Lévy processes and Lévy copulas with an application in insurance

Martin Hunting

Thesis for the degree of Master of Science
Mathematical statistics
University of Bergen, Norway
June 1, 2007

UNIVERSITETET I BERGEN
Matematisk institutt

This thesis is written in $\LaTeX\,2_\varepsilon$ using the "uib-mi-master" document class, developed by Karl Ove Hufthammer. It was compiled by pdfeTeX-1.21a on June 7, 2007. The body text is 11 point Bitstream Charter. The math font is Math Design Charter, the heading font is HV Math, and the computer code font is Bera Mono.

Acknowledgments

I would like to express deep gratitude to my advisor, Trygve S. Nilsen, for coping with my foolhardiness and, more than once, forcing me back to the right track when I got lost. I would also like to thank Marius Fredheim for making me aware of the Danish fire insurance dataset and Karl Ove Hufthammer for helping me solve a multitude of technical problems and issues. Not least would I like to thank my father for invaluable proofreading assistance.

This thesis deals with infinitely divisible distributions and Lévy processes. Key-
words: Infinitely divisible distributions, stable distributions, Levy copula, compound Poisson distribution, generalized Pareto distribution.

Contents

Notation

$\emptyset$	the empty set
w.r.t	with respect to
a.s.	almost surely
a.e.	almost everywhere with respect to the Lebesgue measure
i.i.d.	independent and indentically distributed
$\cdot$	unspecified set
$\#A$	the number of elements of a set A
$:=$	defined as
$\stackrel{d}{=}$	equal in distribution
$\stackrel{d}{\to}$	convergence in distribution
Dom	domain
Ran	range
$\mathbb{R}$	the set of real numbers
$\overline{\mathbb{R}}$	the extended set of real numbers $\mathbb{R}\bigcup\{-\infty,+\infty\}$
t	time.
T	transpose
$A^{\complement}$	the complement of the set A
$\mathbb{C}$	the set of complex numbers
$\lvert c\rvert$	absolute value of c or, if c is complex, the modulus of c
$\bar{z}$	The conjugate of the complex number z
$\times$	Cartesian product
$\mathbb{R}^d$	the d-dimensional space $\underbrace{\mathbb{R}\times\mathbb{R}\times\ldots\times\mathbb{R}}_{d \text{ times}}$
$\mathscr{B}(\mathbb{R}^d)$	the Borel σ-algebra of $\mathbb{R}^d$, i.e. the σ-algebra generated by the open sets of $\mathbb{R}^d$
$m(B)$	the Lebesgue measure of a set B. m(dx) is written dx
$\bigotimes_1^n \mathscr{A}_j$	product σ-algebra of the σ-algebras $\mathscr{A}_1\ldots\mathscr{A}_n$.
$\overline{\mathbb{R}}^d$	d-dimensional space $\underbrace{\overline{\mathbb{R}}\times\overline{\mathbb{R}}\times\ldots\times\overline{\mathbb{R}}}_{d \text{ times}}$
card	cardinality
$\langle\mathbf{a},\mathbf{b}\rangle$	scalar product of vectors a and b
$\mathbf{0}$	$(0,0,\ldots,0)$ generalized origo vector
$\widehat{\mathbb{P}}$	Fourier transformation of the probability distribution $\mathbb{P}$
$f^+(x)$	$\max(0,f(x))$
$f^-(x)$	$-\min(0,f(x))$

$f(x) \sim g(x)$	$\lim_{x\to a} \frac{f(x)}{g(x)} = 1$ for some limit point $a \in \overline{\mathbb{R}}$.
	In this thesis a will be ∞ unless otherwise specified.
$f(x) = O(g(x))$	$\lim_{x\to a} \frac{\lvert f(x)\rvert}{\lvert g(x)\rvert} < \infty$ for some limit point $a \in \overline{\mathbb{R}}$.
$f(x) = o(g(x))$	$\lim_{x\to a} \frac{\lvert f(x)\rvert}{\lvert g(x)\rvert} = 0$ for some limit point $a \in \overline{\mathbb{R}}$.
$1_A(x)$	indicator function:
	$= 1$ if $x \in A$
	$= 0$ otherwise
$h \in \mathscr{R}_\gamma$	h is a regular varying function with index γ.
C^n	Class of all functions whose partial derivatives of order $\leq n$
	all exist and are continuous.
$\text{sign}(x)$	sign of x, i.e. $+1$ if $x \geq 0$, $\quad -1$ if $x < 0$.
$C(f)$	points where f is continuous
sup	supremum, the least upper bound
inf	infemum, the greatest lower bound
$\lim_{x\downarrow a} g(x)$	limit of $g(x)$, letting x decrease towards a
$\lim_{x\uparrow a} g(x)$	limit of $g(x)$, letting x increase towards a

1
Introduction

1.1 Topics covered in the thesis

This thesis discusses Lévy processes and Lévy copulas. In connection with Lévy processes we treat some of the theory behind infinitely divisible distributions, acknowledging that the two classes are equivalent. Within the class of Lévy processes we will mostly look at stable processes and compound Poisson processes.

1.1.1 Lévy processes

Origin of Lévy processes

The theory of Lévy processes dates back to the late 1920's, after de Finetti first introduced the class of infinitely divisible distributions. In 1934 those distributions were shown by Paul Lévy to have characteristic functions of the form given by the Lévy-Khintchine formula. Since then Lévy processes have become popular tools for modelling in finance, insurance and physics.

1.1.2 Lévy copulas

Copulas are functions that can be regarded as (a) functions that join or "couple" a multidimensional distribution to its one-dimensional margins or (b) as multivariate distributions whose one-dimensional margins are uniform on the interval $(0,1)$.

In Tankov (2003b) Peter Tankov introduced Lévy copulas to model the dependencies between components of a multidimensional spectrally positive Lévy process. Lévy copulas for more general Lévy processes are discussed in Cont and Tankov (2004). Lévy copulas have many similarities with other copula functions, but have the domain $[0,\infty]^d$ for $d = 2, 3, \ldots$ rather than $[0,1]^d$.

1.1.3 Stable processes

In this thesis we implement an algorithm given on page 202 in Cont and Tankov (2004) for simulation of a two-dimensional Lévy process whose components are stable processes with stable distributions. Stable distributions are characterized by:

- Having the stability property (see definition 3.3.2 on page 27).
- The fact that a distribution has a domain of attraction (defined in definition 2.1.6 on page 13) if and only if it is stable.
- Having infinite variance (except for the Gaussian distribution).
- Having an index $\alpha \in (0,2]$. Thix index will be explained in section 3.3.2 on page 27.

Stable processes are stochastic processes whose increments obey a stable distribution. For a stable process the stability property translates into the concept of self-similarity (defined in definition 4.2.1 on page 39).

1.2 Compound Poisson processes

A favoured approach in insurance is to model a risk process as a compound Poisson process, with positive jumps representing the insurance claims. Classical ruin theory is based on the assumption that all the claims are independent and identically distributed. The assumption of all claims being independent is dropped in Bregman and Klüppelberg (2005). Discussed there are two dependent compound Poisson processes X_t and Y_t with positive jumps and whose dependence is described by a Clayton Lévy copula. The sum $X_t + Y_t$ is identified as a compound Poisson process with new Poisson intensity and claim distribution.

Several new ruin probability formulas are given in Bregman and Klüppelberg (2005). In this thesis we will estimate the parameters for one of these formulas

using the multivariate Danish fire insurance claims dataset provided by Alexander McNeil and available from
`http://www.ma.hw.ac.uk/~mcneil/data.html`.

1.3 Applications of Lévy processes

1.3.1 Lévy processes in finance

As described in the introduction of Schoutens (2003), modelling financial markets with stochastic processes began in 1900 with Bachelier (1900). He modelled the prices of stocks listed at the Paris Bourse as a *Brownian motion*. Also known as a *Wiener-process*, Brownian motion is a stochastic process with independent, stationary increments that obey a Gaussian distribution. 65 years later another, more appropriate model was suggested in Samuelson (1965), where the *logarithms* of the stock prices were modelled as a Wiener process. This model is known as *geometric Brownian motion*. In Black and Scholes (1973) it was demonstrated how to price European options based on the geometric Brownian model. This stock-price model has been widely acclaimed and is now known as the Black-Scholes model. As pointed out in chapter 1 in Cont and Tankov (2004) there are, however, a number of flaws with the Black-Scholes model. Some of the most serious are the following:

- Continuity:
  Brownian motion is inherently *continuous*, while compelling empirical evidence has made it clear that the trajectories of log-prices have a large number of discontinuities.

- Scale invariance:
  The statistical properties of Brownian motion are the same at all time resolutions. On page 2 in Cont and Tankov (2004), the path of the log-price of SLM[1] in the period 1993-1997 is compared with the path of a simulated Brownian motion. While the Brownian path looks the same over a one-month period as over three years or three months, the price behavior over this period is clearly dominated by a large downward jump, which accounts for half of the monthly return. On an *intra-day* scale the price moves essentially through jumps, while the Brownian model retains the same continuous behavior as over long horizons. As noted on page 4 in Cont and Tankov (2004), "Assuming that prices move in a continuous manner amounts to neglecting the abrupt movements in which most of the *risk* is concentrated."

- Light tails:
  High variability is a constantly observed feature of financial asset returns.

> The empirical distribution of returns decays slowly at infinity and very large moves have a significant probability of occuring. As an example, six-standard deviation market moves are commonly observed in all markets. As noted by Cont and Tankov, the Gaussian distribution, in the other hand, is a light-tailed distributiion, and in a Gaussian model a daily return of such magnitude occurs on average less than once in a million years.

