) = \mathrm{o}\big(\overline{F}(x)\big)$. Then $A * F \in \mathcal{S}$ and $\overline{A * F}(x) \sim \overline{F}(x)$*

Proof: Take $A_1 = A, A_2 = F$ so that $a_1 = 0, a_2 = 1$. □

Proposition 3.15 *Let $Y_1, Y_2, \ldots$ be i.i.d. with common distribution $G \in \mathcal{S}$ and let N be an independent integer–valued r.v. with $\mathbb{E}z^N < \infty$ for some $z > 1$. Then $\mathbb{P}(Y_1 + \cdots + Y_N > u) \sim \mathbb{E}N\,\overline{G}(u)$.*

Proof: Using $\overline{G^{*n}}(u) \sim n\overline{G}(u)$, $u \to \infty$, one can use the basic decomposition (3.7) in a similar manner as in the proofs of Propositions 3.9 and 3.10 to show thta for each $z > 1$ there is a $D < \infty$ such that $\overline{G^{*n}}(u) \le \overline{G}(u)Dz^n$ for all u. Therefore we can use dominated convergence with $\sum \mathbb{P}(N = n)\,Dz^n$ as majorant to obtain

$$\frac{\mathbb{P}(Y_1 + \cdots + Y_N > u)}{\overline{G}(u)} \;=\; \sum_{n=0}^{\infty} \mathbb{P}(N = n)\frac{\overline{G^{*n}}(u)}{\overline{G}(u)} \;\to\; \sum_{n=0}^{\infty} \mathbb{P}(N = n)\cdot n \;=\; \mathbb{E}N. \quad \square$$

Exercises

3.1. Let $L(x) > 0$ be a differentiable function and let $M(x) = \log L(\mathrm{e}^x)$. Show that if $M'(x) \to 0$ as $x \to \infty$ then L is slowly varying. Show thereby that $L(x) = \exp\{(\log x)^\alpha \cos(\log x)^\beta\}$ is slowly varying with $\liminf L(x) = 0$, $\limsup L(x) = \infty$ when $\alpha, \beta > 0$, $\alpha + \beta < 1$.
3.2. Verify (3.4).
3.3. Use Pitman's criterion to show that the Pareto distribution with tail $(1 + x)^{-\alpha}$ is subexponential without referring to Proposition 3.5.
3.4. Show that the distribution with tail $\mathrm{e}^{-x/\log(x+\mathrm{e})}$ is subexponential.
3.5. Show that if $F \in \mathcal{S}$, then

$$\mathbb{P}(X_1 > x \mid X_1 + X_2 > x) \to \frac{1}{2}, \quad \mathbb{P}(X_1 \le y \mid X_1 + X_2 > x) \to \frac{1}{2}F(y).$$

That is, given $X_1 + X_2 > x$, the r.v. X_1 is w.p. 1/2 'typical' (with distribution F) and w.p. 1/2 it has the distribution of $X_1 | X_1 > x$.
3.6. Show that if $X_1, X_2 \ge 0$ are independent and X_1 subexponential, then $\mathbb{P}(X_1 - X_2 > x) \sim \mathbb{P}(X_1 > x)$.

4 Extreme Value Theory

We consider a sequence $X_1, X_2, \ldots$ of random variables and are interested in properties of the maxmum

$$M_n \ = \ \max(X_1, \ldots, X_n)$$

and the minimum; note, however, that it is sufficient to consider maxima instead of minima because of the simple relation

$$\min(X_1, \ldots, X_n) \ = \ -\max(-X_1, \ldots, -X_n)\,.$$

The X_i will be i.i.d. with common distribution F. The exact distribution of M_n is then easily available: the c.d.f. of M_n is

$$\mathbb{P}(M_n \le x) \ = \ \mathbb{P}\big(X_1 \le x, \ldots, X_n \le x\big) \ = \ F^n(x)\,.$$

Our concern will be to obtain more summary information, for example the growth rate or asymptotic distribution of M_n. It is important to distinguish between the theories of extreme values and rare events studied elsewhere in the notes. Extreme value theory, the topic of this secton, is about the *typical* behavior of the extremes as defined by the maxima and minima, not about calculating probabilities of order say 10^{-5} or less. For example, when studying the risk of earthquakes in Chile in the next 50 years we will get qualified answers to questions such as what is the median or the 90% quantile of the maximal size of the earthquakes in a year, but not to what is the risk of a catastrophe which would normally only occur in a time span of hundred thousands of years.

The topic has to do with convergence of sequences of the form $\mathbb{P}(M_n \le u_n)$. Namely, if we want to show that $\Phi_n(M_n)$ has a limit in distribution, the relevant choice is $u_n = \Phi_n^{-1}(x)$ for continuity points x of the limit. To this end, the following simple observation is extremely useful (note that any limit in $(0, \infty)$ can be written as $\mathrm{e}^{-\tau}$).

Proposition 4.1 $\mathbb{P}(M_n \le u_n) \to \mathrm{e}^{-\tau} \iff n\overline{F}(u_n) \to \tau$.

Proof: Just note that

$$\log \mathbb{P}(M_n \le u_n) \ = \ \log\big(1 - \overline{F}(u_n)\big)^n \ \sim \ -n\overline{F}(u_n)$$

if $\overline{F}(u_n) \to 0$. □

The case of a standard uniform F plays in some sense a special role, so we introduce a special notation for it: if $U_1, \ldots, U_n$ are i.i.d. uniform$(0,1)$, we write $M_n^U \ = \ \max(U_1, \ldots, U_n)$.

Example 4.2 Consider the standard uniform case $F(u) = u$, $0 < u < 1$. Then $\overline{F}(u) = 1-u$ for $0 < u < 1$, and to get a limit τ of $n\overline{F}(u_n)$, the choice $u_n = 1-y/n$ will work to get the limit $\tau = y$ of $n\overline{F}(u_n)$ corresponding to the limit $\mathrm{e}^{-\tau} = \mathrm{e}^{-y}$ of $F^n(1-y/n)$. This means that

$$\mathbb{P}\big(n(1-M_n^U) \geq y\big) \;=\; \mathbb{P}(M_n^U \leq 1-y/n) \;=\; F^n(1-y/n) \;\to\; \mathrm{e}^{-y} \;=\; \mathbb{P}(V \geq y)$$

where V is standard exponential. I.e., $n(1-M_n^U) \xrightarrow{\mathcal{D}} V$ which we can rewrite as $M_n^U \sim 1 - V/n$.

A direct proof without reference to Proposition 4.1 is of course only a small variant:

$$\mathbb{P}(M_n^U \leq u_n) \;=\; F^n(u_n) \;=\; (1-y/n)^n \;\to\; \mathrm{e}^{-y}\,,$$

◇

Example 4.3 Let F be standard exponential, $F(x) = 1-\mathrm{e}^{-x}$ for $x > 0$. To get a limit of $F^n(u_n)$ or equivalently of $n\overline{F}(u_n)$, we take u_n such that $\mathrm{e}^{-u_n} = z/n$. Then

$$n\overline{F}(u_n) = z\,, \qquad F^n(u_n) \;=\; \Big(1-\frac{z}{n}\Big)^n \;\to\; \mathrm{e}^{-z}\,.$$

This means

$$\mathrm{e}^{-z} \;=\; \lim_{n\to\infty} \mathbb{P}(M_n \leq u_n) \;=\; \lim_{n\to\infty} \mathbb{P}(M_n \leq \log n - \log z)$$

which, taking $z = \mathrm{e}^{-y}$, we can rewrite as

$$\mathrm{e}^{-\mathrm{e}^{-y}} \;=\; \lim_{n\to\infty} \mathbb{P}(M_n \leq \log n + y)\,.$$

I.e., $M_n - \log n$ converges in distribution to a r.v. Y with c.d.f. $G(y) = \mathrm{e}^{-\mathrm{e}^{-y}}$, $y \in \mathbb{R}$. This distribution is known as the *Gumbel distribution.* ◇

The uniform case is in some sense the most fundamental one since the general case follows from the uniform one by transformation with the inverse c.d.f. In more detail, assume F is supported by some open real interval I, and that F is strictly increasing and continuous on I. For $u \in (0,1)$, we define $x = F^{\leftarrow}(u)$ as the unique solution $x \in I$ of $u = F(x)$. Then $F^{\leftarrow}$ is a strictly increasing and continuous bijection $(0,1) \to I$ [the setup can be generalized at the cost of some technicalities; we omit these here]. Further, a r.v. X with distribution F can be represented as $X \overset{\mathscr{D}}{=} F^{\leftarrow}(U)$ with U uniform$(0,1)$:

$$\mathbb{P}\big(F^{\leftarrow}(U) \leq x\big) \;=\; \mathbb{P}\big(U \leq F(x)\big) \;=\; F(x)\,.$$

It follows that

$$M_n \stackrel{\mathscr{D}}{=} \max\bigl(F^{\leftarrow}(U_1),\dots,F^{\leftarrow}(U_n)\bigr) \;=\; F^{\leftarrow}(M_n^U)\,. \tag{4.1}$$

