Extremes and large deviations in the presence of heavy tails

Filip Lindskog, KTH

2nd Northern Triangular Seminar, 15 March 2010, KTH

lindskog@kth.se, www.math.kth.se/∼lindskog/

The outline of the tutorial is as follows.

Part I introduces and explains the concept of regularly varying random variables.

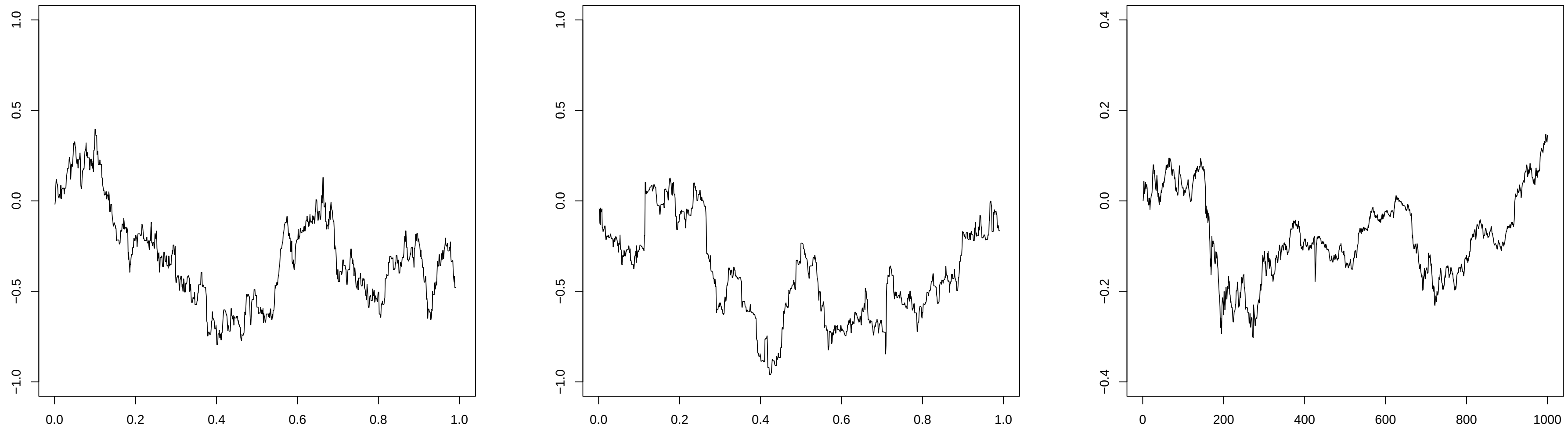
This begins with the Pareto distribution and ends with random variables taking values in rather general metric spaces.

One important point is the close connection between regular variation and weak convergence of measures. The whole box of tools from weak convergence theory is available.

Part II builds upon Part I and investigates extremes for stochastic processes.

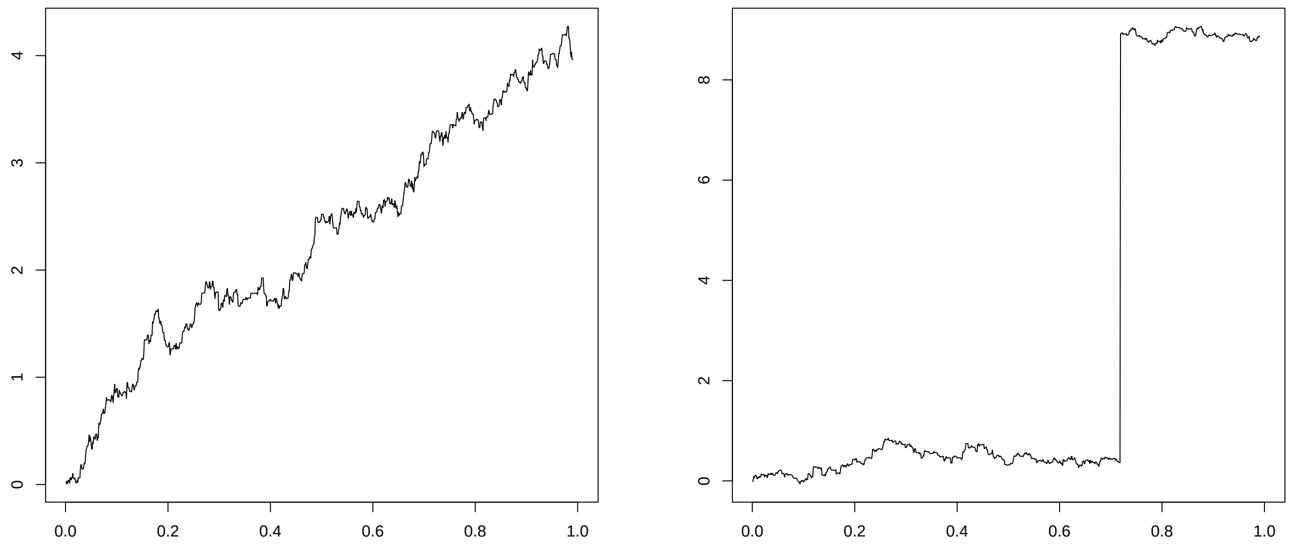
This begins with Lévy processes, continues to stochastic integral processes and ends with stochastic differential equations driven by Lévy noise.

Simulations of Compound Poisson processes fitted to log-price data.



Simulated sample paths of Compound Poisson processes with normally distributed jumps (left) and Student's t(4) distributed jumps (middle), and the log-price process of the DAX (right).

Large deviations heuristics: rare events occur in the most likely way.

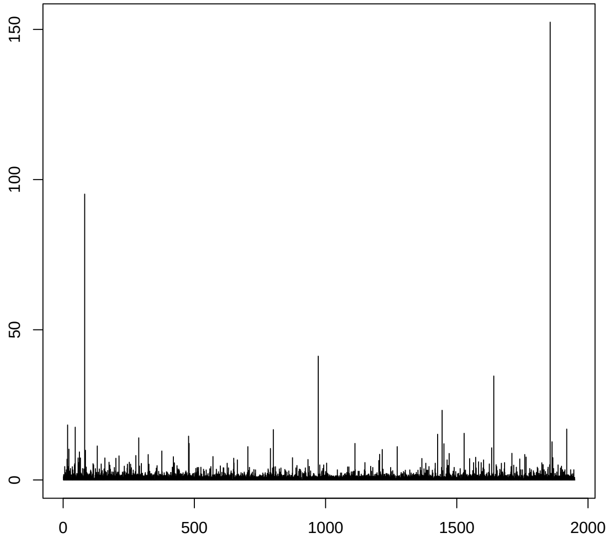


Simulated sample paths of Compound Poisson processes with normally distributed jumps (left) and Student's t(4) distributed jumps (right) conditioned on the supremum being large ($>$ 4). Comparing the conditional tail probabilities we find, for $\lambda > 1$,

$$\mathbb{P}(X > \lambda x \mid X > x) \sim \frac{1}{\lambda} e^{-(\lambda^2-1)x^2/2\sigma^2} \quad \text{vs.} \quad \mathbb{P}(X > \lambda x \mid X > x) \sim \lambda^{-4}$$

as $x \to \infty$. Many small upwards jumps vs. one big jump.

Dataset consisting of claims in million Danish Kroner from fire insurance in Denmark. We observe that there are a few claims much larger the 'every-day' claim. This suggests that the claims have a heavy-tailed distribution.