Many Lévy process models allow both discontinuities and heavy tails and have therefore been suggested by several authors as candidates for option pricing models (see chapter 4 in Cont and Tankov (2004)).

1.3.2 Applications of stable processes in physics

While stable process models remain controversial in finance (for two discussions on the matter see section 7.3 in Cont and Tankov (2004) and section 17.7 in Uchaikin and Zolotarev (1999)), they are routinely applied in several branches of physics. Common textbook examples where “the basic physical mechanism inexorably leads to a description in terms of an α-stable law with a particular α” (Woyczyńscki (2001)) include the following (see Woyczyńscki (2001)) :

Example 1.3.1: The first hitting time for the Brownian particle

Consider a Brownian particle moving in $\mathbb{R}$ whose trajectory $X_t, t \geq 0$, starts at $X_0 = 0$. The first time, $T_b > 0$, it hits the barrier located at $x = b > 0$ is a random variable that can be defined by the formula

$$T_b = \inf\left\{t \geq 0 : X_t = b\right\}.$$

It is shown in Woyczyńscki (2001) that T_b obeys the Lévy distribution defined in equation 3.20 on page 33. This Lévy distribution is a stable distribution with index $\alpha = 1/2$.

[2] The SLM corporation is listed at the New York Stock Exchange and is a member of Standard & Poor’s S&P 500 index.

Example 1.3.2: Particles emitted from a point source

Consider a source located at the point $(0, \eta)$ in the $\mathbb{R}^2$ plane, emitting particles into the right half-space with random directions (angles), Θ, uniformly distributed on the interval $[-\pi/2, \pi/2]$. The particles are detected by a flat panel device represented by the vertical line $x = \tau$ at the distance τ from the source. In Woyczyńscki (2001) the probability distribution function of the random variable representing the position Y of particles on the detecting device is shown to obey a one-dimensional Cauchy distribution, defined in equation 3.3 on page 24. Cauchy distributions of any dimension are stable distributions with index $\alpha = 1$.

Example 1.3.3: Stars, uniformly distributed in space

Consider a model of the universe in which the stars with masses $M_1 \geq 0, i = 1, 2, \ldots$ located at positions $\mathbf{X}_i \in \mathbb{R}^3, i = 1, 2, \ldots,$ interact via the Newtonian gravitational potential, exerting force

$$\mathbf{G}_i = gM_i \frac{\mathbf{X}_i}{\left|\mathbf{X}_i\right|^3} \in \mathbb{R}^3, \quad i = 1, 2, \ldots,$$

on a unit mass located at the origin $(0, 0, 0)$. Here g is the universal gravitational constant. Make the assumptions that

- The locations $\mathbf{X}_i, i = 1, 2, \ldots$ form a Poisson point process in $\mathbb{R}^3$ with density ρ.
- The masses $M_i, i = 1, 2, \ldots$ are i.i.d. variables.

Let $\mathbf{G}_R$ be the total gravitational force on a unit mass located at the origin, exerted by stars located inside a ball B_R, centered at $(0, 0, 0)$ and of radius R, that is

$$\mathbf{G}_R = \sum_{i:|X_i| \leq R} \mathbf{G}_i$$

It is then shown in Woyczyńscki (2001) that the limit $\lim_{R\to\infty} \mathbf{G}_R$ obeys a three-dimensional, spherically symmetric stable distribution with index $\frac{3}{2}$. In astrophysics this distribution is known as the *Holtsmark distribution*.

More examples of applications of stable distributions/stable processes are found in chapter 10-17 in Uchaikin and Zolotarev (1999).

2

Basic definitions and results

This chapter is a collection of definitions and results, mostly taken from chapters 1 and 2 in Sato (1999) and included here to be used as a reference.

2.1 Probability measure

Definition 2.1.1: Probability space

> Let Ω be a set, $\mathscr{F}$ a σ-algebra of subsets in Ω, and $\mathbb{P}$ a measure on $\mathscr{F}$. The triplet $(\Omega, \mathscr{F}, \mathbb{P})$ is then called a *measure space*. If $\mathbb{P}(\Omega) = 1$ then $(\Omega, \mathscr{F}, \mathbb{P})$ is called a *probability space*.

Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, any set $A \in \mathscr{F}$ is called an *event*, and $\mathbb{P}[A]$ is called the *probability* of the event A. The σ-algebra generated by the open sets in $\mathbb{R}^d$ is called the *Borel σ-algebra*. A real valued function $f(\mathbf{x})$ on $\mathbb{R}^d$ is called *measurable*[1] if it is $\mathscr{B}(\mathbb{R}^d)$-measurable. We shall say that F is a probability measure on $\mathbb{R}^d$ if F is a probability measure on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$.

[1] Let Ω and Θ be two abstract spaces , $\mathscr{M}$ be a σ-algebra on Ω and $\mathscr{N}$ be a σ-algebra on Θ. A function $f : \Omega \to \Theta$ is called *measurable* if for any set $E \in \mathscr{N}$ the set $\{\omega : f(\omega) \in E\}$ is included in the σ-algebra $\mathscr{M}$.

Definition 2.1.2: Random variable

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. A mapping X from Ω into $\mathbb{R}^d$ is called an $\mathbb{R}^d$-valued *random variable* (or random variable on $\mathbb{R}^d$) if X is $\mathscr{F}$-measurable.

Let $B \in \mathscr{B}\left(\mathbb{R}^d\right)$. We write $\mathbb{P}(\omega : X(\omega) \in B)$ as $\mathbb{P}(X \in B)$. As a mapping of B this is a probability measure on $\mathscr{B}(\mathbb{R}^d)$, which we denote by $\mathbb{P}_X$ and call the distribution (or law) of X.

In general, probability measures on $\mathscr{B}(\mathbb{R}^d)$ are called *distributions* on $\mathbb{R}^d$. If two random variables X, Y on $\mathbb{R}^d$ (not necessarily on the same probability space) have an identical distribution, i.e. $\mathbb{P}_X = \mathbb{P}_Y$, we write $X \stackrel{d}{=} Y$.

Definition 2.1.3: Weak convergence

Let F_n and F be probability measures on $\mathbb{R}^d$. The sequence $\{F_n\}$ converges *weakly* to F if $\left\{\int_{\mathbb{R}^d} f(\mathbf{x})F_n(d\mathbf{x})\right\}$ converges to $\int_{\mathbb{R}^d} f(\mathbf{x})F(d\mathbf{x})$ for every function f which is real-valued, continuous and bounded on $\mathbb{R}^d$.

Definition 2.1.4: Convergence in distribution

Let $\left\{(X_n)_{n\geq 1}\right\}$ be a sequence of $\mathbb{R}^d$-valued random variables. We say $\{X_n\}$ *converges in distribution* to X if $\mathbb{P}_{\{X_n\}}$ converges weakly to $\mathbb{P}_X$. We write $X_n \xrightarrow{d} X$.

Definition 2.1.5: Random walk

Let $\{Z_n : n = 1, 2, \ldots,\}$ be a sequence of independent and identically distributed $\mathbb{R}^d$-valued random varibles. Let $S_0 = \mathbf{0}, S_n = \sum_{j=1}^{n} Z_j$ for $n = 1, 2, \ldots$. Then $\{S_n : n = 0, 1, \ldots\}$ is a *random walk* on $\mathbb{R}^d$.

Definition 2.1.6: Domain of attraction

Let S_n be a random walk and F be the common distribution. Then F is said to belong to the *domain of attraction* of a probability measure R if there are constants $b_n > 0$ and constant vectors $\mathbf{c}_n$ such that the series $\{b_n S_n + \mathbf{c}_n\}$ converges to R in distribution.

A random variable X on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is said to have a property A *almost surely* (abbreviated a.s.) if there is a measurable set $\Omega_0 \in \mathscr{F}$ with $\mathbb{P}[\Omega_0] = 1$ such that, for every element $\omega \in \Omega_0$, $X(\omega)$ has the property A.

If $X = \mathbf{c}$ a.s., where $\mathbf{c}$ is constant vector in a euclidean space, we say that the distribution of X is *trivial*.

If X is a real-valued random variable and if $\left|\int_\Omega X(\omega)\mathbb{P}_X(d\omega)\right| < \infty$, then the integral is called the *expectation* of x and is denoted by $\mathbb{E}[X]$ or $\mathbb{E}X$. If in addition X is a random variable on $\mathbb{R}^d$, and $f(x)$ is a bounded measurable function on $\mathbb{R}^d$, then

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f(\mathbf{x})\mathbb{P}_X(d\mathbf{x}).$$

Definition 2.1.7: Independence

Let X_j be an $\mathbb{R}^{d_j}$-valued random variable for $j = 1,\dots,n$. The family $\{X_1,\dots,X_n\}$ is *independent* if, for every set $B_j \in \mathscr{B}(\mathbb{R}^{d_j})$, $j = 1,\dots,n$,

$$\mathbb{P}\left(X_1 \in B_1,\dots,X_n \in B_n\right) = \mathbb{P}\left(X_1 \in B_1\right)\mathbb{P}\left(X_2 \in B_2\right)\dots\mathbb{P}\left(X_n \in B_n\right).$$

We say that $X_1,\dots,X_n$ are independent if the family $\{X_1,\dots,X_n\}$ is independent. An infinite family of random variables is independent, if every finite subfamily of it is independent.