Example 4.3 continued. In the exponential case, $I = (0,\infty)$ and $F^{\leftarrow}(u) = -\log(1-u)$. Using $M_n^U \sim 1 - V/n$ leads to

$$M_n \;\approx\; -\log\bigl(1-(1-V/n)\bigr) \;=\; \log n - \log V\,.$$

But the c.d.f. of $-\log V$ is

$$\mathbb{P}(-\log V \le y) \;=\; \mathbb{P}(V \ge \mathrm{e}^{-y}) \;=\; \mathrm{e}^{-\mathrm{e}^{-y}}\,,$$

so $-\log V$ is Gumbel and we are back to the conclusion of Example 4.3. ◇

It turns out, that when considering only *affine* transformations of M_n, i.e. of the form $(M_n - a_n)/b_n$ with $b_n > 0$, only three types[2] are possible: the Gumbel distribution with c.d.f. $G(y) = \mathrm{e}^{-\mathrm{e}^{-y}}$, $y \in \mathbb{R}$, the *Fréchet distribution* with c.d.f. $G(y) = \mathrm{e}^{-1/y^\alpha}$ for $y > 0$, and the *Weibull distribution* with c.d.f. $G(y) = \mathrm{e}^{-|y|^\alpha}$ for $y < 0$. This results is usually referred to as the *Fisher-Tippett theorem.* A Weibull limit can only occur if F has finite support, and so it will not concern us here since we are concerned with heavy right tails. The key result relating to heavy tails is then that a Fréchet limit occurs precisely for F RV and a Gumbel limit for F lognormal and Weibull. If for example $(M_n - a_n)/b_n$ has a Gumbel limit, we write $F \in \mathrm{MDA}(\mathrm{Gumbel})$.

The simplest case is convergence to the Fréchet:

Theorem 4.4 *Assume that $\overline{F}(x)$ is regularly varying at $x = \infty$, $\overline{F}(x) = L(x)/x^\alpha$ with $\alpha > 0$ and L slowly varying. Then $F \in \mathrm{MDA}(\text{Fréchet}_\alpha)$ and one may take $a_n = F^{\leftarrow}(1-1/n)$, $b_n = 0$.*

Proof: With $a_n = F^{\leftarrow}(1-1/n)$, we have for $t < 1$ that

$$\liminf_{n\to\infty} \frac{n\overline{F}(a_n)}{n\overline{F}(a_n-)} \;\ge\; \liminf_{n\to\infty} \frac{\overline{F}(a_n)}{\overline{F}(ta_n)} \;=\; t^\alpha$$

which tends to 1 as $t \uparrow 1$ (we used here that $L(ta_n)/L(a_n) \to 1$). It thus follows from $n\overline{F}(a_n) \le 1 \le n\overline{F}(a_n-)$ that $n\overline{F}(a_n) \to 1$. We then get

$$n\overline{F}\bigl(a_n x)\bigr) \;\sim\; n\overline{F}(a_n)/x^\alpha \;\sim\; 1/x^\alpha$$

[2] Two r.v.'s Y, Z are of the same type if $Z \stackrel{\mathscr{D}}{=} aY + b$ with $a > 0$.

(using $L(a_nx)/L(a_n) \to 1$) so that $\mathbb{P}(M_n \le a_nx) \to \mathrm{e}^{-1/x^\alpha}$ by Proposition 4.1. □

Recall that the Gumbel distribution has c.d.f. $\mathrm{e}^{-\mathrm{e}^{-x}}$. MDA(Gumbel) turns out to be the class of distributions F such that there exists a function $e(t)$ such that if $X \sim F$, then the conditional distribution of the overshoot $X - t$ given $X > t$ asymptotically is that of $e(t)V$ where V is a standard exponential r.v. That is, the distribution of $(X-t)/e(t)$ given $X > t$ should convergence to that of V or equivalently, one should have

$$\lim_{t\to\infty} \mathbb{P}\bigl((X-t)/e(t) > y \bigm| X > t\bigr) \;=\; \lim_{t\to\infty} \frac{\overline{F}\bigl(t + ye(t)\bigr)}{\overline{F}(t)} \;=\; \mathrm{e}^{-y} \tag{4.2}$$

for all $y > 0$ (below, we require (4.2) for all $y \in \mathbb{R}$ which is usually as easy to verify).

Theorem 4.5 *Assume* (4.2) *holds for all* $y \in \mathbb{R}$*. Then* $F \in$ *MDA(Gumbel) and one may take* $b_n = F^{\leftarrow}(1-1/n)$*,* $a_n = e(b_n)$*.*

Condition (4.2) may appear rather special, but the examples below show the flexibility.

Proof: Let $b_n = F^{\leftarrow}(1-1/n)$. By part 3) of Proposition **??**, $n\overline{F}(b_n) \le 1 \le n\overline{F}(b_n-)$ and for $y > 0$

$$\liminf_{n\to\infty} \frac{n\overline{F}(b_n)}{n\overline{F}(b_n-)} \;\ge\; \liminf_{n\to\infty} \frac{\overline{F}(b_n)}{\overline{F}\bigl(b_n - ye(b_n)\bigr)} \;=\; \mathrm{e}^{y}\,.$$

Letting $y \downarrow 0$ gives $n\overline{F}(b_n) \to 1$. We then get

$$n\overline{F}\bigl(b_n + xe(b_n)\bigr) \;\sim\; n\overline{F}(b_n)\mathrm{e}^{-x} \;\sim\; \mathrm{e}^{-x}$$

so that $\mathbb{P}\bigl(M_n \le b_n + xe(n_n)\bigr) \to \mathrm{e}^{-\mathrm{e}^{-x}}$ by Proposition 4.1. □

Remark 4.6 It can be shown that $e(x)$ may be chosen as the mean overshoot

$$\mathbb{E}\bigl[(X-x) \,|\, X > x\bigr] \;=\; \frac{1}{\mathbb{P}(X > x)}\mathbb{E}[X-x]^+ \;=\; \frac{1}{\overline{F}(x)} \int_x^\infty \overline{F}(t)\,\mathrm{d}t \tag{4.3}$$

If a density f exists, then by L'Hôpital's rule the r.h.s. is asymptotically $\overline{F}(x)/f(x)$ (the inverse failure rate) and so this is another possible choice. ◇

Example 4.7 Consider the Weibull distribution on $(0,\infty)$ with $\overline{F}(x) = \mathrm{e}^{-x^\beta}$ and corresponding density $f(x) = \beta x^{\beta-1}\mathrm{e}^{-x^\beta}$. Using Remark 4.6 we take $e(x) = 1/\beta x^{\beta-1}$ as trial. Now by Taylor expansion

$$\Bigl(x + \frac{y}{\beta x^{\beta-1}}\Bigr)^\beta \;=\; x^\beta\Bigl(1+\frac{y}{\beta x^\beta}\Bigr)^\beta \;=\; x^\beta\Bigl(1+\frac{y}{x^\beta} + \mathrm{o}(1/x^\beta)\Bigr)$$

which easily gives (4.2).

Note that if $\beta = 1$, $e(x) \sim 1$ is immediate from the memoryless property of the exponential distribution. If $\beta < 1$, F is more heavy-tailed than the exponential distribution and $e(x) \to \infty$, whereas if $\beta > 1$, F has a lighter tail and $e(x) \to 0$.

By continuity of F, we can compute b_n as solution of $1/n = \overline{F}(b_n) = \mathrm{e}^{-b_n^\beta}$ which gives $b_n = \log^{1/\beta} n$. We then get $a_n = e(b_n) = \log^{1/\beta-1} n/\beta$. ◇

Example 4.8 In the lognormal distribution with $\mu = 0$, $\sigma^2 = 1$, we take $e(t) = t/\log t$ as trial, motivated by Remark 4.6 and Example 3.7. With this choice,

$$\frac{\overline{F}\bigl(t+ye(t)\bigr)}{\overline{F}(t)} \;=\; \frac{\log t}{\log\bigl(t+yt/\log t)\bigr)}\exp\bigl\{(\log t)^2/2 - \bigl[\log\bigl(t+yt/\log t)\bigr)\bigr]^2/2\bigr\}$$

Using $\log\bigl(t+yt/\log t)\bigr) = \log t + \log(1+y/\log t) = \log t + y/\log t + \mathrm{o}(1/t)$, this becomes

$$(1+\mathrm{o}(1))\exp\bigl\{-\bigl(y/\log t + \mathrm{o}(1/\log t)\bigr)^2/2 - 2\cdot\log t\cdot\bigl(y/\log t + \mathrm{o}(1/\log t)\bigr)/2\bigr\}$$
$$= \exp\{-y+\mathrm{o}(1)\}$$

$$\frac{1}{n} \;=\; \frac{1}{\log b_n\sqrt{2\pi}}\exp\{-(\log b_n)^2/2\}$$

TBC ◇

Exercises

4.1. Assuming F has a density f, find the density of M_n and the joint density of M_n and second largest X, i.e. $\max\{X_k : k \le n,\, X_k < M_n\}$.