How do we estimate $\mathbb{P}(X > 100)$? Idea: scaling of an empirical estimate.

$$\mathbb{P}(X > \lambda x) = \mathbb{P}(X > \lambda x \mid X > x)\,\mathbb{P}(X > x) \approx c(\lambda) n_x/n,$$

where n_x/n is the empirical estimate of $\mathbb{P}(X > x)$. We must take x large enough so that $\mathbb{P}(X > \lambda x \mid X > x) \approx c(\lambda)$ but also small enough so that the empirical estimate n_x/n has small variance.

Regularly varying random variables

If X is a non-negative random variable satisfying

$$\lim_{x\to\infty} \mathbb{P}(X > \lambda x \mid X > x) = c(\lambda) > 0$$

for all λ in some interval (a, b) with $a \geq 1$, then there exists an $\alpha > 0$ such that

$$\lim_{x\to\infty} \mathbb{P}(X > \lambda x \mid X > x) = \lambda^{-\alpha} \quad \text{for all} \quad \lambda > 1$$

and X is said to be regularly varying with index α. In this case we can write

$$\mathbb{P}(X > x) = L(x)x^{-\alpha},$$

where $L \geq 0$ has the property $\lim_{x\to\infty} L(\lambda x)/L(x) = 1$.

The canonical X is a Pareto(α)-distributed X ($L(x) = 1$ for $x > 1$).

It holds that $\mathbb{E}[X^\beta] < \infty$ if $\beta < \alpha$ and $\mathbb{E}[X^\beta] = \infty$ if $\beta > \alpha$.

Sums

If $X_1, \ldots, X_n$ are independent and regularly varying, then

$$\mathbb{P}(X_1 + \cdots + X_n > x) \sim \mathbb{P}(X_1 > x) + \cdots + \mathbb{P}(X_n > x) \quad \text{as} \quad x \to \infty,$$

where $\sim$ means that RHS/LHS $\to 1$. In particular, if the X_ks are iid, then

$$\mathbb{P}(X_1 + \cdots + X_n > x) \sim n\,\mathbb{P}(X_1 > x) \sim \mathbb{P}(\max_k X_k > x) \quad \text{as} \quad x \to \infty.$$

A similar result holds if we let the number of terms grow: if $\alpha > 1$, then

$$\mathbb{P}(X_1 + \cdots + X_n - n\mathbb{E}[X_1] > x) \sim n\,\mathbb{P}(X_1 > x) \quad \text{as} \quad n \to \infty$$

for $x \geq \gamma n$ with $\gamma > 0$ arbitrary.

If a sum takes a large value x^*, then precisely one of its terms is large $\approx x^*$ and the rest take normal values:

$$\mathbb{P}\left(\max_k X_k > x \mid \sum_k X_k > x\right) = \frac{\mathbb{P}(\max_k X_k > x)}{\mathbb{P}(\sum_k X_k > x)} \to 1 \quad \text{as } x \to \infty.$$

Products

If $A \geq 0$ and X are independent, $\mathbb{E}[A^{\alpha+\delta}] < 0$ and $\mathbb{P}(X > x) = L(x)x^{-\alpha}$, then

$$\begin{aligned}\frac{\mathbb{P}(AX > x)}{\mathbb{P}(X > x)} &= \int \frac{\mathbb{P}(aX > x)}{\mathbb{P}(X > x)} dF_A(a) \\ &\sim \int a^{\alpha} dF_A(a) = \mathbb{E}[A^{\alpha}] \quad \text{as } x \to \infty.\end{aligned}$$

For iid copies $X_1, \ldots, X_n$ and independent $A_1, \ldots, A_n$ independent of the X_ks,

$$\begin{aligned}\mathbb{P}\Big(\sum_{k=1}^{n} A_k X_k > x\Big) \sim \sum_{k=1}^{n} \mathbb{P}(A_k X_k > x) &\sim \sum_{k=1}^{n} \mathbb{E}[A_k^{\alpha}]\, \mathbb{P}(X > x) \\ &= \Big(\sum_{k=1}^{n} \mathbb{E}[A_k^{\alpha}] \frac{1}{n}\Big) n\, \mathbb{P}(X > x) \\ &\sim \mathbb{E}[A_U^{\alpha}]\, \mathbb{P}(X_1 + \cdots + X_n > x) \quad \text{as } x \to \infty,\end{aligned}$$

where U is uniformly distributed on $\{1, \ldots, n\}$.

A closer look at the definition

The definition of regular variation was the convergence

$$\lim_{x\to\infty} \frac{\mathbb{P}(X \in xA)}{\mathbb{P}(X \in xE)} = \lim_{x\to\infty} \mathbb{P}(X \in xA \mid X \in xE) = c(A)$$

for $E = (1, \infty)$ and all $A = (\lambda, \infty)$, $\lambda > 1$, and that $c(A) > 0$ for these As.

This is equivalent to the convergence

$$\lim_{x\to\infty} \frac{\mathbb{P}(X \in xA)}{\mathbb{P}(X \in xE)} = \mu(A)$$

for all Borel sets $A \subset [0, \infty)$ bounded away from 0 (means $0 \notin A^-$, the closure of A) and some non-zero measure μ on $(0, \infty)$.

Furthermore, it follows that μ is given by $\mu(A) = \int_A x^{-\alpha-1} dx$.

Regular variation in $\mathbf{R}^n$

For a random vector X in $\mathbf{R}^n$ we can define regular variation in the same way: there exists a non-zero measure μ on $\mathbf{R}^n \setminus \{0\}$ and a Borel set E bounded away from 0 such that

$$\lim_{x\to\infty} \frac{\mathbb{P}(X \in xA)}{\mathbb{P}(X \in xE)} = \mu(A)$$

for all Borel sets A bounded away from 0 such that $\mu(\partial A) = 0$ (μ-continuity sets, $\partial A =$ boundary $=$ closure $-$ interior).

We can always take E to be the complement of a ball centered at 0.

It follows that μ has the following scaling property: there exists an $\alpha > 0$ such that

$$\mu(\lambda A) = \lambda^{-\alpha}\mu(A)$$

for all $\lambda > 0$ and all Borel sets A. In particular, μ is an infinite measure.

Why only convergence for μ-continuity sets?

Take $X = (X_1, X_2)$ with X_1 and X_2 independent and Pareto(α)-distributed. Then μ concentrates its mass on the coordinate axis:

$$\frac{\mathbb{P}(X \in x[(-\varepsilon,\varepsilon)\times(\lambda,\infty)])}{\mathbb{P}(|X|>x)} = \frac{\mathbb{P}(|X_1|<x\varepsilon)\,\mathbb{P}(X_2>\lambda x)}{\mathbb{P}(|X|>x)} \sim \frac{\mathbb{P}(X_2>\lambda x)}{\mathbb{P}(X_2>x)}$$

so $\mu((-\varepsilon,\varepsilon)\times(\lambda,\infty)) = \mu(\{0\}\times(\lambda,\infty)) = \lambda^{-\alpha}$. However,

$$\frac{\mathbb{P}(X \in x[\{0\}\times(\lambda,\infty)])}{\mathbb{P}(|X|>x)} = 0 \neq \mu(\{0\}\times(\lambda,\infty)).$$