Definition 2.1.8: Convolution

The *convolution* F of two distributions F_1 and F_2 on $\mathbb{R}^d$, denoted by $F = F_1 * F_2$, is a distribution defined by

$$F(B) := \iint_{\mathbb{R}^d\times\mathbb{R}^d} 1_B(\mathbf{x}+\mathbf{y})F_1(d\mathbf{x})F_2(d\mathbf{y}). \tag{2.1}$$

2.2 Probability density

2.2.1 Probability density

It can be shown (see chapter 1 in Sato (1999)) that if X_1 and X_2 are independent random variables on $\mathbb{R}^d$ with distributions F_1 and F_2 respectively, then $X_1 + X_2$ has the distribution $F_1 * F_2$.

Definition 2.2.1: Probability density

A probability measure $\mathbb{P}$ on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ is said to have a *probability density* (or just *density*) f if f is a non-negative measurable function on $\mathbb{R}^n$ such that $\mathbb{P}(A) = \int_A f(\mathbf{x})d\mathbf{x}$ [2] for all $A \in \mathscr{B}(\mathbb{R}^n)$.

If a probability measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ has a density, then this density is uniquely determined up to a null-set, as stated in the following theorem:

Theorem 2.2.2

A non-negative Borel-measurable function f is the density of a probability measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ if and only if it satisfies $\int_{\mathbb{R}^n} f(\mathbf{x})\,\mathrm{d}\mathbf{x} = 1$. In this case f entirely determines the probability measure. That is, for any other non-negative Borel measurable function f', if $m^n(f \neq f') = 0$ then f' is also a density for the same probability measure.
Conversely, a probability measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ determines its density (when a density exists) up to a set of Lebesgue measure zero. That is, if f and f' are two densities for this probability, then $m^n(f \neq f') = 0$.

A proof is found in chapter 12 in Jacod and Protter (2004).

We shall sometimes denote as a random vector $\mathbf{X} = \left(X_1, \ldots, X_d\right)^T$, where $\mathbf{X} \in \mathbb{R}^d$ and each $X_k, k = 1, \ldots, d$, is a $\mathbb{R}$-valued random variable. We shall say that $\mathbb{P}_{\mathbf{X}}$, i.e. the distribution of $\mathbf{X}$, is the *joint distribution* of $(X_1, \ldots, X_d)^T$. Conversely, $\mathbb{P}_{X_1}, \ldots, \mathbb{P}_{X_d}$ will be referred to as the *marginal* probability distributions of $\mathbb{P}_{\mathbf{X}}$.

Two families of random vectors $\{\mathbf{X}_t\}$ and $\{\mathbf{Y}_s\}$ are said to be independent if, for any choice of $t_1, \ldots, t_n$ and $s_1, \ldots, s_m$, the random vectors $(X_{t_j})_{j=1,\ldots,n}$ and $(Y_{s_k})_{k=1,\ldots,m}$ are independent.

[2] The Lebesgue measure m on $(\mathbb{R}, \mathscr{B}(\mathbb{R}) \bigcup \{\text{all subsets of nullsets}\})$ measures an interval as its length. The Lebesgue measure m^n is the completion of the n-fold product of m with itself on $\bigotimes_{j=1}^n \mathscr{B}(\mathbb{R})$, i.e. $m^n(A_1 \times \cdots \times A_n) = \prod_{j=1}^n m(A_j)$ for $A_j \in \mathscr{B}(\mathbb{R})$.Since $\mathbb{R}$ is a separable space (see proposition 1.5 in Folland (1999)) we have that $\bigotimes_{j=1}^n \mathscr{B}(\mathbb{R}) = \mathscr{B}(\mathbb{R}^n)$.

Theorem 2.2.3

Let $\mathbf{X} = (Y, Z)$ be a random vector on $\mathbb{R}^2$ with a density f. Then

(a) Both the components Y and Z have densities on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, given by:

$$f_Y(y) = \int_{\mathbb{R}} f(y,z)\,\mathrm{d}z; \quad f_Z(z) = \int_{\mathbb{R}} f(y,z)\,\mathrm{d}y.$$

(b) Y and Z are independent if and only if

$$f(y,z) = f_Y(y) f_Z(z)$$

for all $(y,z) \in \mathbb{R}^2 \setminus E$, where E is an m^2-null set.

This theorem can be generalized to $\mathbb{R}^n, n = 3, 4, \ldots$ (see chapter 12 in Jacod and Protter (2004)).

2.3 Stochastic process

2.3.1 Stochastic process on a euclidean space

Definition 2.3.1: Stochastic process

A family $\{X_t : t \geq 0\}$ of probability distributions on $\mathbb{R}^d$ with parameter $t \in [0, \infty)$, defined on a common probability space, is called a *stochastic process*. It is written as $\{X_t\}$.

It can be shown (see chapter 1 in Sato (1999)) that, for any fixed $\{0 \leq t_1 < t_2 < t_n\}$,

$$\mathbb{P}\left(X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n\right)$$

determines a probability measure on $\mathscr{B}((\mathbb{R}^d)^n)$. The family of probability measures over all choices of n and $t_1, \ldots, t_n$ is called the *system of finite-dimensional distributions* of $\{X_t\}$.

Definition 2.3.2: Cylinder set and Kolmogorov σ-algebra

Let $\Omega = (\mathbb{R}^d)^{[0,\infty)}$. Let ω be the collection of all functions $\omega = (\omega(t))_{t\in[0,\infty)}$ from $[0,\infty)$ into $\mathbb{R}^d$. Define X_t by $X_t(\omega) = \omega(t)$. A set

$$C = \{\omega : X_{t_1}(\omega) \in B_1, X_{t_2}(\omega) \in B_2, \ldots, X_{t_n}(\omega) \in B_n\}$$

for $0 \le t_1 < \cdots < t_n$ and $B_1, \ldots, B_n \in \mathscr{B}(\mathbb{R}^d)$ is called a *cylinder set*.

Let $\mathscr{F}$ be the σ-algebra generated by the cylinder sets. Then $\mathscr{F}$ is called the *Kolmogorov σ-algebra*.

The following theorem by Kolmogorov ensures that a suitable "consistent" system of finite-dimensional distributions will define a stochastic process.

Theorem 2.3.3: Kolmogorov's extension theorem

Suppose that, for any choice of n and $0 \le t_1 < \cdots < t_n$, a distribution $F_{t_1,\ldots,t_n}$ is given. Suppose further that, if $B_1, \ldots, B_n \in \mathscr{B}(\mathbb{R}^d)$ and $B_k = \mathbb{R}^d$, then

$$F_{t_1,\ldots,t_n}\left(B_1 \times \cdots \times B_n\right) = F_{t_1,\ldots,t_{k-1},t_{k+1},\ldots t_n}(B_1 \times \cdots \times B_{k-1} \times B_{k+1} \cdots \times B_n).$$

Then there exists a unique probability measure $\mathbb{P}$ on $\mathscr{F}$ that has $\left\{F_{t_1,\ldots,t_n}\right\}$ as its system of finite-dimensional distributions.
The theorem is stated on page 4 in Sato (1999). A proof can be found on page 489 in Billingsley (1986).

A stochastic process $\{\mathbf{Y}_t : t \ge 0\}$ on the probability space $(\Omega, \mathscr{F}, P)$ is called a *modification* of the stochastic process $\{X_t : t \ge 0\}$ on the same probability space, if $P(X_t = Y_t) = 1$ for $t \in [0,\infty)$.
Two stochastic processes $\{X_t\}$ and $\{Y_t\}$ are *identical in law*, written as

$$\{X_t\} \stackrel{d}{=} \{Y_t\},$$

if the systems of their finite-dimensional distributions are identical. Considered as a function of t, $X(t,\omega)$ is called a *sample function*, or *sample path*, of $\{X_t\}$.

2.3.2 Càdlàg processes

When we get to chapter 4 we shall see that stochastic continuity and the càdlàg property are two of the defining properties of a Lévy process. In this section we define these concepts and introduce a measure for the discontinuities (jumps) of càdlàg processes (stochastic processes with the càdlàg property).

Definition 2.3.4: Stochastic continuity

A stochastic process $\{\mathbf{X}_t : t \geq 0\}$ on a probability space $(\mathbb{R}^d, \mathscr{F}, \mathbb{P})$ is said to be *stochastically continuous* if, for every $t \geq 0$ and every $\epsilon > 0$,

$$\lim_{s \to t} \mathbb{P}(|\mathbf{X}_s - \mathbf{X}_t| \geq \epsilon) = 0. \tag{2.2}$$

'

Definition 2.3.5: Càdlàg

Let $\mathbf{X}_t$ be a stochastic process on the probability space $(\mathbb{R}^d, \mathscr{F}, \mathbb{P})$. We say that $\mathbf{X}_t$ has the *càdlàg* property if there exists $\Omega_0 \in \mathscr{F}$ with $\mathbb{P}(\Omega_0) = 1$ such that, for every $\omega \in \Omega_0, X_t(\omega)$ is right-continuous in $t \geq 0$ and has left limits in $t > 0$.

Definition 2.3.6: Jump times of a càdlàg process

Let $\mathbf{X}$ be a stochastic process with the càdlàg, property. For a given time t we shall denote the left limit $\lim_{s \uparrow t} \mathbf{X}_s$, by $\mathbf{X}_{t-}$ and the difference $\mathbf{X}_t - \mathbf{X}_{t-}$ by $\Delta\mathbf{X}_t$. For a given time interval (a, b) we shall call the set $\{t \in (a, b) : \Delta\mathbf{X}_t \neq 0\}$ the *jump times* of $\{\mathbf{X}_t : t \geq 0\}$ in (a, b).