5 Heavy-Tailed Random Walks

We consider a random walk $S_n = X_1 + \cdots + X_n$ where $X, X_1, X_2, \ldots$ are i.i.d. and heavy-tailed with distribution F, and are interested in the tail of $M = \sup_{n=0,1,\ldots} S_n$. With light tails, one has under some conditions that $\mathbb{P}(M > x) \sim C\mathrm{e}^{-\gamma x}$ where $\gamma > 0$ solves $\mathbb{E}\mathrm{e}^{\gamma X} = 1$ and C is a constant. With heavy tails:

Theorem 5.1 *Consider a random walk such that $\mu = \mathbb{E}X < 0$ and that $\overline{F}(x) \sim B(x)$, $x \to \infty$, for some distribution B on $(0,\infty)$ which is long–tailed and satisfies $B_0 \in \mathcal{S}$. Then, writing $\overline{F}_I(x) = \int_x^\infty \overline{F}(y)\,\mathrm{d}y$, it holds that*

$$\mathbb{P}(M > x) \;\sim\; \frac{1}{|\mu|}\overline{F}_I(x), \quad x \to \infty; \tag{5.1}$$

Remark 5.2 The intuition behind Theorem 5.1 is the one big jum heuristics: the random walk evolves in its typical way until the big jump taking it above x. In this context, the meaning of 'typical' should be taken as drift at rate $\mu = -|\mu|$ so that $S_n \approx -n|\mu|$ A jump at time $n+1$ taking the random walk above x therefore should have size at least $x + n|\mu|$. This suggests that

$$\mathbb{P}(M > x) \;\approx\; \sum_{n=0}^{\infty} \overline{F}(x + n|\mu|) \;\approx\; \int_0^\infty \overline{F}(x + t|\mu|)\,\mathrm{d}t \;=\; \frac{1}{|\mu|}\int_x^\infty \overline{F}(y)\,\mathrm{d}y$$

as asserted. One often refers to $\overline{F}_I$ as the *integrated tail.* ◇

Remark 5.3 The conditions on B in Theorem 5.1 are easily verified to hold in the three main classes RV - lognormal - Weibull of subexponential distribution. Counterexamples going either way show, however, that these conditions are not equivalent to $B \in \mathcal{S}$. A useful sufficient condition is that $B \in \mathcal{S}^*$, where $\mathcal{S}^* \subset \mathcal{S}$ is the socalled *Klüppelberg class* defined by the requirement

$$\int_0^{x/2} \overline{B}(x-y)\overline{B}(y)\,\mathrm{d}y \;\sim\; \mu_B \overline{F}(x)$$

where $\mu_B = \int_0^\infty \overline{B}(y)\,\mathrm{d}y$ is the mean of B. ◇

To make the one big jum heuristics in Remark 5.2 rigorous requires quite some work and ideas of a different type coming from fluctuation theory. Define $\tau_+ = \inf\{n \geq 1 : S_n > 0\}$, $\tau_- = \inf\{n \geq 1 : S_n \leq 0\}$, and for $A \subseteq [0,\infty)$, $B \subseteq (-\infty, 0]$, let

$$G_+(A) \;=\; \mathbb{P}(S_{\tau_+} \in A;\, \tau_+ < \infty), \quad G_-(B) \;=\; \mathbb{P}(S_{\tau_-} \in B;\, \tau_- < \infty)$$

$$U_-(B) \;=\; \mathbb{E}\sum_{n=0}^{\tau_+} \mathbb{1}(S_n \in B)\,.$$

One refers to G_+ as the (strict) *ascending ladder height distribution*, to G_- as the (weak) *ascending ladder height distribution* and to U_- as the *pre-τ_+ occupation measure*. Alternatively, G_+, G_- are called the *Wiener-Hopf factors* of F because of the identity

$$1 - \varphi_F(s) \;=\; \big(1 - \varphi_{G_+}(s)\big)\big(1 - \varphi_{G_-}(s)\big) \tag{5.2}$$

where φ_F, φ_{G_+}, φ_{G_-} are the characteristic functions (Fourier transforms) of F, G_+, G_-.

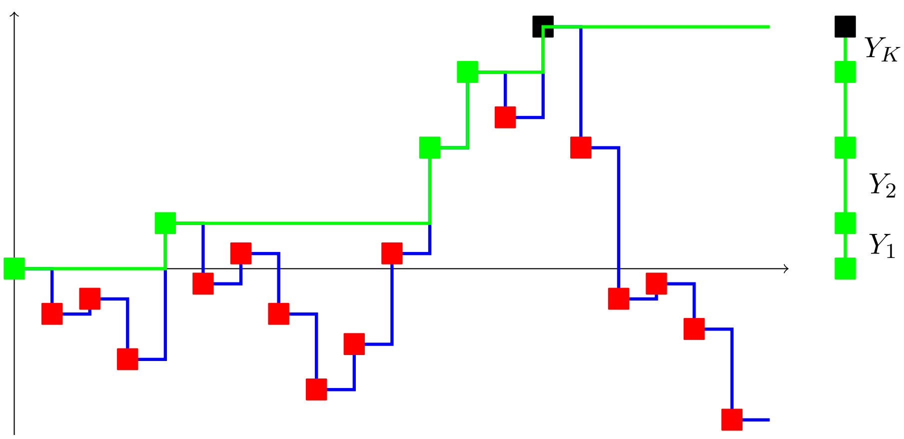

Figure 0.1: Random walk ■, ladder heights ■ and the maximum ■

Lemma 5.4 (a) $\mathbb{P}(M \in \cdot) = \sum_{n=0}^{\infty} G_+^{*n}$. *Equivalently,* $M \stackrel{\mathscr{D}}{=} Y_1 + \cdots + Y_K$ *where* K *is the number of ladder steps and* $Y_1, Y_2, \ldots$ *are i.i.d. with common distribution* $G = G_+/\|G_+\|$.
(b) K *is geometric with* $\mathbb{P}(K = k) = (1-\theta)\theta^k$ *where* $\theta = \|G_+\|$
(d) $U_- = \sum_{n=0}^{\infty} G_-^{*n}$

The starting point of the proof of Theorem 5.1 is the representation of M as a geometric sum of ladder heights, cf. Lemma 5.4 (a),(b) and Figure 0.1. After showing that the tail of G_+ is proportional to $\overline{F}_I(x)$, this gives that so is the tail of M, cf. Proposition 3.15 on random sums of subexpoentials. The remaining difficulty is that the parameter $\theta = \|G_+\|$ is not explicit so one can not directly identify the constant $1/|\mu|$ in 5.1.

Write $\overline{G}_+(x) = G_+(x, \infty) = \mathbb{P}(S_{\tau_+} > x,\, \tau_+ < \infty)$ and let μ_{G_-} be the mean of G_-.

Lemma 5.5 $\overline{G}_+(x) \sim \overline{F}_I(x)/|\mu_{G_-}|$, $x \to \infty$.

Proof: Clearly,

$$\overline{G}_+(x) = \int_{-\infty}^{0} \overline{F}(x-y)\, U_-(dy).$$

The heuristics is now that the contribution from the interval $(-N, 0]$ to the integral is $\mathrm{O}(\overline{F}(x))$ which by long–tailedness is $\mathrm{o}(\overline{F}_I(x))$, whereas for large y, $U_-(\mathrm{d}y)$

is close to Lebesgue measure on $(-\infty, 0]$ normalized by $|\mu_{G_-}|$ so that we should have

$$\overline{G}_+(x) \;\sim\; \frac{1}{|\mu_{G_-}|}\int_{-\infty}^{0} \overline{F}(x-y)\,\mathrm{d}y \;=\; \frac{1}{|\mu_{G_-}|}\overline{F}_I(x)\,.$$

We now make this precise. If G_- is nonlattice, then by Blackwell's renewal theorem $U_-(-n-1,-n] \to 1/|\mu_{G_-}|$. In the lattice case, we can assume that the span is 1 and then the same conclusion holds since then $U_-(-n-1,-n]$ is just the probability of a renewal at $-n$.