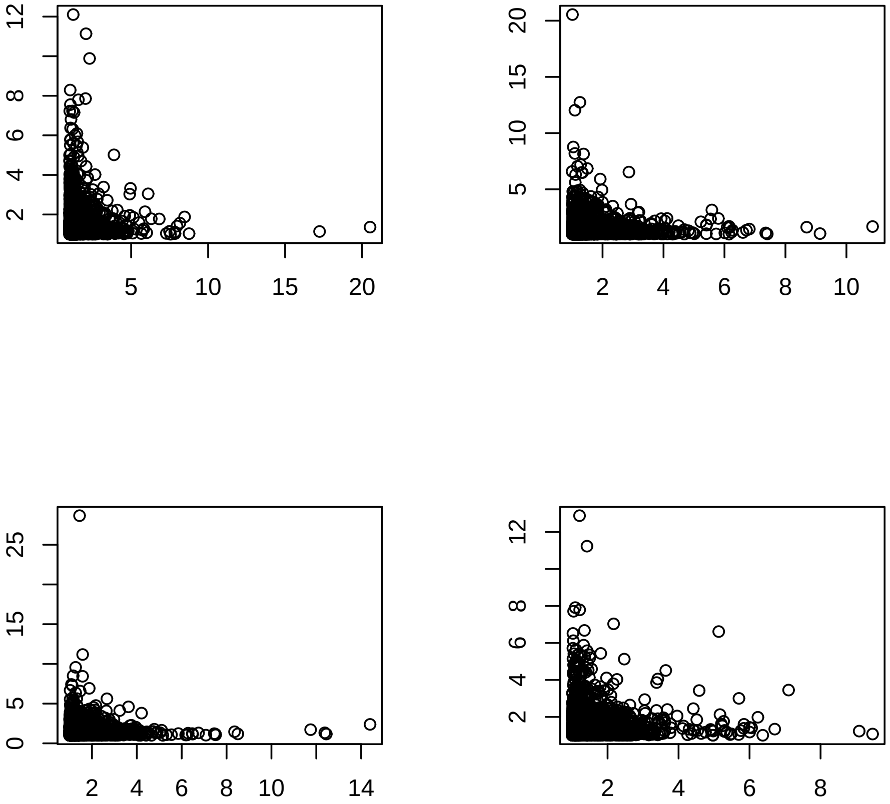
With $A = \{0\}\times(\lambda,\infty)$ we have

$$\partial A = A^- \setminus A^\circ = A^- = \{0\}\times[\lambda,\infty),$$

$\mu(\partial A) = \lambda^{-\alpha} > 0$ and $\mathbb{P}(X \in \partial A) = 0$.

The limit measure μ can assign positive mass to rays from the origin even though this is not so for the distribution of X.

Simulations of a bivariate Pareto(3)-distribution (independent components).



Convergence determining classes

How do you show the convergence

$$\lim_{x\to\infty} \frac{\mathbb{P}(X \in xA)}{\mathbb{P}(X \in xE)} = \mu(A)$$

for all μ-continuity sets bounded away from 0?

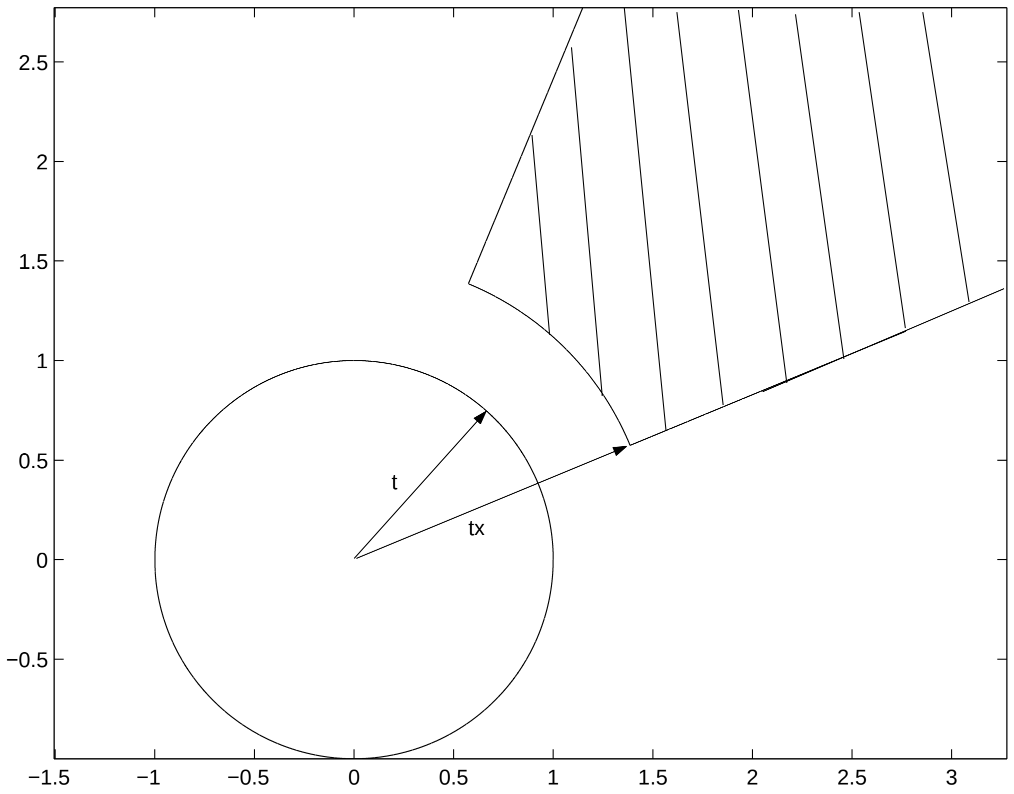
There are a lot of such sets...we cannot go through them all.

Fortunately it is sufficient to check the convergence for sets A is some suitable convergence determining class.

Examples of convergence determining classes are sets of μ-continuity sets of the form

$$V_{\lambda,S} = \{x : |x| > \lambda, x/|x| \in S\} \quad \text{for } \lambda > 0, \quad S \subset \{x : |x| = 1\},$$

or rectangles, balls, etc.



A set $V_{tx,S}$ in the convergence determining class.

Polar coordinates

X is regularly varying if and only if

$$\lim_{x\to\infty} \frac{\mathbb{P}(X \in xV_{\lambda,S})}{\mathbb{P}(|X| > x)} = \mu(V_{\lambda,S})$$

for all μ-continuity sets $V_{\lambda,S}$.

This says that X is regularly varying if and only if

$$\lim_{x\to\infty} \mathbb{P}(|X| > \lambda x, X/|X| \in S \mid |X| > x) = \lambda^{-\alpha}\,\mathbb{P}(\theta \in S),$$

for all $\lambda > 1$ and Borel sets S with $\mathbb{P}(\theta \in \partial S) = 0$, where θ is some random variable taking values in the sphere $\{x : |x| = 1\}$.

We are free to choose any norm $|\cdot|$ we want. We will get the same α regardless of the choice of norm.

Example: elliptical distributions

Let X have a centered non-degenerate elliptical distribution. This means that $X \stackrel{\mathrm{d}}{=} RAU$, where $R \geq 0$ and U are independent, U is uniformly distributed on the unit sphere in $\mathbf{R}^n$ wrt the usual norm, and A is an invertible $(n \times n)$-matrix.