For a càdlàg process $\{X_t : t \geq 0\}$ on $\mathbb{R}^d$ we introduce a measure $J_{\mathbf{X}}$. For every Borel measurable set $A \in \mathbb{R}^d$, $J_{\mathbf{X}}([t_1, t_2] \times A)$ counts the number of jump times of $\mathbf{X}_t$ between t_1 and t_2 with jump sizes in A.

2.4 Characteristic functions

The principal analytical tool in this thesis is the Fourier transform, which in the statistical community is known under the name *characteristic function*.

Definition 2.4.1: Characteristic function

The *characteristic function* of a probability measure F on $\mathbb{R}^d$ is defined as

$$\widehat{F}(\mathbf{u}) := \int_{\mathbb{R}^d} e^{i\langle \mathbf{u}, \mathbf{x} \rangle} F(\mathrm{d}\mathbf{x}).$$

Definition 2.4.2

The characteristic function of the distribution $\mathbb{P}_X$ of a random variable X on $\mathbb{R}^d$ is defined as

$$\widehat{\mathbb{P}}_{\mathbf{X}}(\mathbf{u}) := \int_{\mathbb{R}^d} e^{i\langle \mathbf{u}, \mathbf{x}\rangle} \mathbb{P}_{\mathbf{X}}(\mathrm{d}\mathbf{x}) = \mathbb{E}\left[e^{i\langle \mathbf{u}, \mathbf{X}\rangle}\right]. \tag{2.3}$$

It follows immediately from definition 2.4.2 that if X is a random vector on $\mathbb{R}^d$, a is a real constant and $\mathbf{b} \in \mathbb{R}^d$ is a constant vector, then

$$\widehat{\mathbb{P}}_{a\mathbf{X}+\mathbf{b}}(\mathbf{u}) = e^{i\langle \mathbf{u}, \mathbf{b}\rangle} \widehat{\mathbb{P}}_{\mathbf{X}}(a\mathbf{u}). \tag{2.4}$$

Theorem 2.4.3

The following theorem sums up some of the most important properties of characteristic functions.
Let F and $F_1, F_2, \ldots, F_n$ be distributions on $\mathbb{R}^d$.

(i) (Bochner's theorem) Then $\widehat{F}(\mathbf{0}) = 1$ and $|\widehat{F}(\mathbf{u})| \leq 1$. Also $\widehat{F}(\mathbf{u})$ is uniformly continuous nonnegative-definite in the sense that, for each $n = 1, 2, \ldots$,

$$\sum_{j=1}^{n}\sum_{k=1}^{n} \widehat{F}(\mathbf{u}_j - \mathbf{u}_k) z_j \bar{z}_k \geq 0 \quad \text{for all } \mathbf{u}_1, \ldots \mathbf{u}_n \in \mathbb{R}^d,\ z_1, \ldots z_n \in \mathbb{C}.$$

Conversely, if a complex-valued function $\varphi(\mathbf{u})$ on $\mathbb{R}^d$ with $\varphi(\mathbf{0}) = 1$ is continuous at $\mathbf{u} = \mathbf{0}$ and is nonnegative-definite, then $\varphi(\mathbf{u})$ is the characteristic function of a distribution on $\mathbb{R}^d$.

(ii) If $\widehat{F_1}(\mathbf{u}) = \widehat{F_2}(\mathbf{u})$ for all $u \in \mathbb{R}^d$ then $F_1 = F_2$.

(iii) If $F = F_1 * F_2$, then $\widehat{F}(\mathbf{u}) = \widehat{F_1}(\mathbf{u})\widehat{F_2}(\mathbf{u})$ for all $\mathbf{u} \in \mathbb{R}^d$.
If X_1 and X_2 are independent random vectors on $\mathbb{R}^d$ then

$$\widehat{\mathbb{P}}_{X_1+X_2}(\mathbf{u}) = \widehat{\mathbb{P}}_{X_1}(\mathbf{u})\widehat{\mathbb{P}}_{X_2}(\mathbf{u}) \quad \text{for all } \mathbf{u} \in \mathbb{R}^d. \tag{2.5}$$

(iv) Let $\mathbf{X} = (\mathbf{X}_j, \dots, \mathbf{X}_n)$ be an $\mathbb{R}^{nd}$-valued random vector, where $\mathbf{X}_1, \dots, \mathbf{X}_n$ are $\mathbb{R}^d$-valued random vectors. Then $\mathbf{X}_1, \dots, \mathbf{X}_n$ are independent if and only if

$$\widehat{\mathbb{P}}_{\mathbf{X}}(\mathbf{u}) = \widehat{\mathbb{P}}_{\mathbf{X}_1}(\mathbf{u}_1) \dots \widehat{\mathbb{P}}_{\mathbf{X}_n}(\mathbf{u}_n) \quad \text{for all } \mathbf{u} = (\mathbf{u}_1, \dots, \mathbf{u}_n),$$

where $\mathbf{u}_j \in \mathbb{R}^d$ for $j = 1, \dots, n$.

(v) Let n be a positive integer. If F has a finite absolute moment of order n, that is if $\int |\mathbf{x}|^n F(d\mathbf{x}) < \infty$, then $\widehat{F}(\mathbf{u})$ is a function of class C^n (continuous n-th derivative) and, for any nonnegative integers $n_1, n_2, \dots n_d$ satisfying $n_1 + \dots + n_d \leq n$,

$$\int x_1^{n_1} \dots x_d^{n_d} F(d\mathbf{x}) = \left[\left(\frac{1}{i} \frac{\partial}{\partial u_1} \right)^{n_1} \cdots \left(\frac{1}{i} \frac{\partial}{\partial u_d} \right)^{n_d} \widehat{F}(\mathbf{u}) \right]_{\mathbf{u}=\mathbf{0}}.$$

(vi) Let n be a positive even integer. If $\widehat{F}(\mathbf{u})$ is of class C^n in a neighborhood of the origin, then F has finite absolute moment of order n.

(vii) (Inversion formula) Let $-\infty < a_j < b_j < \infty$ for $j = 1, \dots, d$ and $B = [a_1, b_1] \times \cdots \times [a_d, b_d]$. If B is an F-continuity set, [3]then

$$F(B) = \lim_{c \to \infty} (2\pi)^{-d} \int_{[-c,c]^d} \widehat{F}(\mathbf{u}) \int_B e^{-i\langle \mathbf{u}, \mathbf{x} \rangle} \, d\mathbf{x} \, d\mathbf{u}$$

(viii) If $\int \left| \widehat{F} \right| d\mathbf{u} < \infty$, then F is absolutely continuous [4]with respect to the Lebesgue measure and has a bounded continuous density $f(\mathbf{x})$, where

$$f(\mathbf{x}) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle \mathbf{u}, \mathbf{x} \rangle} \widehat{F}(\mathbf{u}) \, d\mathbf{u}.$$

Proof: On page 10 in Sato (1999) there is a reference to where proofs can be found.

[3] We define the boundary of a set $B \in \mathbb{R}^d$ as the difference between the smallest closed set in $\mathbb{R}^d$ containing B and the biggest open set in $\mathbb{R}^d$ contained in B. We say that B is a F-*continuity* set if the boundary of B has F-measure 0.

[4] Let P and Q be two finite measures on $(\Omega, \mathscr{F})$. We say that Q is *absolutely continuous* with respect to P if, for any $A \in \mathscr{F}$, $P(A) = 0$ implies $Q(A) = 0$.

When F is a distribution on $[0,\infty)$, the *Laplace* transform of F is defined by

$$L_F(u) = \int_{[0,\infty)} e^{-ux} F(\mathrm{d}x) \quad \text{for } u \geq 0. \tag{2.6}$$

Proposition 2.4.4

Let F, F_1, and F_2 be distributions on $[0,\infty)$.

(i) If $L_{F_1}(u) = L_{F_2}(u)$ for all $u \geq 0$, then $F_1 = F_2$.

(ii) If $F = F_1 * F_2$, then $L_F(u) = L_{F_1}(u) L_{F_2}(u)$.

Proof: For a proof see proposition 2.6 in Sato (1999).

Lemma 2.4.5

Suppose that $\phi(\mathbf{u})$ is a continuous function from $\mathbb{R}^d$ into $\mathbb{C}$ such that $\phi(\mathbf{0}) = 1$ and $\phi(\mathbf{u}) \neq 0$ for any $\mathbf{u} \in \mathbb{R}^d$. Then there is a unique continuous function $f(\mathbf{u})$ from $\mathbb{R}^d$ into $\mathbb{C}$ such that $f(\mathbf{0}) = \mathbf{0}$ and $e^{f(\mathbf{u})} = \phi(\mathbf{u})$. Also for any positive integer n there is a unique continuous function $g_n(\mathbf{u})$ from $\mathbb{R}^d$ into $\mathbb{C}$ such that $g_n(\mathbf{0}) = 1$ and $\left[g_n(\mathbf{u})\right]^n = \phi(\mathbf{u})$. f and g_n have the relation $g_n(\mathbf{u}) = e^{f(\mathbf{u})/n}$.

Proof: For a proof see lemma 7.6 in Sato (1999).

We write $f(\mathbf{u}) = \log \phi(\mathbf{u})$ and $g_n(\mathbf{u}) = [\phi(\mathbf{u})]^{1/n}$. We call f and g_n the *distinguished logarithm* and the *distinguished nth root* of ϕ, respectively. For all $r \geq 0$, $[\phi(\mathbf{u})]^r = e^{rf(\mathbf{u})}$. We call $e^{rf(\mathbf{u})}$ the *distinguished rth power* of ϕ. If $\left[\widehat{F}\mathbf{u}\right]^r$ is the characteristic function of a probability measure, then we denote this probability measure by F^r.