Given ϵ, choose N such that $\overline{F}(n-1)/\overline{F}(n) \le 1+\epsilon$ for $n \ge N$ (this is possible since B is long–tailed), and that $U_-(-n-1,-n] \le (1+\epsilon)/|\mu_{G_-}|$ for $n \ge N$. We then get

$$\begin{aligned}
&\overline{\lim_{x\to\infty}} \frac{\overline{G}_+(x)}{\overline{F}_I(x)} \\
&\le\; \overline{\lim_{x\to\infty}} \int_{-N}^{0} \frac{\overline{F}(x-y)}{\overline{F}_I(x)}\,U_-(\mathrm{d}y) \;+\; \overline{\lim_{x\to\infty}} \int_{-\infty}^{-N} \frac{\overline{F}(x-y)}{\overline{F}_I(x)}\,U_-(\mathrm{d}y) \\
&\le\; \overline{\lim_{x\to\infty}} \frac{\overline{F}(x)}{\overline{F}_I(x)}\,U_-(-N,0] \;+\; \overline{\lim_{x\to\infty}} \frac{1}{\overline{F}_I(x)} \sum_{n=N}^{\infty} \overline{F}(x+n) U_-(-n-1,-n] \\
&\le\; 0 \;+\; \overline{\lim_{x\to\infty}} \frac{1}{\overline{F}_I(x)} \frac{1+\epsilon}{|\mu_{G_-}|} \sum_{n=N}^{\infty} \overline{F}(x+n) \\
&\le\; \frac{(1+\epsilon)^2}{|\mu_{G_-}|}\, \overline{\lim_{x\to\infty}} \frac{1}{\overline{F}_I(x)} \int_{N}^{\infty} \overline{F}(x+y)\,\mathrm{d}y \\
&=\; \frac{(1+\epsilon)^2}{|\mu_{G_-}|}\, \overline{\lim_{x\to\infty}} \frac{\overline{F}_I(x+N)}{\overline{F}_I(x)} \;=\; \frac{(1+\epsilon)^2}{|\mu_{G_-}|}\,.
\end{aligned}$$

Here in the third step we used that $\overline{B}(x)/\overline{B}_0(x) \to 0$ (since B is long–tailed) and hence $\overline{F}(x)/\overline{F}_I(x) \to 0$, and in the last that $\overline{F}_I$ is asymptotically proportional to $B_0 \in \mathcal{S}$. Similarly,

$$\underline{\lim_{x\to\infty}} \frac{\overline{G}_+(x)}{\overline{F}_I(x)} \;\ge\; \frac{(1-\epsilon)^2}{|\mu_{G_-}|}\,.$$

Letting $\epsilon \downarrow 0$, the proof is complete. □

Proof of Theorem 5.1. By Lemma 5.5, $\mathbb{P}(Y_i > x) \sim \overline{F}_I(x)/(\theta|\mu_{G_-}|)$. Hence using dominated convergence, $M = Y_1 + \cdots + Y_K$ yields

$$\mathbb{P}(M > u) \;\sim\; \sum_{k=1}^{\infty} (1-\theta)\theta^k\, k\, \frac{\overline{F}_I(u)}{\theta|\mu_{G_-}|} \;=\; \frac{\overline{F}_I(u)}{(1-\theta)|\mu_{G_-}|}\,.$$

Now just observe that

$$(1-\theta)|\mu_{G_-}| = (1-\|G_+\|)|\mu_{G_-}| = |\mu| \tag{5.3}$$

as may be seen by differentiating (5.2) and letting $s = 0$. □

Now consider the cycle maximum

$$M_\sigma \;=\; \max_{0\le n<\sigma} S_n \qquad \text{where} \;\; \sigma \;=\; \inf\{n\ge 1 : S_n = 0\}\;.$$

Apart from its intrinsic interest, it is relevant for extreme value theory for the reflected random walk W_n defined by $W_0 = 0$, $W_{n+1} = (W_n + X_n)^+$. In fact, the following result is the key to show that that $\max_{0\le k\le n} W_n$ after a suitable normalization has a Fréchet limit distribution as $n\to\infty$ when B is regularly varying and a Gumbel limit distribution B is heavy–tailed Weibull or log–normal.

Theorem 5.6 *Consider a random walk such that $\mu = \mathbb{E}X < 0$ and that $\overline{F}(x) \sim B(x)$, $x\to\infty$, for some $B\in\mathcal{S}^*$. Then*

$$\mathbb{P}(M_\sigma > x) \;\sim\; \mathbb{E}\sigma\overline{F}(x), \quad x\to\infty. \tag{5.4}$$

Remarkably, this extens to general stopping times:

Theorem 5.7 *Consider a random walk such that $\mu = \mathbb{E}X < 0$ and that $\overline{F}(x) \sim B(x)$, $x\to\infty$, for some $B\in\mathcal{S}^*$. Then for any stopping time ω,*

$$\mathbb{P}\Bigl(\max_{n<\omega} > x\Bigr) \;=\sim\; \mathbb{E}\omega\overline{F}(x), \quad x\to\infty. \tag{5.5}$$

Notes and references Theorem 5.1 has a long history associated with the names of (in alphabetical order) von Bahr, Borovkov, Cohen, Pakes and Veraverbeke. These contributions are given a final form in Embrechts & Veraverbeke [11]. There are numerous recent analogues for more general models in queueing theory and insurance risk. The results (5.2), Lemma 5.4 and (5.3) from Wiener Hopf theory can be found in [14] or [2].

Theorem 5.6 is from [1] and Theorem 5.7 from [16].

6 Topics in Dependence and Risk Management

This sextion deals with extensions of the basic subexponential properties to dependent summands. This is a currently very active area, not least because of its importance for assessing quantities such as value-at risk (VaR) in regulation requirements in the banking and insurance industry, cf. [7, 8, 17].

We will often meet the concept of *comonotonicity.* In the case of the same marginal distribution of $X_1, \ldots, X_n$, this just means $X_1 = \cdots = X_n$. For different marginals $F_1, \ldots, F_n$, one possible characterization is

$$(X_1, \ldots, X_n) \stackrel{\mathscr{D}}{=} \bigl(F_1^{\leftarrow}(U), \ldots, F_n^{\leftarrow}(U)\bigr)$$

with $U \sim \text{uniform}(0,1)$ and (as stated) the same for $i = 1, \ldots, n$. For the sake of simplicity, we take $n = 2$ in much of the exposition.

Tail Dependence

In many contexts, one is interested in the extent to which large values of two r.v.s X, Y tend to occur together. A commonly used measure for this is the *tail dependence coefficient* $\lambda(Y|X)$ which for $X \stackrel{\mathscr{D}}{=} Y$ is defined by

$$\lambda(Y|X) \;=\; \lim_{x\to\infty} \mathbb{P}\bigl(Y > x \,\big|\, X > x\bigr)\,.$$

When $X \stackrel{\mathscr{D}}{=} Y$ fails, for example when X, Y are of a different order of magnitude, this definition leads to undesired features such as that symmetry (meaning $\lambda(Y|X) = \lambda(X|Y)$) may fail, that one may have $\lambda(aY|X) \neq \lambda(Y|X)$ etc., and therefore (taking the distributions of X, Y to be continuous for simplicity) the general definition is

$$\lambda(Y|X) \;=\; \lim_{u\uparrow 1} \mathbb{P}\bigl(F_Y(Y) > u \,\big|\, F_X(X) > u\bigr)\,. \qquad (6.1)$$

If in particular $\lambda(Y|X) = 0$ we call X, Y *tail independent.*

Recall that the copula of X, Y is the bivariate c.d.f.

$$C(u_1, u_2) \;=\; \mathbb{P}\bigl(F_X(X) \le u_1, F_Y(Y) \le u_2\bigr)\,.$$

Proposition 6.1 *The tail dependence index depends only on the copula $C_{X,Y}$ of X, Y and not the marginals. More precisely,*

$$\lambda(Y|X) \;=\; \lim_{u\uparrow 1} \frac{1 - 2u + C_{X,Y}(u,u)}{1-u}\,. \qquad (6.2)$$

Further, $\lambda(X|Y) = \lambda(Y|X)$.

Proof: Let $U_1 = F_X(X)$, $U_2 = F_Y(Y)$. Then

$$\begin{aligned}
\mathbb{P}(U_1 > u, U_2 > u) &= 1 - \mathbb{P}(U_1 \le u \text{ or } U_2 \le u) \\
&= 1 - \big[\mathbb{P}(U_1 \le u) + (U_2 \le u) - \mathbb{P}(U_1 \le u, U_2 \le u))\big] \\
&= 1 - 2u + C_{X,Y}(u,u)\,.
\end{aligned}$$

Inserting in (6.1), (6.2) follows immediately, and $\lambda(X|Y) = \lambda(Y|X)$ is clear from (6.2) and $C_{X,Y} = C_{Y,X}$ on the diagonal. □

Proposition 6.2 *Assume that (X,Y) has a bivariate Gaussian distribution with correlation $\rho \in (-1,1)$. Then $\lambda(Y|X) = 0$.*

Proof: Since $\lambda(Y|X)$ depends only on the copula, we may assume that both marginals are $\mathrm{N}(0,1)$. Given $X > x$, X is close to x and the conditional distribution of Y given $X = x$ is $\mathrm{N}(\rho x, 1-\rho^2)$. Thus up to the first order of approximation

$$\lambda(Y|X) \;\approx\; \lim_{x\to\infty} \mathbb{P}(Y > x | X = x) \;=\; \lim_{x\to\infty} \overline{\Phi}\big((1-\rho)x/(1-\rho^2)^{1/2}\big) \;=\; 0.$$

The rigorous verification is left to the reader. □

In view of the properties of the conditional distribution of Y given $X = x$ that were used, Proposition 6.2 is somewhat counterintuitive and indicates that $\lambda(Y|X)$ is only a quite rough measure of tail dependence.