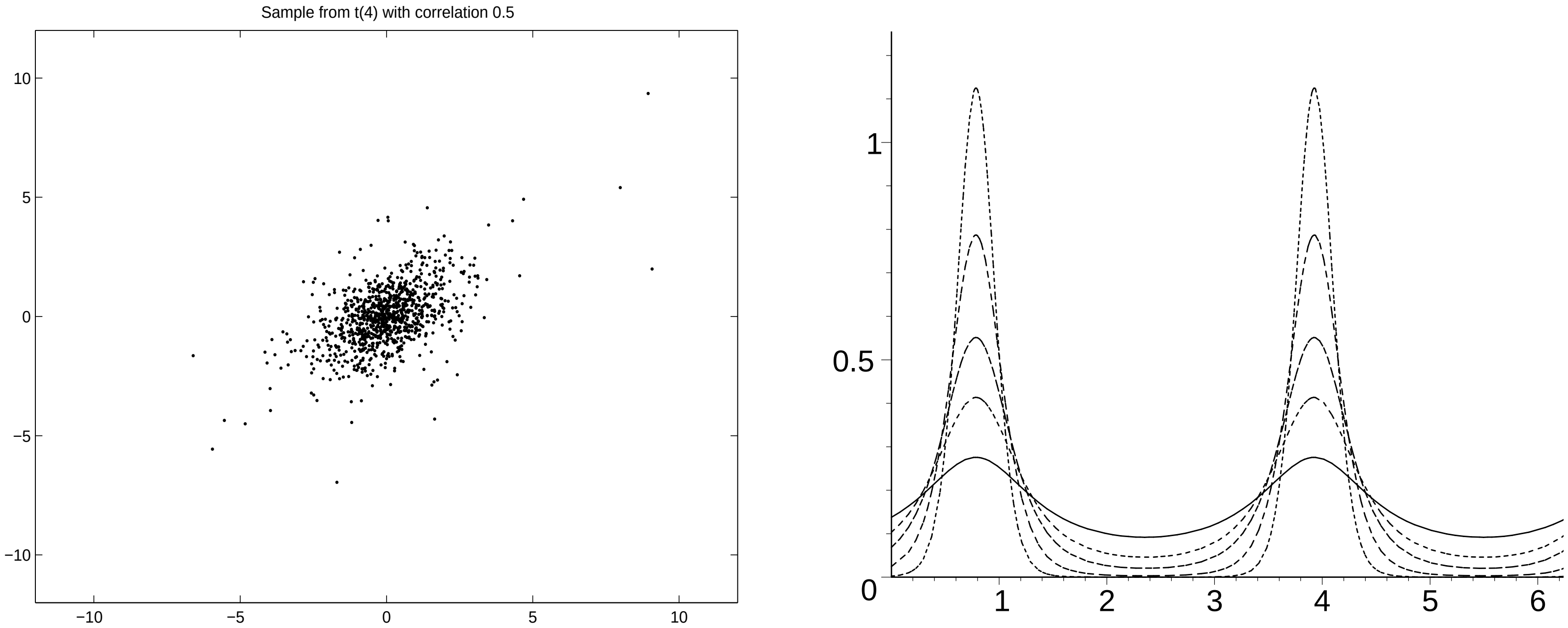
Ex: if R^2 is $\chi^2(n)$-distributed, then $X \sim N(0, AA^\top)$.

Take the Mahalanobis norm $|z| = z^\top \Sigma^{-1} z$ with $\Sigma = AA^\top$. Then

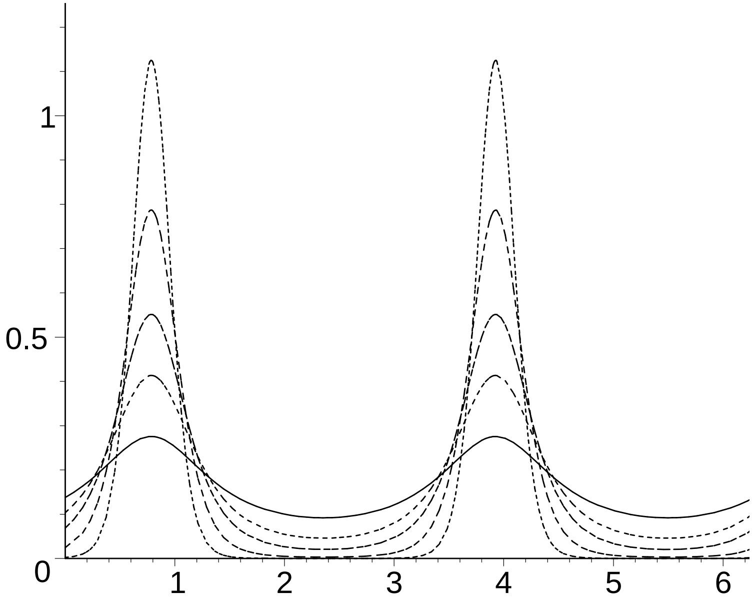
$$\mathbb{P}(|X| > \lambda x, X/|X| \in S \mid |X| > x) = \mathbb{P}(R > \lambda x \mid R > x)\, \mathbb{P}(AU \in S)$$

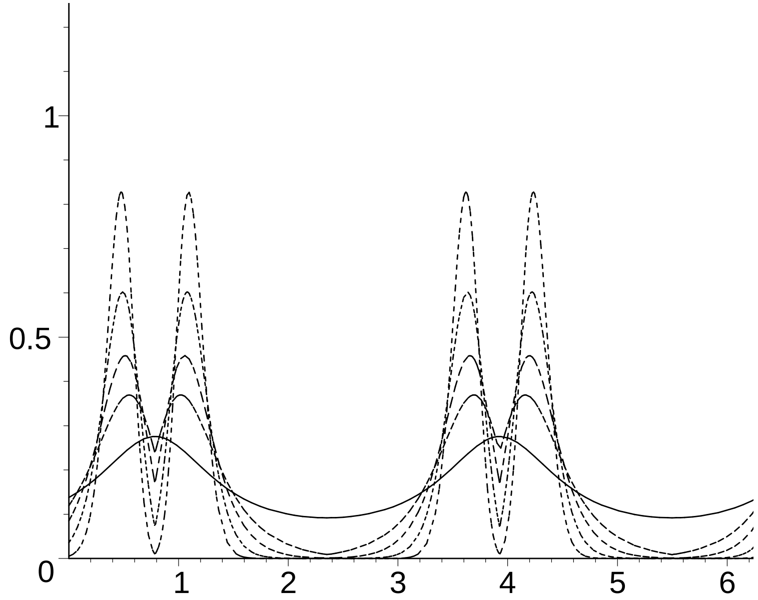
so X is regularly varying if and only if R is regularly varying.

X is regularly varying if and only this holds for $X + \mu$ so there is no loss of generality to consider centered elliptical distributions.



Left: Scatter plot of bivariate standard Student's $t(4)$, $\Sigma_{12} = 0.5$.
Right: Densities of the spectral measure of $\mathbf{X} \sim t(\alpha)$, $\Sigma_{12} = 0.5$,
$\alpha = 0, 2, 4, 8, 16$. Higher peaks corresponds to bigger α (i.e. lighter tails).





Spectral measures wrt usual Euclidean norm (left) and max-norm (right).

$$\lim_{x\to\infty} \mathbb{P}(X/|X| \in S \mid |X| > x), \quad |y|_2 = \sqrt{y_1^2 + y_2^2}, \quad |y|_\infty = \max(|y_1|, |y_2|).$$

$\{y : |y|_2 = x\}$ is a sphere of radius x, $\{y : |y|_\infty = x\}$ is a square with side length $2x$.

Tail indices $\alpha = 0, 2, 4, 8, 16$. $\Sigma_{11} = \Sigma_{22} = 1$ and $\Sigma_{12} = \Sigma_{21} = 0.5$. Larger tail indices corresponds to higher peaks.

Weak convergence and regular variation

The sequence $\{X_k\}$ converges is distribution to X_∞, $X_k \xrightarrow{d} X_\infty$, if

$$\lim_{k\to\infty} \mathbb{P}(X_k \in A) = \mathbb{P}(X_\infty \in A)$$

for all Borel sets A satisfying $\mathbb{P}(X_\infty \in \partial A) = 0$.

If the X_ks take values in $\mathbf{R}$, then this simply means pointwise convergence for the distribution functions $\lim_{k\to\infty} F_k(x) = F_\infty(x)$ for all continuity points x of F_∞.