3
Infinitely divisible distributions

3.1 Definitions and basic examples

This chapter discusses a class of probability distributions known as *infinitely divisible distributions*. We begin by presenting some general results about all distributions in this class before we go on to discuss a sub-class known as α-stable distributions.

Definition 3.1.1

A probability measure F on $\mathscr{B}(\mathbb{R}^d)$ is *infinitely divisible* if for any integer $n \geq 2$ there exist n i.i.d. non-trivial random variables $\mathbf{Y}_1^{(n)}, \ldots \mathbf{Y}_n^{(n)}$ such that $\mathbf{Y}_1^{(n)} + \cdots + \mathbf{Y}_n^{(n)}$ has the distribution F.

Let us now consider an alternative definition of infinitely divisible distributions. The following is shown in Cont and Tankov (2004) and uses the fact that the distribution of sums of i.i.d. variables is given by the convolution of the distribution of the summands. For any $n \geq 2$ let F_n be the distribution of each of the above $\mathbf{Y}_1^{(1)}, \ldots, \mathbf{Y}_n^{(n)}$. Then the n-th convolution of F_n, namely $F_n * \cdots * F_n$ n times, is equal to F.

Therefore an infinitely divisible distribution can also be defined as a distribution F for which, for any $n \geq 2$, there exists a probability measure F_n on $\mathscr{B}\left(\mathbb{R}^d\right)$ such that F is equal to the n-th convolution of F_n with itself:

$$F = F_n^{(n*)}.$$

Example 3.1.2

For $d \geq 2$ let F be the nondegenerate [1]Gaussian distribution on $\mathbb{R}^d$ with mean vector $\boldsymbol{\mu}$ and covariance matrix Q, where Q is a symmetric non-negative definite, invertible matrix . Then F has the probability density

$$(2\pi)^{-d}(\det Q)^{-1/2}e^{-\langle \mathbf{x}-\boldsymbol{\mu}, Q^{-1}(\mathbf{x}-\boldsymbol{\mu})\rangle/2} \tag{3.1}$$

for all $\mathbf{x} \in \mathbb{R}^d$. It can then be shown (see chapter 2 in Sato (1999)) that

$$\widehat{F}(\mathbf{u}) = \exp\left\{-\frac{1}{2}\langle \mathbf{u}, Q\mathbf{u}\rangle + i\langle \mathbf{u}, \boldsymbol{\mu}\rangle\right\}, \quad \mathbf{u} \in \mathbb{R}^d. \tag{3.2}$$

By choosing F_n to be the Gaussian distribution on $\mathbb{R}^d$ with mean vector $\frac{1}{n}\boldsymbol{\mu}$ and covariance matrix $\frac{1}{n}Q$, we trivially get that the Gaussian distribution is infinitely divisible.

The Gaussian distribution owes much of its importance in statistics to its large domain of attraction, formalized in the Central Limit Theorem below.

Theorem 3.1.3: Central Limit Theorem

Let $\mathbf{S}_n$ be a random walk on $\mathbb{R}^d$. Here each i.i.d. step $\mathbf{X}_j, = \left(X_j^{(1)}, \ldots, X_j^{(d)}\right)^T$ has

(a) a finite mean vector $\boldsymbol{\mu}$ and

(b) a finite covariance matrix $Q = \left(q_{k,l}\right)$ with $k, l = 1, \ldots, d$.

Here $q_{k,l} = \mathrm{Cov}\left(X_j^{(k)}, X_j^{(l)}\right)$, where $X_j^{(k)}$ and $X_j^{(l)}$ are the k-th and l-th components of the $\mathbb{R}^d$-valued random variable $\mathbf{X}_j$. Then

$$\left\{\frac{\mathbf{S}_n - n\boldsymbol{\mu}}{\sqrt{n}}\right\} \xrightarrow{d} \mathbf{Z},$$

where $\mathbf{Z}$ is a d-dimensional Gaussian-distributed random vector with mean vector $\mathbf{0}$ and covariance matrix Q.

Proof: A proof can be found on page 238 in Breiman (1968).

[1]A Gaussian distribution is called *degenerate* if the covariance matrix Q is singular, i.e. $\det Q = 0$.

Example 3.1.4

Let F be the d-dimensional Cauchy distribution with parameters $\gamma \in \mathbb{R}^d$ and $c > 0$. That is, let F have the density

$$\Gamma((d+1)/2)\pi^{-(d+1)/2}\left((|\mathbf{x}-\gamma|^2)+c^2\right)^{-(d+1)/2} \quad \text{for } \mathbf{x} \in \mathbb{R}^d. \tag{3.3}$$

It can be shown (see page 11 in Sato (1999)) that the Cauchy distribution has the characteristic function

$$\widehat{F}(\mathbf{u}) = e^{-c|\mathbf{u}|+i\langle\gamma,\mathbf{u}\rangle}. \tag{3.4}$$

Let F_n be the d-dimensional Cauchy distribution with parameters $\frac{1}{n}\gamma \in \mathbb{R}^d$ and $\frac{1}{n}c > 0$. We then see that $\left[\widehat{F_n}(\mathbf{u})\right]^n = \widehat{F}(\mathbf{u})$, so the Cauchy distribution is infinitely divisible.

Example 3.1.5

A trivial example of a function that is not infinitely divisible is the uniform distribution on (a, b), whose characteristic function is

$$\frac{e^{iub} - e^{iua}}{iu}. \tag{3.5}$$

3.2 The Lévy-Khintchine representation

3.2.1 The formula

The most useful analytical tool for studying infinitely divisible distributions is the characteristic function. This is in large part due to a theorem that says that the characteristic function of every infinitely divisible distribution is of a closed form specified by the Lévy-Khintchine representation.

Theorem 3.2.1

Let $D = \{\mathbf{x} : |\mathbf{x}| \leq 1\}$.

If F is an infinitely divisible distribution on $\mathbb{R}^d$, then there exist Q, ν and γ such that

$$\widehat{F}(\mathbf{u}) = \exp\Big[-\frac{1}{2}\langle \mathbf{u}, Q\mathbf{u}\rangle + i\langle \boldsymbol{\gamma}, \mathbf{u}\rangle + \int_{\mathbb{R}^d} \left(e^{i\langle \mathbf{u},\mathbf{x}\rangle} - 1 - i\langle \mathbf{u}, \mathbf{x}\rangle 1_D(\mathbf{x})\right) \nu\,(\mathrm{d}\mathbf{x}) \Big]. \tag{3.6}$$

Here Q is a symmetric nonnegative-definite $d \times d$ matrix, ν is on $\mathbb{R}^d \setminus \{0\}$ with $\int \min(1, |\mathbf{x}|^2)\nu(\mathrm{d}\mathbf{x}) < \infty$ and $\boldsymbol{\gamma}$ a vector in $\mathbb{R}^d$.

The three parameters Q, ν and $\boldsymbol{\gamma}$ are unique.

Proof: A proof of the one-dimensional case is found on page 192-194 in Breiman (1968).

The triplet $(Q, \nu, \boldsymbol{\gamma})$ is called the *generating triplet* of the infinitely divisible random variable X.

If $\nu(B) = 0$ for any Borel set B and $\boldsymbol{\gamma} = \mathbf{0}$, then the Lévy-Khintchine representation gives the characteristic function of a centered d-variate Gaussian distribution with covariance matrix Q (or the variance Q if $d = 1$). We shall therefore refer to the parameter Q as the *Gaussian coefficient*. We shall refer to ν as the *Lévy measure*.

3.2.2 Drift and center

As shown in Remark 8.4 in Sato (1999) , if $\int_{|\mathbf{x}|\leq 1} |\mathbf{x}|\nu(\mathrm{d}\mathbf{x}) < \infty$, then equation 3.6 can be written as

$$\widehat{F}(\mathbf{u}) = \exp\left[-\frac{1}{2}\langle \mathbf{u}, Q\mathbf{u}\rangle + i\langle \boldsymbol{\gamma}_0, \mathbf{u}\rangle + \int_{\mathbb{R}^d} \left(e^{i\langle \mathbf{u},\mathbf{x}\rangle} - 1\right) \nu\,(\mathrm{d}\mathbf{x}) \right].$$

where
$\boldsymbol{\gamma}_0 \in \mathbb{R}^d$ is defined as

$$\boldsymbol{\gamma}_0 := \boldsymbol{\gamma} - \int_{\mathbb{R}^d} \mathbf{x} 1_{|\mathbf{x}|\leq 1}\nu(\mathrm{d}\mathbf{x}). \tag{3.7}$$

For reasons that will become clear in section 4.2.2 we shall then call $\boldsymbol{\gamma}_0$ the *drift* of $\mathbb{P}$.

Similarly, if $\int_{|\mathbf{x}|>1} |\mathbf{x}|\nu(\mathrm{d}\mathbf{x}) < \infty$

then equation 3.6 on the preceding page can be written as

$$\widehat{F}(\mathbf{u}) = \exp\left[-\frac{1}{2}\langle \mathbf{u}, Q\mathbf{u}\rangle + i\langle \boldsymbol{\gamma}_1, \mathbf{u}\rangle + \int_{\mathbb{R}^d} \left(e^{i\langle \mathbf{u},\mathbf{x}\rangle} - 1 - i\langle \mathbf{u}, \mathbf{x}\rangle\right) \nu(\mathrm{d}\mathbf{x})\right], \tag{3.8}$$

where $\boldsymbol{\gamma}_1 \in \mathbb{R}^d$ is defined as

$$\gamma_1 := \gamma + \int_{\mathbb{R}^d} \mathbf{x} 1_{|\mathbf{x}|>1} \nu(\mathrm{d}\mathbf{x}). \tag{3.9}$$

Let F_j, with $j \in 1, \ldots, d$, denote the marginal distributions on $\mathbb{R}$ of F.