Proposition 6.3 *Let $\boldsymbol{X} = (X_1\, X_2)$ be multivariate regularly varying with parameters α, L, μ so that*

$$\mathbb{P}\big(|X_1| + |X_2| > x\big) = L(x)/x^\alpha\,, \quad \mathbb{P}(X_i > x) \sim w_i L(x)/x^\alpha$$

*for $i = 1,2$ where $w_i = \displaystyle\int_{\mathcal{B}_1} \theta_i^{+\alpha}\,\mu(\mathrm{d}\theta_1, \mathrm{d}\theta_2)$, cf. Proposition **??**. Then*

$$\lambda(X_1|X_2) \;= \lambda(X_2|X_1) \;= \int_{\mathcal{B}_1^+} \big[\min\big(\theta_1^\alpha/w_1, \theta_2^\alpha/w_2\big)\big]^\alpha \mu(\mathrm{d}\theta_1, \mathrm{d}\theta_2)$$

where $\mathcal{B}_1^+ \;=\; \big\{(\theta_1,\theta_2) \in \mathcal{B}_1 :\, \theta_1 > 0, \theta_2 > 0\big\}$.

Bounds

In this and the next subsections, we consider the question of giving estimates for $\mathbb{P}(S > x)$ where $S = X_1 + \cdots + X_d$ with the components of the vector $(X_1 \ldots X_d)$ possibly dependent and having marginals $F_1, \ldots, F_d$. I.e., $F_i(x) = \mathbb{P}(X_i \le x)$. For simplicity, we will assume that the F_i are continuous.

Rüschendorf [26] refers to the following inequality as one of the 'standard bounds'.

Proposition 6.4

$$\mathbb{P}\big(X_1 + \cdots + X_d > x\big) \;\le\; \inf_{x_1+\cdots+x_d=x} \big(\overline{F}_1(x_1) + \cdots + \overline{F}_d(x_d)\big)\,.$$

Proof: If $S > x = x_1 + \cdots + x_d$, then at least one X_i must exceed x_i (otherwise $S \le x$). Hence

$$\mathbb{P}(S > x) \le \; \mathbb{P}\Big(\bigcup_{i=1}^{n} \{X_i > x_i\}\Big) \;\le\; \sum_{i=1}^{d} \mathbb{P}(X_i > x_i) \;=\; \sum_{i=1}^{d} \overline{F}(x_i)\,.$$

Taking the inf gives the result. □

It is remarkable that despite the simplicity of its proof, the inequality in Proposition 6.4 is sharp, i.e. attained for a certain dependence structure. We verify this for $d = 2$.

Proposition 6.5 *Given F_1, F_2 and x, there exists a random vector (X_1, X_2) with marginals F_1, F_2 such that*

$$\mathbb{P}\big(X_1 + X_2 > x\big) \;=\; \inf_{x_1+x_2=x} \big(\overline{F}_1(x_1) + \overline{F}_2(x_2)\big) \tag{6.3}$$

Proof: Denote the r.h.s. of (6.3) by p^*. Then for x_1 such that $\overline{F}_1(x_1) \le p^*$

$$\begin{aligned} x_1 + \overline{F}_2^{\leftarrow}\big(p^* - \overline{F}_1(x_1)\big) \;&\ge\; x_1 + \overline{F}_2^{\leftarrow}\big(\overline{F}_2(x - x_1)\big) \\ &\ge\; x_1 + (x - x_1) \;=\; x \end{aligned} \tag{6.4}$$

where in the first step we used that $p^* \le \overline{F}_1(x_1) + \overline{F}_2(x - x_1)$ and that $\overline{F}_2^{\leftarrow}$ is non–increasing. Now let U_1 be uniform$(0, 1)$ and let $U_2 = p^* - U_1$ for

$U_1 \le p^*$, $U_2 = U_1$ for $U_1 > p^*$. Then U_2 is again uniform$(0,1)$ since clearly $\mathbb{P}(U_2 \in A) = \mathbb{P}(U_1 \in A)$ for $A \subseteq (p^*, 1)$, whereas also

$$\begin{aligned}\mathbb{P}(U_2 \in A) &= \mathbb{P}(U_2 \in A,\, U_2 \le p^*) \;=\; \mathbb{P}(p^* - U_1 \in A,\, U_1 \le p^*)\\ &=\; \mathbb{P}(U_1 \in A,\, U_1 \le p^*) \;=\; \mathbb{P}(U_1 \in A)\end{aligned}$$

for $A \subseteq (0, p^*]$. Hence $X_1 = \overline{F}_1^{\leftarrow}(U_1)$ and $X_2 = \overline{F}_1^{\leftarrow}(U_2)$ have marginals F_1, F_2. On the set $U_1 \le p^*$, we have

$$U_1 = \overline{F}_1(X_1) \le p^*\,, \quad X_2 = \overline{F}_2^{\leftarrow}\big(p^* - \overline{F}_1(X_1)\big)\,,$$

and hence $X_1 + X_2 \ge x$ by (6.4). Hence $\mathbb{P}(X_1 + X_2 > x) \ge \mathbb{P}(U_1 \le p^*) = p^*$, and by the definition of p^* equality must hold. □

Proposition 6.6 *Assume $F_1 = F_2 = F$ and that F has a density f that is non-increasing on some interval $[a, \infty)$. Then $\mathbb{P}(X_1 + X_2 > x) \le 2\overline{F}(x/2)$ for all large x.*

Proof: The assumption ensures that $\overline{F}$ is convex on $[a, \infty)$. Thus for $x_1 + x_2 = x$, $x_2 \ge x_1 \ge a$ we have $\overline{F}(x_1) + \overline{F}(x_2) \ge 2\overline{F}(x/2)$. For $x_1 + x_2 = x$, $x_1 < a$ we have $\overline{F}(x_1) + \overline{F}(x/2) \ge \overline{F}(a)$. Hence the inf in Proposition 6.4 equals $\overline{F}(x/2)$ when x is so large that $2\overline{F}(x/2) \le \overline{F}(a)$. □

Remark 6.7 The value $2\overline{F}(x/2)$ is to be compared with the smaller $\overline{F}(x/2)$ one gets in the comonotonic case, and gives a key example that comonotonicity is not necessarily the dependence structure leading to the maximal risk.

A further illustration of this comes from F being regular varying, $\overline{F}(x) = L(x)/x^\alpha$. Here $\mathbb{P}(X_1 + X_2 > x) = \overline{F}(x/2) \sim 2^\alpha L(x)/x^\alpha$ in the comonotonic case, whereas if X_1, X_2 are independent, then by Proposition 3.4

$$\mathbb{P}(X_1 + X_2 > x) \;\sim\; 2\overline{F}(x) \;=\; 2L(x)/x^\alpha\,.$$

Thus if $\alpha < 1$ so that $2^\alpha < 2$, the asymptotic order of $\mathbb{P}(X_1 + X_2 > x)$ is larger with independence than with comonotonicity, again contrary to the naive first guess.

With independence and regular variation, we have $\mathbb{P}(X_1 + X_2 > x) \sim 2\overline{F}(x)$. This is of the same rough order of magnitude $1/x^\alpha$ as the standard bound and as in comonotonic case. With lighter tails than $1/x$, the situation is, however, typically the opposite. For example, if F is standard exponential, then $\mathbb{P}(X_1+X_2>x) \sim x\mathrm{e}^{-x}$ in the independent case, whereas $\mathbb{P}(X_1+X_2>x) = \overline{F}(x/2) = \mathrm{e}^{-x/2}$ is of larger magnitude in the comonotonic case (though still not attaining the upper bound $2\mathrm{e}^{-x/2}$!). See also Proposition 6.9 below. ◇

Asymptotics

For simplicity, we will concentrate on the case $d=2$ and positive r.v.s X_1, X_2 with continuous marginals F_1, F_2 such that F_1 is more heavy-tailed than F_2, more precisely in the sense that

$$c \;=\; \lim_{x\to\infty} \frac{\overline{F}_2(x)}{\overline{F}_1(x)} \tag{6.5}$$

exists and is in $[0,1]$. Tail dependence as measured by $\lambda(Y|X)$ defined by (6.1) is a less relevant concept since it is scale-free whereas $\mathbb{P}(X_1+X_2>x)$ is changed if one of the r.v.s is scaled. We shall therefore instead work with

$$\widehat{\lambda} \;=\; \lim_{x\to\infty} \mathbb{P}(X_2>x|X_1>x)\,. \tag{6.6}$$

A key question is asymptotic comparison of the tails of S and X_1: does $\mathbb{P}(S>x)/\overline{F}_1(x)$ go to 1, to a constant in $(1,\infty)$, or to ∞? A relatively simple result in this frame is the following:

Proposition 6.8 *Assume that F_1 is regularly varying and that $\widehat{\lambda}=0$. Then $\mathbb{P}(X_1+X_2>x) \sim (1+c)\overline{F}_1(x)$.*

This is the same conclusion as in Proposition **??**.3.12, with independence relaxed to $\widehat{\lambda}=0$. The proof is given later.