If the X_ks take values in $\mathbf{R}^n$, then $X_k \xrightarrow{d} X_\infty$ if and only if $z \cdot X_k \xrightarrow{d} z \cdot X_\infty$ for all $z \in \mathbf{R}^n$. This is called the Cramér-Wold device.

Cramér-Wold device for regular variation on $\mathbf{R}^n$

Suppose that for some $z_0 \in \mathbf{R}^n$

$$b(z) := \lim_{x \to \infty} \frac{\mathbb{P}(z \cdot X > x)}{\mathbb{P}(z_0 \cdot X > x)} \quad \text{exists for all } z \in \mathbf{R}^n.$$

It follows that $b(\lambda z) = \lambda^\alpha b(z)$ for some $\alpha > 0$ and all z.

If X has non-negative components or if α is not an integer, then X is regularly varying,

$$\lim_{x \to \infty} \frac{\mathbb{P}(X \in xA)}{\mathbb{P}(X \in xE)} = \mu(A),$$

where $E = \{y : z_0 \cdot y > 1\}$ and μ is determined by $b(z) = \mu\{y : z \cdot y > 1\}$.

The result need not hold when X takes values in all of $\mathbf{R}^n$ and α is an integer.

Regular variation in a general setting

Not many of the properties of $\mathbf{R}^n$ are necessary in order to define regular variation. The definition in $\mathbf{R}^n$ actually qualifies as a definition of regular variation for measures on all sorts of natural spaces.

Let $(\mathbf{S}, d)$ be a complete separable metric space with $0 \in \mathbf{S}$ such that

The map $[0, \infty) \times \mathbf{S} \ni (\lambda, x) \mapsto \lambda x \in \mathbf{S}$ is well-defined and continuous (multiplication by a scalar is accepted and unproblematic), and

$d(0, \lambda_1 x) < d(0, \lambda_2 x)$ if $x \neq 0$ and $\lambda_1 < \lambda_2$ (multiplication with larger scalars pushes points further away from the origin along a ray).

Let $\mathbf{M}_0$ be the class of Borel measures on $\mathbf{S} \setminus \{0\}$ whose restrictions to $\mathbf{S} \setminus B_{0,r}$ are finite for every $r > 0$.

Definition A measure $\nu \in \mathbf{M}_0$ is regularly varying if there exists a non-zero measure $\mu \in \mathbf{M}_0$ and a Borel set $E \subset \mathbf{S}$ bounded away from 0 such that

$$\lim_{x \to \infty} \frac{\nu(xA)}{\nu(xE)} = \mu(A)$$

for all Borel sets $A \subset \mathbf{S}$ bounded away from 0 with $\mu(\partial A) = 0$.

It follows that $\mu(\lambda A) = \lambda^{-\alpha}\mu(A)$ for some $\alpha > 0$ and all $\lambda > 0$ and Borel sets A.

An $\mathbf{S}$-valued random variable is regularly varying if its distribution is regularly varying.

Note that the definition of regular variation is essentially a weak convergence statement: probability measures ν_x on $\mathbf{S}$ converges weakly to a probability measure ν_∞ on $\mathbf{S}$ if $\nu_x(A) \to \nu_\infty(A)$ for all Borel sets $A \subset \mathbf{S}$ with $\nu_\infty(\partial A) = 0$.

More general classes of heavy-tailed distributions

A distribution function F, with right tail $\overline{F} = 1 - F$, is long tailed ($\mathcal{L}$) if

$$\lim_{x\to\infty} \frac{\overline{F}(x-y)}{\overline{F}(x)} = 1 \quad \text{for all } y > 0,$$

and subexponential ($\mathcal{S}$) if

$$\lim_{x\to\infty} \frac{\overline{F^{n*}}(x)}{\overline{F}(x)} = n \quad \text{for all } n \geq 2.$$

Regular variation ($\mathcal{R}$) $\Rightarrow$ subexponentiality $\Rightarrow$ long tailedness.

Many of the one-dimensional properties of $\mathcal{R}$ hold more generally for $\mathcal{L}$ and $\mathcal{S}$. However, $\mathcal{L}$ and $\mathcal{S}$ are one-dimensional concepts with no natural multivariate analogues.

Examples: Lognormal and heavy-tailed Weibull ($\overline{F}(x) = e^{-cx^{\tau}}$, $\tau \in (0,1)$) belong to $\mathcal{S}$ but not $\mathcal{R}$.

Extremes and large deviations in the presence of heavy tails

Part II

Filip Lindskog, KTH

2nd Northern Triangular Seminar, 16 March 2010, KTH

Based on joint work with Henrik Hult.

lindskog@kth.se, www.math.kth.se/∼lindskog/

A suitable space for stochastic processes

We have seen that regular variation can be defined on rather general metric spaces $\mathbf{S}$.

We will now take $\mathbf{S}$ to be a space where many of the interesting continuous time stochastic processes live.

Take $\mathbf{S} = \mathbf{D}$ to be the space of functions from $[0, 1]$ to $\mathbf{R}^d$ that are right-continuous with left limits.

The space $\mathbf{D}$ is equipped with the Skorohod J_1-topology and we can choose a metric d on $\mathbf{D}$, generating this topology, such that $(\mathbf{D}, d)$ has the properties we want.

$d(y, z)$ is small if $\sup_t |y(t) - z(f(t))|$ is small for some increasing (time deformation) f with $f(t) \approx t$. Example: $1_{[\tau,1]}$ and $1_{[\tau+\varepsilon,1]}$ are close if $|\varepsilon|$ is small.

Regular variation for stochastic processes

Suppose that X is a $\mathbf{D}$-valued random variable: $X = \{X_t\}_{t\in[0,1]}$ is a stochastic process with sample paths in $\mathbf{D}$.

X is regularly varying if there exists a non-zero measure μ on $\mathbf{D}\setminus\{0\}$ and a Borel set $E\subset\mathbf{D}$ bounded away from 0 such that

$$\lim_{x\to\infty}\frac{\mathbb{P}(X\in xA)}{\mathbb{P}(X\in xE)} = \mu(A)$$

for all μ-continuity Borel sets $A\subset\mathbf{D}$ bounded away from 0.

We can take $E = \{y : \sup_t |y_t| > 1\}$ and in many cases even $E = \{y : |y_1| > 1\}$.

It seems rather difficult to check the convergence...

Sufficient condition for regular variation.

X is regularly varying if for any $t_1, \ldots, t_k \in [0,1]$ there exists a measure $\mu_{t_1,\ldots,t_k}$, not all of them zero, such that

$$\lim_{x\to\infty} \frac{\mathbb{P}((X_{t_1}, \ldots, X_{t_k}) \in xB)}{\mathbb{P}(X \in xE)} = \mu_{t_1,\ldots,t_k}(B), \quad B \subset \mathbf{R}^k,$$

and the probability of wild oscillations of amplitude $\approx x$ within an arbitrary small time interval is much smaller then $\mathbb{P}(X \in xE)$ as $x \to \infty$.

The analogue for weak convergence of probability measures is convergence of finite dimensional distributions and relative compactness (or tightness).

Lévy processes

If X is a Lévy process (stationary and independent increments), then X is regularly varying if and only if X_1 is regularly varying:

$$\lim_{x\to\infty}\frac{\mathbb{P}(X_1\in xB)}{\mathbb{P}(|X_1|>x)}=\mu_1(B),\quad B\subset\mathbf{R}^n$$

if and only if

$$\lim_{x\to\infty}\frac{\mathbb{P}(X\in xA)}{\mathbb{P}(|X_1|>x)}=\mu(A),\quad A\subset\mathbf{D}.$$

Moreover, $\mu(A)=\mathbb{E}[\mu_1\{z: z1_{[U,1]}\in A\}]$ with U uniformly distributed on $(0,1)$.