It can then be shown (see Example 25.12 in Sato (1999)) that the condition $\int_{|\mathbf{x}|>1} |\mathbf{x}| \ \nu(\mathrm{d}\mathbf{x}) < \infty$ is equivalent to $\int_{\mathbb{R}^d} |\mathbf{x}| F(\mathrm{d}\mathbf{x}) < \infty$, and that for each $j \in 1, \ldots, d$ the component $\gamma_1^{(j)}$ of $\boldsymbol{\gamma}_1$ is the expectation value of F_j. We shall call $\boldsymbol{\gamma}_1$ the *center* of F.

3.3 Stable distributions

In this section we will look at a family of infinitely divisible distributions known as *stable distributions*, defined in chapter 2 in Samorodnitsky and Taqqu (1994) as follows:

3.3.1 Stability and infinite divisibility

Definition 3.3.1

A random vector $\mathbf{X}$ on $\mathbb{R}^d$ is said to have a *stable* distribution if, for every $a > 0$ and every $b > 0$, there exist a positive number c and a vector $\mathbf{d} \in \mathbb{R}^d$ such that

$$a\mathbf{X}^{(1)} + b\mathbf{X}^{(2)} \stackrel{d}{=} c\mathbf{X} + \mathbf{d}, \tag{3.10}$$

where $\mathbf{X}^{(1)}$ and $\mathbf{X}^{(2)}$ are any i.i.d. random vectors independent of $\mathbf{X}$, but with the same distribution as $\mathbf{X}$.

If, for any $a > 0$ and any $b > 0$, equation 3.10 holds with $\mathbf{d} = \mathbf{0}$, then $\mathbf{X}$ is said to be *strictly* stable.

X is called *symmetric stable* if it is stable and

$$F\{\mathbf{X} \in A\} = F\{-\mathbf{X} \in A\} \tag{3.11}$$

for any Borel set A of $\mathbb{R}^d$, where F is the distribution of **X**.

An alternative and equivalent definition of a stable distribution is the following (see page 69 in Sato (1999)):

Definition 3.3.2

Let F be an infinitely divisible probability measure on $\mathscr{B}\left(\mathbb{R}^d\right)$. F is called *stable* if, for any $a > 0$, there exist $b > 0$ and $\mathbf{c} \in \mathbb{R}^d$ such that

$$[\widehat{F}(\mathbf{u})]^a = \widehat{F}(b\mathbf{u})e^{i\langle \mathbf{c},\mathbf{u}\rangle}. \tag{3.12}$$

It is called *strictly stable* if, for any $a > 0$, there exists $b > 0$ such that

$$[\widehat{F}(\mathbf{u})]^a = \widehat{F}(b\mathbf{u}). \tag{3.13}$$

F is called *symmetric stable* if F is stable, and for any Borel set B of $\mathbb{R}^d$, $F\{-x : x \in B\} = F\{x : x \in B\}$.

Stable distributions are also characterized by the fact that a distribution posseses a domain of attraction (see definition 2.1.6 on page 13) if and only if it is stable (see theorem 1 XVII.5 in Feller (1971).

3.3.2 Index of stability

As stated in the theorem below, any linear combination of the components of a stable distribution is stable.

Theorem 3.3.3

Let $\mathbf{X} = (X_1, \ldots, X_n)^T$ be a non-trivial and stable (respectively, strictly stable, symmetric stable) random vector in $\mathbb{R}^d$. Then there is a constant $\alpha \in (0, 2]$ such that, in equation 3.10 on the preceding page,
$c = (a^\alpha + b^\alpha)^{1/\alpha}$. Moreover, any linear combination of the components of **X** of the type $\sum_{k=1}^d b_k X_k$ is a stable (respectively, strictly stable, symmetric stable) random variable. A proof can be found on page 58 in
Samorodnitsky and Taqqu (1994).

As a corollary of theorem 1 in chapter VI.I in Feller (1971) and theorem 3.3.3 above we have the following:

Corollary 3.3.4

Let $\mathbf{X}$ be a non-trivial random vector and $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(n)}$ be any i.i.d random vectors independent of $\mathbf{X}$, but with the same distribution as $\mathbf{X}$.

Then $\mathbf{X}$ is stable if and only if there exists an $\alpha \in (0,2]$ such that, for any $n \geq 2$, there exists a displacement vector $\mathbf{d}_n$ such that

$$\mathbf{X}^{(1)} + \mathbf{X}^{(2)} + \cdots + \mathbf{X}^{(n)} \overset{d}{=} n^{1/\alpha}\mathbf{X} + \mathbf{d}_n. \tag{3.14}$$

The index α is called the *index of stability*, and a stable distribution with index of stability α is called an α*-stable* distribution.

Similarly, theorem 13.11 and theorem 13.15 in Sato (1999) give that in equation 3.12 on the previous page

$$b = a^{\frac{1}{\alpha}}. \tag{3.15}$$

The α in equation 3.15 is the same as in equation 3.14.

Example 3.3.5: Gaussian distribution (revisited)

It follows from equation 3.2 on page 23 and equation 2.4 on page 19 that if $\mathbf{X}$ is a d-dimensional centered and Gaussian distributed random vector (mean vector $\mathbf{0}$) with covariance matrix Q, and if $\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(n)}$ are n independent copies of $\mathbf{X}$ (n a positive integer), then

$$\begin{aligned}
&\widehat{\mathbb{P}}_{\mathbf{X}^{(1)}+\cdots+\mathbf{X}^{(n)}}(\mathbf{u}) \\
&= \mathbb{E}\left(e^{i\langle \mathbf{u}, \mathbf{X}^{(1)}+\cdots+\mathbf{X}^{(n)}\rangle}\right) \\
&= \left[\mathbb{E}\left(e^{i\langle \mathbf{u}, \mathbf{X}\rangle}\right)\right]^n \\
&= \left[\exp \frac{1}{2}\langle \mathbf{u}, Q\mathbf{u}\rangle\right]^n \\
&= \exp \frac{1}{2}\left[\langle n^{1/2}\mathbf{u}, n^{1/2}Q\mathbf{u}\rangle\right] \\
&= \widehat{\mathbb{P}}_{\mathbf{X}}(n^{1/2}\mathbf{u}) = \widehat{\mathbb{P}}_{n^{1/2}\mathbf{X}}(\mathbf{u}).
\end{aligned}$$

Since two distributions are equal when their characteristic functions are equal we have that

$$\mathbf{X}^{(1)} + \cdots + \mathbf{X}^{(n)} \overset{d}{=} n^{1/2}\mathbf{X}.$$

Therefore a centered Gaussian distribution is a strictly stable distribution with index 2.

Now let us consider the general Gaussian distribution. Let $\mathbf{Y}^{(k)} := \mathbf{X}^{(k)} + \boldsymbol{\mu}$ where $\boldsymbol{\mu} \in \mathbb{R}^d$. From the above the general Gaussian distribution can be expressed as follows:

$$\mathbf{Y}^{(1)} + \cdots + \mathbf{Y}^{(n)} \stackrel{d}{=} \mathbf{X}^{(1)} + \cdots + \mathbf{X}^{(n)} + n\boldsymbol{\mu} \stackrel{d}{=} n^{1/2}\mathbf{X} + n\boldsymbol{\mu}.$$

Therefore a general Gaussian distribution with mean $\boldsymbol{\mu}$ is stable with index $\alpha = 2$ and displacement vector $\mathbf{d}_n = n\boldsymbol{\mu}$.

The following argument shows that all non-trivial stable distributions with finite covariace matrices are Gaussian: Suppose that $\mathbf{X}$ is a stable random vector with a finite covariance matrix $V \neq \mathbf{0}$. From the above (a) the sum of $n \geq 2$ independent copies of $\mathbf{X}$ has a finite covariance matrix $nV \neq \mathbf{0}$ and
(b) $n^{1/\alpha}\mathbf{X}$ has the finite covariance matrix $n^{2/\alpha}V \neq \mathbf{0}$. Because of the stability property of $\mathbf{X}$ we must have $nV = n^{2/\alpha}V$, which implies $\alpha = 2$. It follows that a non-deterministic stable random vector with finite covariance must have index 2. Because of the central limit theorem (theorem 3.1.3 on page 23) $\mathbf{X}$ must be Gaussian distributed.Thus the Gaussian distribution is the only non-trivial stable distribution with finite covariance matrix. On page 77 in Sato (1999) it is also proven that if a stable distribution has index 2 then it is Gaussian.

Example 3.3.6: Cauchy distribution (revisited)

From equation 3.4 on page 24 it follows that if $\mathbf{X}$ is a Cauchy distributed random vector and $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ are independent copies of $\mathbf{X}$, then

$$\begin{aligned}
&\widehat{\mathbb{P}}_{\mathbf{X}^{(1)}+\cdots+\mathbf{X}^{(n)}}(\mathbf{u}) \\
&= \mathbb{E}\left(e^{i\langle \mathbf{u}, \mathbf{X}^{(1)}+\cdots+\mathbf{X}^{(n)}\rangle}\right) \\
&= \left[\mathbb{E}\left(e^{i\langle \mathbf{u}, \mathbf{X}\rangle}\right)\right]^n \\
&= \left[e^{-c|\mathbf{u}|+i\langle \boldsymbol{\gamma}, \mathbf{u}\rangle}\right]^n \\
&= e^{-c|n\mathbf{u}|+i\langle \boldsymbol{\gamma}, n\mathbf{u}\rangle} \\
&= \widehat{\mathbb{P}}_{\mathbf{X}}(n\mathbf{u}).
\end{aligned}$$

Combined with equation 2.4 on page 19 and item ii on page 19 we have that

$$\mathbf{X}^{(1)} + \cdots + \mathbf{X}^{(n)} \stackrel{d}{=} n\mathbf{X}.$$

Thus the Cauchy distribution is a stable distribution with index 1.