Another simple case is the following:

Proposition 6.9 *Assume that F_1 is in MDA(Gumbel) and that*

$$m \;=\; \inf_{a>0}\liminf_{x\to\infty} \mathbb{P}\bigl(X_2>ae(x)\,\big|\,X_1>x\bigr) \;>\; 0 \tag{6.7}$$

where m is the mean excess function. Then $\mathbb{P}(S>x)/\overline{F}_1(x)\to\infty$.

Remark 6.10 The result covers a fairly broad spectrum of situations with lighter than regularly varying marginals. In fact, MDA(Gumbel) gave the main examples of such marginals in Section 4 and one had $e(x)=cx/\log x$

for the lognormal case and typically $e(x) = cx^{1-\beta}$ with $\beta > 0$ otherwise. When $0 < \beta < 1$, (6.7) basically says that X_2 may grow with X_1 but at a possibly slower rate, and when $\beta > 1$, (6.7) roughly can only fail if X_2 becomes small as X_1 becomes large, a quite pathological type of behavior. $\diamond$

Proof of Proposition 6.9. The assumption $F_1 \in$ MDA(Gumbel) implies that $e(x)$ is self-neglecting, i.e. that $e\big(x + ae(x)\big)/e(x) \to 1$ for all $a \in \mathbb{R}$, cf. Exercise 6.1. Thus for a given $a > 0$, we have $e(x) \le 2e\big(x - ae(x)\big)$ for all large x. Hence

$$\begin{aligned}\liminf_{x\to\infty} \frac{\mathbb{P}(S > x)}{\overline{F}_1(x)} \;&\ge\; \liminf_{x\to\infty} \frac{\mathbb{P}\big(X_1 > x - ae(x),\ X_2 > ae(x)\big)}{\overline{F}_1(x)} \\ &= \liminf_{x\to\infty} \frac{\overline{F}_1\big(x - ae(x)\big)\mathbb{P}\big(X_2 > ae(x) \,\big|\, X_1 > x - ae(x)\big)}{\overline{F}_1(x)} \\ &\ge \mathrm{e}^a \liminf_{x\to\infty} \mathbb{P}\big(X_2 > 2ae(x - ae(x)) \,\big|\, X_1 > x - ae(x)\big) \;\ge\; m\mathrm{e}^a .\end{aligned}$$

Letting $a \to \infty$ completes the proof. $\square$

In the regularly varying case (i.e., $F_1 \in$MDA(Fréchet)) it can not occur that $\mathbb{P}(S > x)/\overline{F}_1(x) \to \infty$. More precisely:

Proposition 6.11 *Assume that F_1 is regularly varying, i.e. $\overline{F}_1(x) = L(x)/\alpha$. Then*

$$\limsup_{x\to\infty} \frac{\mathbb{P}(S > x)}{\overline{F}_1(x)} \;\le\; \begin{cases} \Big(\widehat{\lambda}^{1/(\alpha+1)} + (1 + c - 2\widehat{\lambda})^{1/(\alpha+1)}\Big)^{\alpha+1} & 0 \le \widehat{\lambda} \le (1+c)/3 \\ 2^\alpha(1 + c - \widehat{\lambda}) & (1+c)/3 \le \widehat{\lambda} \le 1 \end{cases}$$

Proof: A straightforward extension of the proof of Proposition **??**,3.4, to be given in a moment below, gives

$$\limsup_{x\to\infty} \frac{\mathbb{P}(S > x)}{\overline{F}_1(x)} \;\le\; \frac{1 + c - 2\widehat{\lambda}}{(1-\delta)^\alpha} + \frac{\widehat{\lambda}}{\delta^\alpha} \tag{6.8}$$

for $0 < \delta < 1/2$. Here the first term on the r.h.s. increases with δ and the second decreases. Easy calculus shows that the minimizer δ^* of the r.h.s. is given by

$$\delta^* \;=\; \begin{cases} \dfrac{1}{1 + \big((1+c)/3 - 2\big)^{1/(\alpha+1)}} & 0 \le \widehat{\lambda} \le (1+c)/3 \\ 1/2 & (1+c)/3 \le \widehat{\lambda} \le 1 \end{cases}$$

Inserting this in (6.8) gives the result.

For the proof of (6.8), note as in the proof of Proposition **??**.3.4 that $\{S > x\} \subseteq A_1 \cup A_2 \cup A_3$ where

$$A_1 = \bigl\{X_1 > (1-\delta)x\bigr\}\,,\ A_2 = \bigl\{X_2 > (1-\delta)x\bigr\}\,,\ A_3 = \bigl\{X_1 > \delta x, X_2 > \delta x\bigr\}\,.$$

Here

$$\mathbb{P}\bigl((A_1 \cup A_2) \cap A_3\bigr) \ \geq\ \mathbb{P}(A_2 \cap A_3) \ \geq\ \mathbb{P}(A_2 \cap A_1)$$

whenever $0 < \delta < 1/2$. Hence

$$\begin{aligned}\mathbb{P}(S > x) &= \mathbb{P}(A_1 \cup A_2) + \mathbb{P}A_3 - \mathbb{P}\bigl((A_1 \cup A_2) \cap A_3\bigr)\\ &\geq \mathbb{P}A_1 + \mathbb{P}A_2 + \mathbb{P}A_3 - 2\mathbb{P}(A_1 \cap A_2)\\ &= \overline{F}_1\bigl((1-\delta)x\bigr) + \overline{F}_2\bigl((1-\delta)x\bigr) \ +\ \mathbb{P}(X_1 > \delta x, X_2 > \delta x)\\ &\quad - 2\mathbb{P}\bigl(X_1 > (1-\delta)x, X_2 > (1-\delta)x\bigr)\end{aligned}$$

so that the lim sup in the Proposition is bounded by the lim sup of

$$(1-2\widehat{\lambda})\frac{\overline{F}_1\bigl((1-\delta)x\bigr)}{\overline{F}_1(x)} + \frac{\overline{F}_2\bigl((1-\delta)x\bigr)}{\overline{F}_1(x)} + \frac{\overline{F}_1(\delta x)}{\overline{F}_1(x)}\mathbb{P}(X_2 > \delta x \,|\, X_1 > \delta x)$$

which equals

$$(1-2\widehat{\lambda})\frac{1}{(1-\delta)^\alpha} + \frac{c}{(1-\delta)^\alpha} + \frac{\widehat{\lambda}}{\delta^\alpha}$$

as asserted. □

Proof of Proposition 6.8. That $1 + c$ is an asymptotic upper bound for $\mathbb{P}(S > x)/\overline{F}_1(x)$ follows immediately by inserting $\widehat{\lambda} = 0$ in the bound of Proposition 6.11. That it is also a lower bound follows from

$$\begin{aligned}\mathbb{P}(S > x) &\geq \mathbb{P}\bigl(\max(X_1, X_2) > x\bigr)\\ &= \mathbb{P}(X_1 > x) + \mathbb{P}(X_2 > x) - \mathbb{P}(X_1 > x, X_2 > x)\\ &= \mathbb{P}(X_1 > x) + \mathbb{P}(X_2 > x) - \mathbb{P}(X_1 > x)\mathbb{P}(X_2 > x | X_1 > x)\\ &\sim \overline{F}_1(x)(1 + c - \widehat{\lambda}) \ =\ \overline{F}_1(x)(1+c)\,.\end{aligned}$$

□

Exercises

6.1. Let F be a distribution and $e(x) > 0$ a function such that $\overline{F}\bigl(x + ae(x)\bigr) \sim \mathrm{e}^{-a}\overline{F}(x)$ for all $a \in \mathbb{R}$. Show that $e\bigl(x + ae(x)\bigr) \sim e(x)$.

7 Appendix: Tails of Sums of Light-Tailed Random Variables

We consider here the asymptotics of $\overline{F^{*n}}(x) = \mathbb{P}(S_n > x)$ and the corresponding density $f^{*n}(x)$ for an i.i.d. light-tailed sum as x goes to ∞.

The most traditional set-up for this problem is to let x depend on n as $x = ny$ where $y > \mathbb{E}X$; the LLN then implies $\mathbb{P}(S_n > ny) \to 0$. Define $\kappa(\theta) = \log \mathbb{E}\mathrm{e}^{\theta X}$ as the cumulant function of X.

Theorem 7.1 *Let $y > \kappa'(0) = \mathbb{E}X_1$, assume that the equation $\kappa'(\theta) = y$ has a solution $\theta = \theta(y)$, and define $\kappa^*(y) = \theta y - \kappa(\theta)$. Then*

$$\begin{aligned} \mathbb{P}(S_n > ny) &\leq \mathrm{e}^{-n\kappa^*(y)}, & (7.1)\\ \frac{1}{n}\log \mathbb{P}(S_n > ny) &\to -\kappa^*(y), \quad n \to \infty, & (7.2)\\ \mathbb{P}(S_n > ny) &\sim \frac{1}{\theta\sqrt{2\pi\sigma_\theta^2 n}}\mathrm{e}^{-n\kappa^*(y)}, \quad n \to \infty, & (7.3) \end{aligned}$$

provided in addition for (7.2) that $\sigma_\theta^2 = \kappa''(\theta) < \infty$ and for (7.3) that $|\kappa'''(\theta)| < \infty$ and that F satisfies Cramér's condition (C).