Note that, as expected, μ concentrates on the set of step functions with one step (a heavy-tailed Lévy process takes extreme values by making one big jump).

Moreover, the time of the big jump is uniformly distributed on $(0,1)$ (due to the stationary increments).

Illustration: Let X be a Lévy process with X_1 regularly varying, i.e.

$$\lim_{x\to\infty} \frac{\mathbb{P}(X_1 \in xB)}{\mathbb{P}(|X_1| > x)} = \mu_1(B).$$

What can be said about the tail behavior of the random variable $\int_0^1 X_t dt$?

Let $A = \{y \in \mathbf{D} : \int_0^1 y_t dt > 1\}$. Then A is bounded away from 0, $\mu(\partial A) = 0$, and

$$\begin{aligned}
\mu(A) &= \mu\{y = z1_{[u,1]} : y \in A\} \\
&= \mu\{y = z1_{[u,1]} : z(1-u) > 1\} \\
&= \int_0^1 \mu_1\{z : z > 1/(1-u)\} du \\
&= \mu_1(1,\infty) \int_0^1 (1-u)^\alpha du \\
&= \mu_1(1,\infty)/(\alpha+1),
\end{aligned}$$

where $\mu_1(1,\infty) = \lim_{x\to\infty} \mathbb{P}(X_1 > x)/\mathbb{P}(|X_1| > x)$.

This gives

$$\lim_{x\to\infty} \frac{\mathbb{P}(\int_0^1 X_t dt > \lambda x)}{\mathbb{P}(X_1 > x)} = \lim_{x\to\infty} \frac{\mathbb{P}(X \in \lambda x A)}{\mathbb{P}(|X_1| > \lambda x)} \frac{\mathbb{P}(|X_1| > \lambda x)}{\mathbb{P}(|X_1| > x)} \frac{\mathbb{P}(|X_1| > x)}{\mathbb{P}(X_1 > x)}$$
$$= \frac{\lambda^{-\alpha}}{\alpha + 1}$$

which gives the asymptotic approximation

$$\mathbb{P}\left(\int_0^1 X_t dt > \lambda x \right) \approx \frac{\lambda^{-\alpha}}{\alpha + 1} \mathbb{P}(X_1 > x).$$

The message here is that deriving the tail behavior of functionals of heavy-tailed Lévy processes is surprisingly easy!

Extremes for stochastic integral processes

Let $X = \{X_t\}_{t \in [0,1]}$ be a process of the type

$$X_t = \int_0^t V_s dY_s.$$

Here Y and V are defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$,

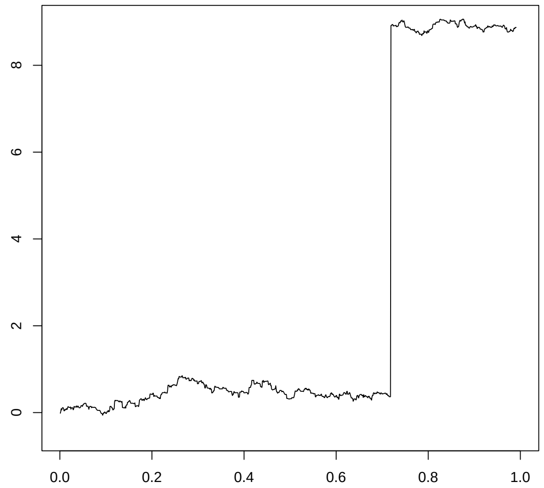
Y is a Lévy process (adapted, meaning that Y_t is $\mathcal{F}_t$-measurable, and with right-continuous sample paths with left limits) and

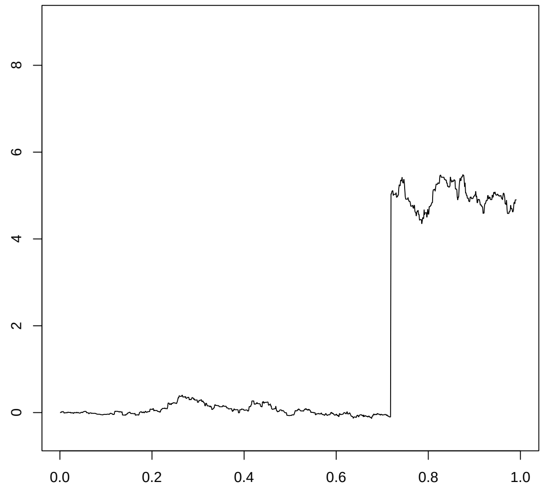
V is predictable (V_t is $\mathcal{F}_{t-}$-measurable) and has left-continuous sample paths with right limits, and $\min_k \sup_t |V_t^{(k)}| > 0$ a.s.

The essential point (modulo technicalities) is that V_t and dY_t are independent. However, V and Y may be strongly dependent as one may take e.g. $V_t = \sqrt{Y_{t-}}$.

Moreover, it will be assumed that Y is regularly varying (equivalent to Y_1 being regularly varying) and that the tails of V are lighter than those of Y.

As an example take the integrator Y to be a Compound Poisson processes with Student's t(4) distributed jumps and the integrand V given by $V_t = \sqrt{|Y_{t-}|}$.





Left: a simulated sample path of Y conditioned on the supremum being large,
Right: the realization of the stochastic integral process $\int_0^t V_s dY_s$.

In discrete time we can write the stochastic integral $\int_0^t V_s dY_s$ as

$$\sum_{k=1}^{n} V_k Y_k$$

where the Y_ks are iid and regularly varying with index $\alpha > 0$, V_k and Y_k are independent for each k, and $\mathbb{E}[V_k^{\alpha+\delta}] < \infty$. (Example: $V_k = \sqrt{Y_{k-1}}$.)

We have seen that if $\{V_k\}$ and $\{Y_k\}$ are independent, then

$$\mathbb{P}\Big(\sum_{k=1}^{n} V_k Y_k > x\Big) \sim \mathbb{E}[V_U^{\alpha}]\,\mathbb{P}\Big(\sum_{k=1}^{n} Y_k > x\Big) \sim \mathbb{E}[V_U^{\alpha}]\,\mathbb{P}\big(\max_k Y_k > x\big)$$

as $x \to \infty$.

Since only one Y_k will be large when $\sum_{k=1}^{n} V_k Y_k$ is large we guess that the dependence between the V_ks and Y_ks will not mess things up (need only independence between V_k and Y_k for each k).

We have no reason not to expect a similar result for $\int_0^t V_s dY_s$.

The result (after quite a bit of work)

We assume that Y_1 is regularly varying with index $\alpha > 0$:

$$\lim_{x\to\infty} \frac{\mathbb{P}(Y_1 \in xB)}{\mathbb{P}(|Y_1| > x)} = \mu_1(B), \quad B \subset \mathbf{R}^n.$$

This implies that Y is regularly varying:

$$\lim_{x\to\infty} \frac{\mathbb{P}(Y \in xA)}{\mathbb{P}(|Y_1| > x)} = \mathbb{E}[\mu_1\{z : z\mathbf{1}_{[U,1]} \in A\}], \quad A \subset \mathbf{D}$$

Moreover, if the process V is lighter-tailed than Y, $\mathbb{E}[\sup_t |V_t|^{\alpha+\delta}] < \infty$ for some $\delta > 0$, then also X is regularly varying

$$\lim_{x\to\infty} \frac{\mathbb{P}(X \in xA)}{\mathbb{P}(|Y_1| > x)} = \mathbb{E}[\mu_1\{z : zV_U\mathbf{1}_{[U,1]} \in A\}],$$

where U is uniformly distributed and independent of V.