In the remaining discussion of stable distributions we shall assume that the random variables involved are all non-trivial.

3.3.3 Stable linear combinations, unstable joint distribution

The converse of theorem 3.3.3 on page 27 is not true. There exist random vectors that are not stable although any linear combination of their components is stable. An example of such an non-stable random vector is given in chapter 2 in Samorodnitsky and Taqqu (1994).

If some additional conditions are satisfied, however, stability of a random vector can be inferred.

Theorem 3.3.7: Conditions for joint stability

(a) If all linear combinations $Y = \sum_{k=1}^{d} b_k X^{(k)}$ have strictly stable distributions, then $\mathbf{X}$ is a strictly stable random vector in $\mathbb{R}^d$.

(b) If all linear combinations are symmetric stable, then $\mathbf{X}$ is a symmetric stable random vector in $\mathbb{R}^d$.

(c) If all linear combinations are stable with index of stability greater than or equal to one, then $\mathbf{X}$ is a stable random vector in $\mathbb{R}^d$.

(d) If all linear combinations are stable and $\mathbf{X}$ is infinitely divisible then $\mathbf{X}$ is stable.

Proof: Proofs of the assertions above except the last one can be found on page 59 in Samorodnitsky and Taqqu (1994). A reference to a proof of item d is listed on page 65 in Samorodnitsky and Taqqu (1994).

3.4 Characteristic function of a stable distribution

In this section we state some results for the characteristic function of an α-stable random vector.

In addition to the Lévy-Khintchine -representation, another popular representation of the characteristic function of an α-stable distribution is by means of a finite measure on a unit sphere $S_d \in \mathbb{R}^d$. This representation stems from the following theorem:

Theorem 3.4.1: Characteristic function of a stable distribution

F is an α-stable distribution on $\mathbb{R}^d$ with $0 < \alpha < 2$ if and only if there exists a finite non-zero measure Γ on $\mathbb{R}^d$ and a vector $\boldsymbol{\tau}$ in $\mathbb{R}^d$ such that

(i) If $\alpha \neq 1$

$$\widehat{F}(\mathbf{u}) = \exp\left[-\int_{S_d} |\langle \mathbf{u}, \mathbf{s}\rangle|^{\alpha} \left(1 - i \tan\frac{\pi\alpha}{2} \operatorname{sign}(\langle \mathbf{u}, \mathbf{s}\rangle)\right) \Gamma(\mathrm{d}\mathbf{s}) + i\langle \mathbf{u}, \boldsymbol{\tau}\rangle\right]. \tag{3.16}$$

If $\mathbf{X}$ is α-stable with $\alpha \neq 1$ then $\mathbf{X}$ is strictly α-stable if and only if $\boldsymbol{\tau} = \mathbf{0}$.

(ii) If $\alpha = 1$ then

$$\widehat{F}(\mathbf{u}) = \exp\left[-\int_{S_d} |\langle \mathbf{u}, \mathbf{s}\rangle| \left(1 + i\frac{2}{\pi}(\langle \mathbf{u}, \mathbf{s}\rangle) \ln|\langle u, \mathbf{s}\rangle|\right) \Gamma(\mathrm{d}\mathbf{s}) + i\langle \mathbf{u}, \boldsymbol{\tau}\rangle\right]. \tag{3.17}$$

If F is 1-stable then F is strictly 1-stable if and only if

$$\int_{S_d} s_k \Gamma(\mathrm{d}s) = 0 \quad \text{for } k = 1, 2, \ldots, d.$$

For both cases the following hold: (a) The measure Γ and the vector $\boldsymbol{\tau}$ above are unique. (b) F is a symmetric α-stable distribution on $\mathbb{R}^d$ if and only if Γ is a symmetric measure, i.e. $\Gamma(B) = \Gamma(-B)$ for all $B \in \mathscr{B}(\mathbb{R}^d)$, and $\boldsymbol{\tau} = \mathbf{0}$.

Proof: This is a combination of theorem 2.3.1,theorem 2.4.1 and 2.4.3 in Samorodnitsky and Taqqu (1994).

3.4.1 Simplified Lévy measure

Theorem 3.4.2

A probability measure F on $\mathbb{R}^d$ is α-stable with $0 < \alpha < 2$ if and only if it is infinitely divisible with generating triplet $(0, \nu, \gamma)$ and there exists a finite measure λ on S_d, a unit sphere of $\mathbb{R}^d$, such that

$$\nu(B) = \int_{S_d} \lambda(\mathrm{d}\xi) \int_0^\infty 1_B(r\xi) \frac{\mathrm{d}r}{r^{1+\alpha}}. \tag{3.18}$$

Further, if $\alpha = 1$, then F is strictly 1-stable if and only if either F is 1-stable, $\nu \neq 0$, and the measure λ in equation 3.18 satisfies

$$\int_{S_d} \xi \lambda \mathrm{d}\xi = 0,$$

or $Q = 0, \nu = 0$, and $\gamma \neq 0$.

Proof: For a proof of the first assertion in this theorem see page 78 in Sato (1999). For a proof of the second (involving $\alpha = 1$) see Remark 14.6 and theorem 14.7 in Sato (1999).

In the one-dimensional case $S_d = \{-1, 1\}$, so equation 3.18 reduces to (see page 80 in Sato (1999))

$$\nu(x) = \frac{A}{x^{\alpha+1}} 1_{x>0} + \frac{B}{|x|^{\alpha+1}} 1_{x<0}, \tag{3.19}$$

where A and B are positive constants.

3.4.2 Probability densities of stable distributions

Before going on to Lévy processes in the next chapter, we remark the following. Except for the Gaussian distribution and the Cauchy distribution there is unfortunately only one known case where the probability density of a stable distribution can be expressed analytically. That case is the Lévy distribution $S_{\frac{1}{2}}(0, 1, \mu)$ (see chapter 3 in Cont and Tankov (2004)) which has index $\frac{1}{2}$ and the following probability density:

$$\left(\frac{\sigma}{2\pi}\right)^{1/2} \frac{1}{(x-\mu)^{3/2}} \exp\left\{-\frac{\sigma}{2(x-\mu)}\right\} 1_{x>\mu}. \tag{3.20}$$

Furthermore it can be shown that the Lévy distribution above has the Lévy measure $\nu(x) = \frac{\sigma}{2\sqrt{\pi}} \frac{1}{x^{3/2}} 1_{x>0}$.

4
Lévy processes

4.1 Basic properties and one example

Definition 4.1.1

A stochastic process $\{\mathbf{X}_t, t \geq 0\}$ on the probability space $(\mathbb{R}^d, \mathscr{F}, \mathbb{P})$ is a Lévy process if the following conditions are satisfied:

- For any choice of $n \geq 1$ and $0 \leq t_o \leq t_1 \cdots \leq t_n$ the random vectors $\mathbf{X}_{t_0}, \mathbf{X}_{t_1} - \mathbf{X}_{t_0}, \ldots, \mathbf{X}_{t_n} - \mathbf{X}_{t_{n-1}}$ are independent (independent increments property).
- $\mathbf{X}_0 = 0$ a.s.
- The distribution of $\mathbf{X}_{s+t} - \mathbf{X}_s$ does not depend on s (stationary increments property)
- $\mathbf{X}_t$ is stochastically continuous.

- There exists $\Omega_0 \in \mathscr{F}$ with $\mathbb{P}(\Omega_0) = 1$ such that, for every $\omega \in \Omega_0$, $X_t(\omega)$ is right-continuous in $t \geq 0$ and has left limits in $t > 0$. (The "càdlàg" property.)

In the following the σ-algebra $\mathscr{F}$ will be on $\mathscr{B}(\mathbb{R}^d)$ (the Borel sets in $\mathbb{R}^d$).

4.1.1 Lévy processes are infinitely divisible

Looking at an arbitrary Lévy process $(\mathbf{X}, \mathbb{P})$ and using the decomposition

$$\mathbf{X}_1 = \mathbf{X}_{1/n} + (\mathbf{X}_{2/n} - \mathbf{X}_{1/n}) \cdots + (\mathbf{X}_{n/n} - \mathbf{X}_{n-1/n}),$$

we see that the distribution $\mathbb{P}(\mathbf{X}_1 \in \boldsymbol{.})$ is infinitely divisible. A similar decomposition can be done (see Bertoin (1996)) to show that, for any rational number $t \geq 0$, the probability measure $P(\mathbf{X}_t \in \boldsymbol{.})$ is infinitely divisible as well, and

$$\widehat{\mathbb{P}}_{\mathbf{X}_t}(\mathbf{u}) = \mathbb{E}\left(e^{i\langle \mathbf{u}, \mathbf{X}_t \rangle}\right) = \left[\mathbb{E}\left(e^{i\langle \mathbf{u}, \mathbf{X}_1 \rangle}\right)\right]^t = \left[\widehat{\mathbb{P}}_{\mathbf{X}_1}(\mathbf{u})\right]^t, \quad \text{where } \mathbf{u} \in \mathbb{R}^d. \tag{4.1}$$

4.1.2 Infinitely divisible distributions viewed as Lévy processes

Example 4.1.2

Poisson process:

Let $c > 0$. The Poisson distribution with mean c is defined by the probability measure

$$F(\{k\}) = \frac{c^k}{k!} e^{-c} \quad \text{for } k \in 0, 1, 2, \ldots,$$

where $F(B) = 0$ for any B not containing a non-negative integer.
We have that

$$\widehat{F}(u) = \exp\left\{c(e^{iu} - 1)\right\}, \quad \text{where } u \in \mathbb{R},$$

so the Poisson distribution is obviously infinitely divisible.