The inequality (7.1) goes under the name of the *Chernoff bound*, (7.3) is the *saddlepoint approximation*, and (7.2) is often referred to as *Cramér' large deviations theorem*, though this also covers some extensions to $\mathbb{P}(S_n \in A)$ for more general sets than half-lines.

The proof of Theorem 7.1 will not be given here, since it can be found in several standard texts (e.g. [18], [23] or [2]). In the rest of the section, we consider instead the somewhat less standard topic of the asymptotics of $\overline{F^{*n}}(x) = \mathbb{P}(S_n > x)$ as x goes to ∞ with n fixed. This is similar to the set-up for subexponential limit theory developed in Section 3. The theory which has been developed applies to distributions with a density or tail close to e^{-cx^β}. .

The cases $\beta = 1$ and $\beta > 1$ are intrinsically different (of course, $\beta < 1$ leads to a heavy tail). The difference may be understood from by looking at two specific examples with $n = 2$, the exponential distribution where $\beta = 1$ and the normal distribution where $\beta = 2$ by Mill's ratio in (**??**). If F is exponential, Example 3.2 shows that if $X_1 + X_2 > x$, then approximately X_1, X_2 are both uniform on between 0 and x, with the joint distribution concentrating on the line $x_1 + x_2 = x$. On the other hand, if F is standard normal and $X_1 + X_2 > x$, then X_1, X_2 are both approximately equal to $x/2$. This follows from noting that $X_1 + X_2 \approx x$ when $X_1 + X_2 > x$ (cf. Remark **??**) and that the marginal distributions of X_1 and X_2 given $X_1 + X_2 = x$ are both normal $(x/2, 1/2)$.

For ease of exposition, we make some assumptions that are simplifying but not all crucial. In particular, we take F to be concentrated on $(0, \infty)$ and having a density $f(x)$, and exemplify 'close to' by allowing a modifying regularly varying prefactor to e^{-cx^β}. For the density, this means

$$f(x) \;\sim\; L(x)x^\gamma \mathrm{e}^{-cx^\beta}\,, x > 0, \tag{7.4}$$

with $L(\cdot)$ slowly varying. However, in the rigorous proofs we only take $n = 2$ and $L(x) \equiv d$ will be constant. We shall need the following simple lemma (cf. Exercise 7.1):

Lemma 7.2 *If (7.4) holds with $\beta \geq 1$, then $\overline{F}(x) = \mathbb{P}(X > x) \sim L(x)x^\gamma \mathrm{e}^{-cx^\beta}/c$.*

The case $\beta = 1$ of a close-to-exponential tail is easy:

Proposition 7.3 *Assume $f(x) \sim L(x)x^{\alpha-1}\mathrm{e}^{-cx}$ with $\alpha > 0$. Then the density and the tail of an i.i.d. sum satisfy*

$$f^{*n}(x) \;\sim\; \frac{\Gamma(\alpha)^n}{\Gamma(n\alpha)} L(x)^n x^{n\alpha-1}\mathrm{e}^{-cx}\,, \tag{7.5}$$

$$\overline{F^{*n}}(x) \;=\; \mathbb{P}(S_n > x) \;\sim\; \frac{\Gamma(\alpha)^n}{c\Gamma(n\alpha)} L(x)^n x^{n\alpha-1}\mathrm{e}^{-cx}\,. \tag{7.6}$$

Proof: By Lemma 7.2, it suffices to show (7.3). We give the proof for the case $n = 2$ only where

$$f^{*2}(x) \;=\; \int_0^x f(z)f(x-z)\,\mathrm{d}z \tag{7.7}$$

and also take $L(x) \equiv d$ for simplicity. Denote by $\widetilde{f}^{*2}(x)$ the rh.s. of (7.5) and write $\log_2 x = \log\log x$. Then

$$\mathbb{P}(X_1 > x - \log_2 x) \;\sim\; dx^{\alpha-1}(\log x)^c \mathrm{e}^{-cx} \;=\; \mathrm{o}\big(\widetilde{f}^{*2}(x)\big)$$

by Lemma 7.2 since $\alpha > 0$, so the contribution to $f^{*2}(x)$ from the event $X_1 > x - \log_2 x$ can be neglected. The same is true for the event $X_1 < \log_2 x$

since if that occurs, we can only have $X_1 + X_2 = x$ if $X_2 > x - \log_2 x$. It thus suffices to show that

$$\int_{\log_2 x}^{x-\log_2 x} f(z)f(x-z)\,\mathrm{d}z \;\sim\; \widetilde{f}^{*2}(x)\,.$$

Since here $z \ge \log_2 x$ and $x - z \ge \log_2 x$ and $\log_2 x \to \infty$, the asymptotics of the l.h.s. is the same as that of

$$\int_{\log_2 x}^{x-\log_2 x} d^2 z^{\alpha-1}\mathrm{e}^{-cz}(x-z)^{\alpha-1}\mathrm{e}^{-c(x-z)}\,\mathrm{d}z.$$

But this equals

$$d^2\mathrm{e}^{-cx}x^{2\alpha-2}\int_{\log_2 x/x}^{1-\log_2 x/x} y^{\alpha-1}(1-y)^{\alpha-1}\,x\,\mathrm{d}y$$
$$\sim\; d^2x^{2\alpha-1}\mathrm{e}^{-cx}\mathrm{Beta}(\alpha,\alpha) \;=\; d^2x^{2\alpha-1}\mathrm{e}^{-cx}\frac{\Gamma(\alpha)^2}{\Gamma(2\alpha)} \;=\; \widetilde{f}^{*2}(x)\,.$$

□

Tails that are lighter than exponential present more substantial difficulties. The main example is $\overline{F}(x)$ being close to a Weibull tail e^{-cx^β} with $\beta > 1$, and we give the rigorous proof of some results in that setting in Proposition 7.4 below. We give, however, first some intuition for a slightly more general case, $f(x) = h(x)\mathrm{e}^{-\Lambda(x)}$, with h in some sense much less variable than Λ. Assume that Λ is such that the same principle applies as for the normal distribution, that if $X_1 + X_2 > x$, then X_1, X_2 are both close to $x/2$. Then the main contribution to the convolution integral (7.7) comes from a neighbourhood of $z = x/2$ where

$$\Lambda(z) \;\approx\; \Lambda(x/2) + \Lambda'(x/2)(z-x/2) + \frac{\Lambda''(x/2)}{2}(z-x/2)^2$$
$$\Lambda(x-z) \;\approx\; \Lambda(x/2) + \Lambda'(x/2)(x/2-z) + \frac{\Lambda''(x/2)}{2}(x/2-z)^2 \qquad (7.8)$$

Also, the assumption on h indicates that h is roughly constant in this neighbourhood. Noting that the $\Lambda'(x/2)$ terms in (7.8) cancel when adding, the expression (7.7) for $f^{*2}(x)$ therefore approximately should equal

$$h(x/2)^2\int \exp\bigl\{-2\Lambda(x/2) - \Lambda''(x/2)(z-x/2)^2\bigr\}$$
$$=\; h(x/2)^2\exp\bigl\{-2\Lambda(x/2)\bigr\}\sqrt{\frac{\pi}{\Lambda''(x/2)}}$$

Here is a precise result in that direction:

Proposition 7.4 *Assume that* $f(x) \sim L(x)x^{\alpha+\beta-1}\mathrm{e}^{-cx^\beta}$ *with* $\beta > 1$*. Then the density and the tail of an i.i.d. sum satisfy*

$$f^{*n}(x) \;\sim\; L(x/n)^n k_n x^{\alpha_n+\beta-1}\mathrm{e}^{-c_n x^\beta}\,, \tag{7.9}$$

$$\overline{F^{*n}}(x) \;=\; \mathbb{P}(S_n > x) \;\sim\; L(x/n)^n k_n x^{\alpha_n}\mathrm{e}^{-c_n x^\beta}\,, \tag{7.10}$$

where $\alpha_n = n\alpha + (n+1)\beta/2$ *and*

$$k_n = \Big[\frac{\pi}{\beta(\beta-1)c}\Big]^{(n-1)/2} n^{-n\alpha-(n+1)\beta/2+1}. \tag{7.11}$$

Proof: We give again the proof for the case $n = 2$ only and take $L(x) \equiv d$, $c = 1$. Lemma 7.2 gives in particular that $\overline{F}(x) \sim dx^\alpha \mathrm{e}^{-x^\beta}/\beta$ and also shows that (7.10) follows from (7.9), so it suffices to show (7.9).