Hence, the again the extremal behavior is described by a one-step function.

Example. Assume that $V_t > 0$ and determine the tail behavior of

$$\sup_{t\in[0,1]} X_t = \sup_{t\in[0,1]} \int_0^t V_s dY_s.$$

With $A = \{y : \sup_{t\in[0,1]} y_t > 1\}$ we find that (assuming that $\mu_1(1,\infty) > 0$)

$$\begin{aligned}
\lim_{x\to\infty} \frac{\mathbb{P}(\sup_{t\in[0,1]} X_t > x)}{\mathbb{P}(Y_1 > x)} &= \lim_{x\to\infty} \frac{\mathbb{P}(X \in xA)}{\mathbb{P}(|Y_1| > x)} \frac{\mathbb{P}(|Y_1| > x)}{\mathbb{P}(Y_1 > x)} \\
&= \mu(A)/\mu_1(1,\infty) \\
&= \mathbb{E}[\mu_1\{z : zV_U 1_{[U,1]} \in A\}]/\mu_1(1,\infty) \\
&= \mathbb{E}[\mu_1\{z : zV_U > 1\}]/\mu_1(1,\infty) \\
&= \mathbb{E}[\mu_1(V_U^{-1},\infty)]/\mu_1(1,\infty) \\
&= \mathbb{E}[V_U^\alpha]\mu_1(1,\infty)/\mu_1(1,\infty) \\
&= \int_0^1 \mathbb{E}[V_t^\alpha]dt.
\end{aligned}$$

Similarly

$$\lim_{x\to\infty} \frac{\mathbb{P}(\inf_{t\in[0,1]} X_t < -x)}{\mathbb{P}(Y_1 < -x)} = \int_0^1 \mathbb{E}[V_t^\alpha]dt.$$

Therefore the following asymptotic approximations hold

$$\mathbb{P}\left(\sup_{t\in[0,1]} \int_0^t V_s dY_s > \lambda x\right) \sim \lambda^{-\alpha} \int_0^1 \mathbb{E}[V_t^\alpha]dt\, \mathbb{P}(Y_1 > x)$$

and

$$\mathbb{P}\left(\inf_{t\in[0,1]} \int_0^t V_s dY_s < -\lambda x\right) \sim \lambda^{-\alpha} \int_0^1 \mathbb{E}[V_t^\alpha]dt\, \mathbb{P}(Y_1 < -x)$$

as $x \to \infty$.

Ruin probabilities under optimal investments

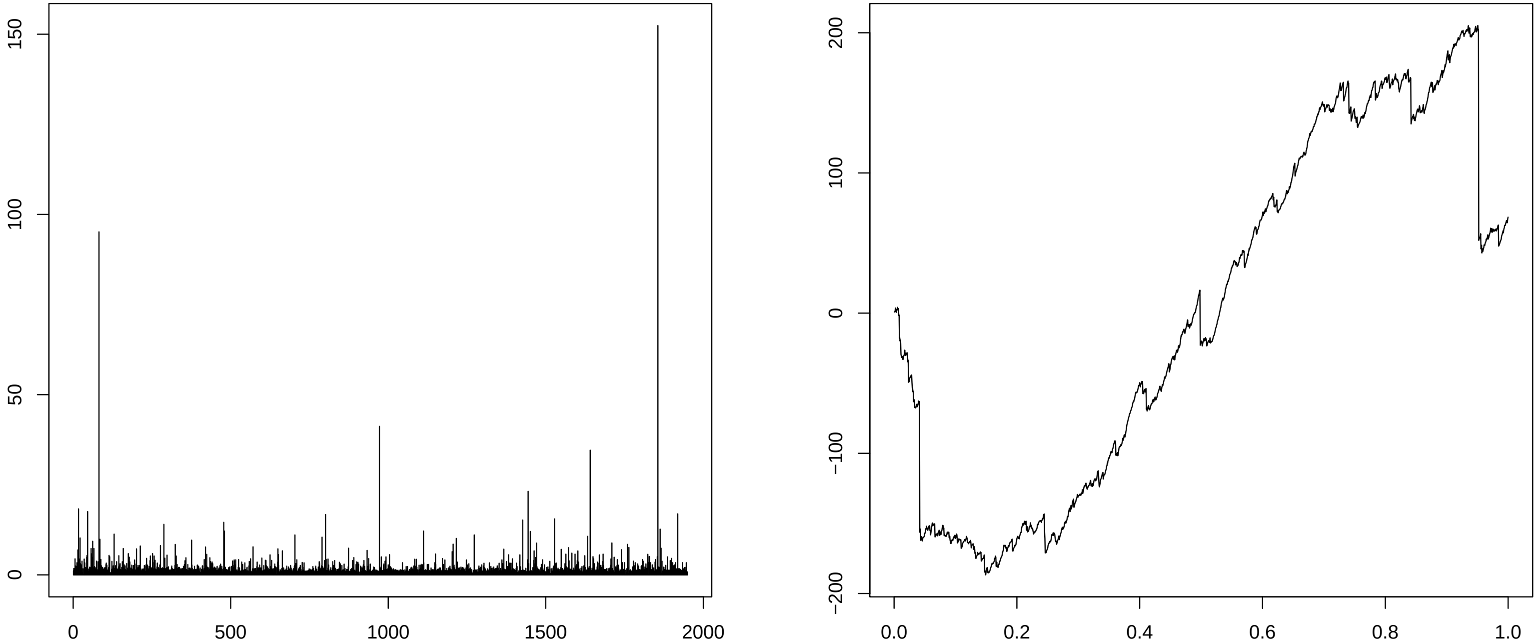
Consider an insurance company depositing a part of a its capital on a bank account and investing the rest in risky assets with spot prices S_t^k. The evolution of the capital is given by

$$X_t^\varepsilon = x + \int_0^t \pi_s^0 X_{s-}^\varepsilon r_{s-} ds + \sum_{k=1}^n \int_0^t \pi_s^k X_{s-}^\varepsilon \frac{dS_s^k}{S_{s-}^k} + \varepsilon Y_t,$$

where εY_t is the aggregated premium income minus claims, $\varepsilon > 0$ is small, and $\pi_t = (\pi_t^0, \dots, \pi_t^n)$ with $\pi_t^0 + \dots + \pi_t^n = 1$ is the investment strategy.

We will study the ruin probability $\mathbb{P}(\inf_{t\in[0,1]} X_t^\varepsilon < 0)$ without the strong assumption of a certain parametric model. Instead we will look for asymptotic approximations.

Fire insurance claims C_k and the process $Y_t = pt - \sum_{k=1}^{N_t} C_k$ based on these claims.



Technical assumptions

We assume that Y is a Lévy process: adapted, right-continuous sample paths with left limits, Y has stationary and independent increments (and is continuous in probability).

We assume that $\lim_{x\to\infty} \mathbb{P}(Y_1 < -\lambda x)/\,\mathbb{P}(Y_1 < -x) = \lambda^{-\alpha}$ for some $\alpha > 0$ and all $\lambda > 0$ (regular variation).

We assume that the S^k's are positive semimartingales.

This means that S^k is adapted, right-continuous sample paths with left limits, and that $S^k \approx$ martingal + process with sample paths having finite variation on finite time intervals.

Example: Lévy processes, sub- and supermartingales, etc. There is a well-established theory for stochastic integration with respect to semimartingales.

We assume that π is predictable (π_t is $\mathcal{F}_{t-}$-measurable) and r is adapted.

The solution to the SDE

We may write the evolution of the capital over time as

$$X_t^\varepsilon = x + \int_0^t X_{s-}^\varepsilon dZ_s + \varepsilon Y_t$$

where $Z = Z^\pi$ is the semimartingale

$$Z_t = \int_0^t \pi_s^0 r_{s-} ds + \sum_{k=1}^n \int_0^t \pi_s^k \frac{dS_s^k}{S_{s-}^k}.$$

We assume that the quadratic covariation $[S^k, Y] = 0$ for all k (e.g. if Y is compound Poisson with no jump time coinciding with those for the price processes S_t^k). Then

$$X_t^\varepsilon = \mathcal{E}(Z)_t \left(x + \varepsilon \int_0^t \frac{dY_s}{\mathcal{E}(Z)_{s-}} \right),$$

where $\mathcal{E}(Z)$ is the Doléans-Dade exponential of Z:

$$\mathcal{E}(Z)_t = e^{Z_t - \frac{1}{2}[Z,Z]_t^c} \prod_{s \in (0,t]} (1 + \Delta Z_s) e^{-\Delta Z_s}, \quad \Delta Z_s = Z_s - Z_{s-}.$$

The ruin probability

If $\Delta Z_t > -1$ for all t, then $\mathcal{E}(Z)_t > 0$ and the ruin probability can be written

$$\mathbb{P}\left(\inf_{t\in[0,1]} X_t^{\varepsilon} < 0\right) = \mathbb{P}\left(\inf_{t\in[0,1]} \mathcal{E}(Z)_t \left\{x + \varepsilon \int_0^t \frac{dY_s}{\mathcal{E}(Z)_{s-}}\right\} < 0\right)$$
$$= \mathbb{P}\left(\inf_{t\in[0,1]} \int_0^t \frac{dY_s}{\mathcal{E}(Z)_{s-}} < -\frac{x}{\varepsilon}\right).$$

Note: letting $\varepsilon \to 0$ or $x \to \infty$ gives the same asymptotic analysis.

Note: $\Delta S_t = S_t - S_{t-} > -S_{t-}$ so $\Delta Z_t = \sum_{k=1}^n \pi_t^k \Delta S_t / S_{t-} > -\sum_{k=1}^n \pi_t^k = -1$ if $\pi_t^k \geq 0$.

We see that ruin occurs due to a combination of bad investments ($\mathcal{E}(Z)_t$ close to zero) and either a lot of average size claims (the case of a light-tailed claim size distribution) or a few very large claims (the case of a heavy-tailed claim size distribution).

Asymptotic approximation of the ruin probability

Assume that the probability that only bad investments reduces the capital to almost zero is sufficiently small (OK for all Lévy processes Z with jumps which are greater than -1):

$$\mathbb{E} \sup_{t\in[0,1]} \mathcal{E}(Z)_t^{-\alpha-\delta} < \infty \quad \text{for some } \delta > 0.$$

Then (as we have already seen) it holds that

$$\begin{aligned} \mathbb{P}\Big(\inf_{t\in[0,1]} X_t^{\varepsilon} < 0\Big) &\sim \mathbb{P}\Big(\int_0^t \frac{dY_s}{\mathcal{E}(Z)_{s-}} < -\frac{x}{\varepsilon}\Big) \\ &\sim x^{-\alpha} \int_0^1 \mathbb{E}\mathcal{E}(Z)_t^{-\alpha} dt\, \mathbb{P}(Y_1 < -1/\varepsilon) \end{aligned}$$

as $\varepsilon \to 0$.

The integral often rather easy to compute (e.g. when Z is a Lévy process).

Optimal investment strategies

Let's now look at optimal investment strategies (optimal in the sense of minimizing the ruin probability). This means that for the integrator

$$Z_t^\pi = \int_0^t \pi_s^0 r_{s-} ds + \sum_{k=1}^n \int_0^t \pi_s^k \frac{dS_s^k}{S_{s-}^k}$$

we choose $\pi_t = \pi_t(\varepsilon)$ at time $t-$ optimally given ε, information on the current reserve, prices, claim amounts, etc.

Let Π be a family of investment strategies such that it holds that $\inf_{t\in[0,1]} \Delta Z_t^\pi > -1$ for all $\pi \in \Pi$ and

$$\sup_{\pi\in\Pi} \mathbb{E} \sup_{t\in[0,1]} \mathcal{E}(Z^\pi)_t^{-\alpha-\delta} < \infty \quad \text{for some } \delta > 0.$$

This means that only bad investments do not lead to ruin and that it is sufficiently unlikely that only bad investments lead to almost ruin.

Robust asymptotics

Then it holds that

$$\lim_{\varepsilon\to 0}\sup_{\pi\in\Pi}\left|\frac{\mathbb{P}(\inf_{t\in[0,1]} X_t^{\varepsilon,\pi}<0)}{\mathbb{P}(Y_1<-1/\varepsilon)} - x^{-\alpha}\int_0^1 \mathbb{E}\mathcal{E}(Z^\pi)_t^{-\alpha}dt\right| = 0$$

and in particular that

$$\begin{aligned}\lim_{\varepsilon\to 0}\inf_{\pi\in\Pi}\frac{\mathbb{P}(\inf_{t\in[0,1]} X_t^{\varepsilon,\pi}<0)}{\mathbb{P}(Y_1<-1/\varepsilon)} &= \inf_{\pi\in\Pi}\lim_{\varepsilon\to 0}\frac{\mathbb{P}(\inf_{t\in[0,1]} X_t^{\varepsilon,\pi}<0)}{\mathbb{P}(Y_1<-1/\varepsilon)} \\ &= \inf_{\pi\in\Pi} x^{-\alpha}\int_0^1 \mathbb{E}\mathcal{E}(Z^\pi)_t^{-\alpha}dt.\end{aligned}$$

Asymptotic approximation under optimal investments

This gives the asymptotic approximation for the ruin probability under the optimal investment strategy $\widehat{\pi}(\varepsilon)$,

$$\mathbb{P}\Big(\inf_{t\in[0,1]} X_t^{\varepsilon,\widehat{\pi}(\varepsilon)} < 0\Big) = \inf_{\pi\in\Pi} \mathbb{P}\Big(\inf_{t\in[0,1]} X_t^{\varepsilon,\pi} < 0\Big).$$

We have

$$\mathbb{P}\Big(\inf_{t\in[0,1]} X_t^{\varepsilon,\widehat{\pi}(\varepsilon)} < 0\Big) \sim x^{-\alpha} \inf_{\pi} \int_0^1 \mathbb{E}\mathcal{E}(Z^\pi)_t^{-\alpha} dt\, \mathbb{P}(Y_1 < -1/\varepsilon)$$

as $\varepsilon \to 0$.

Note that the right-hand side is a lot simpler to evaluate than the essentially impossible left-hand side.

Thanks for your attention!

If you haven’t had enough of this, more details can be found in:

Boman, Lindskog (2009) *Support theorems for the Radon transform and Cramér-Wold theorems*, J. Theoret. Probab. 22, 683-710.

Hult, Lindskog (2006) *Regular variation for measures on metric spaces*, Publ. Inst. Math. (Beograd) (N.S.) 80, 121-140.

Hult, Lindskog (2007) *Extremal behavior of stochastic integrals driven by regularly varying Lévy processes*, Ann. Probab. 35, 309-339.

Hult, Lindskog (2010) *Ruin probabilities under general investments and heavy-tailed claims*, Finance Stoch., accepted for publication.