Denote by

$$\widetilde{f}^{*2}(x) \;=\; d^2\sqrt{\frac{\pi}{\beta(\beta-1)}}\,\frac{1}{2^{2\alpha+3\beta/2-1}}x^{2\alpha+3\beta/2-1}\mathrm{e}^{-x^\beta/2^{\beta-1}}$$

the rh.s. of (7.9) and choose $1/2 < a < 1$ such that $a^\beta > c_2 = 1/2^{\beta-1}$ (this is possible since $\beta > 1$). Then

$$\mathbb{P}(X_1 > ax) \;\sim\; d(ax)^\alpha \mathrm{e}^{-a^\beta x^\beta}/\beta \;=\; \mathrm{o}\big(\widetilde{f}^{*2}(x)\big)$$

by Lemma 7.2. The same argument as in the proof of Proposition 7.3 therefore gives that the contribution to $f^{*2}(x)$ from the events $X_1 > ax$ and $X_1 < (1-a)x$ be neglected. It thus suffices to show that

$$\int_{1-a)x}^{ax} f(z)f(x-z)\,\mathrm{d}z \;\sim\; \widetilde{f}^{*2}(x)\,. \tag{7.12}$$

Since $z \geq (1-a)x$ and $x - z \geq ax$ in (7.12) are both bounded below by $(1-a)x$ which goes to ∞, the asymptotics of the l.h.s. is the same as that of

$$\int_{(1-a)x}^{ax} d^2 z^{\alpha+\beta-1}(x-z)^{\alpha+\beta-1}\exp\big\{-z^\beta - (x-z)^\beta\big\}\,\mathrm{d}z. \tag{7.13}$$

We now use the substitution $z \to v$ given by

$$z \;=\; \frac{x}{2}(1 + v/x^{\beta/2}) \;\text{ or equivalently } x - z \;=\; \frac{x}{2}(1 - v/x^{\beta/2})$$

together with the expansion

$$(1+h)^\beta \;=\; 1 + h\beta + \frac{h^2}{2}\beta(\beta-1)\omega(h) \;=\; 1 + h\beta + \frac{h^2 2^\beta}{4\sigma^2}\omega(h)$$

where $1/\sigma^2 \,=\, 2^{1-\beta}\beta(\beta-1)$, $\omega(h) \to 1$ as $h \to 0$ and $\omega(h)$ has the form $(1+h^*)^{\beta-2}$. for some h^* between 0 and h. We can then write $z^\beta + (x-z)^\beta$ as

$$\begin{aligned}
&\frac{x^\beta}{2^\beta}\Big(1 + \frac{v}{x^{\beta/2}} + \frac{v^2 2^\beta}{4\sigma^2 x^\beta}\omega(v/x^{\beta/2})\Big) + \frac{x^\beta}{2^\beta}\Big(1 - \frac{v}{x^{\beta/2}} + \frac{v^2 2^\beta}{4\sigma^2 x^\beta}\omega(v/x^{\beta/2})\Big) \\
&\quad = \;\frac{x^\beta}{2^{\beta-1}} + \frac{v^2}{2\sigma^2}\delta(v/x^{\beta/2})\,,
\end{aligned}$$

where $\delta(h) = (\omega(h) + \omega(-h))/2$, so that (7.13) takes the form

$$\begin{aligned}
&\frac{d^2}{2^{2\alpha+2\beta-2}} x^{2\alpha+2\beta-2}\exp\Big\{-\frac{x^\beta}{2^{\beta-1}}\Big\} \\
&\quad\cdot \int_{(1/2-a)x^{\beta/2}}^{(a-1/2)x^{\beta/2}} \Big(1 - \frac{v^2}{x^\beta}\Big)^{\alpha+\beta-1} \exp\Big\{-\frac{v^2}{2\sigma^2}\delta(v/x^{\beta/2})\Big\}\,\frac{1}{2}x^{1-\beta/2}\,\mathrm{d}v \quad (7.14)
\end{aligned}$$

When z varies between $(1-a)x$ and ax, $v/x^{\beta/2}$ varies between $1/2 - a$ and $a - 1/2$. But the choice $a < 1$ then ensures that $\delta(v/x^{\beta/2})$ is bounded in this range and goes to 1 as $x \to \infty$. Similar estimates for the powers allows us to use dominated convergence to conclude that the asymptotics of (7.14) and hence (7.13) is given by

$$\begin{aligned}
&\frac{d^2}{2^{2\alpha+2\beta-1}} x^{2\alpha+3\beta/2-1}\exp\Big\{-\frac{x^\beta}{2^{\beta-1}}\Big\}\int_{-\infty}^{\infty}\exp\Big\{-\frac{v^2}{2\sigma^2}\Big\}\,\mathrm{d}v \\
&\quad = \;\frac{d^2\sqrt{2\pi\sigma^2}}{2^{2\alpha+2\beta-1}} x^{2\alpha+3\beta/2-1}\exp\Big\{-\frac{x^\beta}{2^{\beta-1}}\Big\} \\
&\quad = \;\frac{d^2}{2^{2\alpha+3\beta/2-1}}\Big(\frac{\pi\,2^\beta}{\beta(\beta-1)}\Big)^{1/2} x^{2\alpha+3\beta/2-1}\exp\Big\{-\frac{x^\beta}{2^{\beta-1}}\Big\}
\end{aligned}$$

which is the same as $\widetilde{f}^{*2}(x)$. □

Exercises

7.1. Prove Lemma 7.2.

7.2. Verify that Proposition 7.4 gives the correct result if F is standard normal.

Notes and references The key reference for Proposition 7.4 is [6]. A somewhat more elementary survey along the present lines is in [4]. The extension from $n = 2$ to $n > 2$ can be carried out by first deriving analogues of Propositions 7.3, 7.4 with different c, L, α, and then using induction. Keeping track of the constants is messy, as may be guessed from the complicated expressions (7.11) ! The more advanced argument in [6] uses convex analysis and some non-standard central limit theory. Section 8 of [4] also contains some references for the more elementary Proposition 7.4.

Bibliography

[1] S. Asmussen (1998) Subexponential asymptotics for stochastic processes: extremal behaviour, stationary distributions and first passage probabilities. *Ann. Appl. Probab.* **8**, 354–374.

[2] S. Asmussen (2003) *Applied Probability and Queues* (2nd ed.). Springer.

[3] S. Asmussen & H. Albrecher (2010) *Ruin Probabilities* (2nd ed.). World Scientific.

[4] S. Asmussen, E. Hashorva, P. Laub & T. Taimre (2017) Tail asymptotics for light-tailed Weibull-like sums. *Probability and Mathematical Statistics* **37**, 235–256.

[5] S. Asmussen & M. Steffensen (2018/19) *Risk and Insurance. A Graduate Text.* Springer.

[6] A.A. Balkema, C. Klüppelberg & S.I. Resnick (1993) Densities with Gaussian tails. *Proc. London Math. Soc.* **66**, 568–588.

[7] Basel II (2004) *Basel II: Revised International Capital Framework.* Bank for International Settlements.

[8] BCBS (2012) *Consultative Document October 2013. Fundamental Review of the Trading Book.* Basel Committee on Banking Supervision.

[9] N. Bingham, C. Goldie & J.L. Teugels (1989) *Regular Variation.* Cambridge University Press.

[10] P. Embrechs, C. Klüppelberg & T. Mikosch (1997) *Modeling Extreme Events for Insurance and Finance.* Springer-Verlag.

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

[12] F. Esscher (1932) On the probability function in the collective theory of risk. *Skand. Akt. Tidsskr.* 175–195.

[13] L. de Haan & A. Ferreira (2006) *Extreme Value Theory.* Springer-Verlag.

[14] W. Feller (1971) *An Introduction to Probability Theory and Its Applications* **I**I (2nd ed.). Wiley.

[15] S. Foss, D. Korshunov & S. Zachary (2011) *An Introduction to Heavy-Tailed and Subexponential Distributions.* Springer-Verlag.

[16] S. Foss & S. Zachary (2003). The maximum on a random time interval of a random walk with long-tailed increments and negative drift. *Ann. Appl. Probab.* **1**3, 37–53.

[17] IAIS (2014) *Consultation Document December 2014. Risk-Based Global Insurance Capital Standard.* International Association of Insurance Supervisors.

[18] J.L. Jensen (1995) *Saddlepoint Approximations.* Clarendon Press, Oxford.

[19] H. Joe (1997) *Multivariate Models and Dependence Concepts.* Chapman & Hall.

[20] A. McNeil, R. Frey & P. Embrechts (2015) *Quantitative Risk Management. Concepts, Techniques and Tools* (2nd ed.). Princeton University Press.

[21] A. Müller & D. Stoyan (2002) *Comparison Methods for Stochastic Models and Risks.* Wiley.

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

[23] V.V. Petrov (1975) *Sums of Independent Random Variables.* Springer-Verlag.

[24] S. Resnick, S. (1987). *Extreme Values, Regular Variation, and Point Processes.* Springer-Verlag.

[25] S. Resnick (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling.* Springer-Verlag.

[26] L. Rüschendorf (2013) *Mathematical Risk Analysis.* Springer-Verlag.