Definition 4.1.3

A stochastic process $\{X_t : t \geq 0\}$ on $\mathbb{R}$ is a *Poisson process* if it is a Lévy process and, for $t > 0$, X_t has Poisson distribution with mean ct.

The sample paths of a Poisson process are characterized by the following theorem (see page 308 in Breiman (1968)).

Theorem 4.1.4

A stochastic process $\{X_t : t \geq 0\}$ with stationary, independent increments has a modification with all sample paths constant except for upward jumps of length one if and only if there is a parameter $c \geq 0$ such that $\widehat{\mathbb{P}}_{X_t}(u) = \exp\left\{c(e^{iu}-1)\right\}$ for all $u \in \mathbb{R}$.

From equation 4.1 on the preceding page and theorem 2.4.3 on page 19 we have that such a stochastic process is a Poisson process.

In the proof of theorem 4.1.4 Breiman also shows that the waiting times between the jumps of a Poisson process are exponentially distributed with parameter c, i.e. are gamma distributed with parameters 1 and c.

Definition 4.1.5

A distribution F on $\mathbb{R}^d$ is *compound Poisson* if, for some $c > 0$ and for some distribution ρ on $\mathbb{R}^d$ with $\rho\{0\} = 0$,

$$\widehat{F}(\mathbf{u}) = \exp\{c(\widehat{\rho}(\mathbf{u})-1)\} \quad \text{for all } \mathbf{u} \in \mathbb{R}^d.$$

Here $\widehat{\rho}(\mathbf{u})$ is the characteristic function of ρ.

Definition 4.1.6

Let $c > 0$ and let ρ be a distribution on $\mathbb{R}^d$ with $\rho(\{0\}) = 0$. A stochastic process $\{\mathbf{X}_t : t \geq 0\}$ on $\mathbb{R}^d$ is a *compound Poisson process* associated with c and ρ if it is a Lévy processes and, for $t > 0$, the characteristic function of the law of $\mathbf{X}_t$ has the following form:

$$\widehat{\mathbb{P}}_{\mathbf{X}_t}(\mathbf{u}) = \exp\left(tc(\widehat{\rho}(\mathbf{u})-1)\right) \quad \text{for all } \mathbf{u} \in \mathbb{R}^d. \tag{4.2}$$

Here c and ρ are uniquely determined by $\mathbf{X}_t$.

Any compound Poisson process can be constructed from a Poisson process and a random walk in the following way:

Theorem 4.1.7

Let $\{N_t : t \geq 0\}$ be a Poisson process with parameter $c > 0$ and let $\{\mathbf{S}_n : n \in 0, 1, \ldots,\}$ be a random walk on $\mathbb{R}^d$ with a common probability space $(\Omega, \mathscr{F}, P)$. Assume that $\{N_t\}$ and that $\{S_n\}$ are independent and $P[\mathbf{S}_1 = \mathbf{0}] = 0$. Define

$$\mathbf{X}_t(\omega) = \mathbf{S}_{N_t(\omega)}(\omega).$$

Then $\{\mathbf{X}_t : t \geq 0\}$ is a compound Poisson process satisfying equation 4.2 on the preceding page, with ρ being the distribution of $\mathbf{S}_1$.

Proof: A proof can be found on page 18 in Sato (1999)

Compound Poisson distributions play a special role in the theory of infinitely divisible distributions, in part because of the following theorem:

Theorem 4.1.8

Every infinitely divisible distribution is the limit of a sequence of compound Poisson distributions.

Proof: For a proof see corollary 8.8 in Sato (1999).

Because of theorem 4.1.8 it is always possible to approach a given infinitely divisible law using a sequence of compound Poisson distributions. Using this idea it is shown in Bertoin (1996) that every infinitely divisible distribution can be viewed as the distribution of a Lévy process evaluated at $t = 1$. This result, combined with the above decomposition of Lévy processes, implies that the classes of infinitely divisible distributions and Lévy processes are equivalent.

Theorem 4.1.9: Infinitely divisible distributions as Lévy processes

Let there be given the following: $\boldsymbol{\gamma} \in \mathbb{R}^d$, a positive nonnegative-definite d-dimensional matrix Q , and a measure ν on $\mathbb{R}^d \setminus \{0\}$ such that $\int \left(\min(1, |\mathbf{x}|^2)\right) \nu(d\mathbf{x}) < \infty$.

For every $\boldsymbol{u} \in \mathbb{R}^d$ let

$$\phi(\boldsymbol{u}) = i\langle \gamma, \boldsymbol{u}\rangle - \frac{1}{2}\langle \mathbf{u}, Q\mathbf{u}\rangle + \int_{\mathbb{R}^d} \left(e^{i\langle \mathbf{u},\mathbf{x}\rangle} - 1 - i\langle \mathbf{u}, \mathbf{x}\rangle 1_{|\mathbf{x}|\leq 1}\right) \nu(\mathrm{d}\mathbf{x}) \qquad (4.3)$$

and let $\widehat{\mathbb{P}}(\boldsymbol{u}) = e^{t\phi(\boldsymbol{u})}$.

Then there exists a a unique probability measure P on $\mathbb{R}^d$ under which the stochastic process $\mathbf{X}_t$ is a Lévy process with characteristic function $\widehat{\mathbb{P}}(\mathbf{u})$. For a proof see page 13 in Bertoin (1996).

The function $\phi(u)$ is called the *characteristic exponent* of $\mathbf{X}_t$.

In the context of Lévy processes we have (see chapter 3 in Cont and Tankov (2004)) that, for a measurable set $A \in \mathbb{R}^d$, the Lévy measure $\nu(A)$ can be interpreted as the expected number, per unit time, of jumps whose size belongs to the set A. I.e.,

$$\nu(A) = \mathbb{E}\left[\#\left\{t \in [0,1] : \Delta \mathbf{X}_t \in A\right\}\right].$$

4.1.3 Vector space property

We refer to the following result from page 65 in Sato (1999). It shows that a linear transformation of a Lévy process is again a Lévy process.

Proposition 4.1.10

Let $\mathbf{X}_t$ be a Lévy process on $\mathbb{R}^d$ with generating triplet (Q, ν, γ) and let U be an $n \times d$ matrix. Then $\mathbf{Y}_t = U\mathbf{X}_t$ is a Lévy process on $\mathbb{R}^d$. Its generating triplet $(Q_\mathbf{Y}, \nu_\mathbf{Y}, \gamma_\mathbf{Y})$ is given as follows:

$$Q_\mathbf{Y} = UQU^T \tag{4.4}$$

$$\nu_\mathbf{Y} = [\nu\, U^{-1}]_{\mathbb{R}^d \setminus \{0\}} \tag{4.5}$$

$$\gamma_\mathbf{Y} = U\gamma + \int U\mathbf{y}\left(1_E(U\mathbf{y}) - 1_D(\mathbf{y})\right)\nu(d\mathbf{x}), \tag{4.6}$$

where $\left(\nu\, U^{-1}\right)(B) = \nu\left(\{\mathbf{x} : U\mathbf{x} \in B\}\right)$, for any $B \in \mathscr{B}(\mathbb{R}^d)$, $D = \left\{\mathbf{x} \in \mathbb{R}^d : |\mathbf{x}| \leq 1\right\}$, and $E = \left\{\mathbf{y} \in \mathbb{R}^d : |\mathbf{y}| \leq 1\right\}$.

4.2 Subclasses of Lévy processes

In this section we will take a closer look at stable Lévy processes and Lévy processes of finite variation.

4.2.1 Stable processes

We shall say that a Lévy process $\mathbf{X}_t$ is a stable, strictly stable or symmetric stable process if the distribution of $\mathbf{X}_t$ at $t = 1$ is respectively stable, strictly stable or symmetric stable.

If $\mathbf{X}$ is a symmetric stable process it follows from equation 3.11 on page 27 that $\{\mathbf{X}_t : t \geq 0\} \stackrel{d}{=} \{-\mathbf{X}_t : t \geq 0\}$.

Definition 4.2.1

Let $\{\mathbf{X}_t : t \geq 0\}$ be a stochastic process on $\mathbb{R}^d$. It is called *self-similar* if , for any $a > 0$, there exists a $b > 0$ such that

$$\{\mathbf{X}_{at}\} \stackrel{d}{=} \{b\mathbf{X}_t\}. \tag{4.7}$$

$\mathbf{X}_t$ is called *broad-sense* similar if for any $a > 0$ there exist $b > 0$ and a function $c(t)$ from $[0, \infty)$ to $\mathbb{R}^d$ such that

$$\mathbf{X}_{at} \stackrel{d}{=} b\mathbf{X}_t + c(t). \tag{4.8}$$

The following reasoning shows that this self-similar property is shared by all strictly stable stochastic processes.

From definition 3.3.2 on page 27 and equation 3.15 on page 28 it follows that if $\mathbf{X}_t$ is a strictly α-stable process, then for any constant $a > 0$,

$$\left[\widehat{\mathbb{P}}_{\mathbf{X}_t}(\mathbf{u})\right]^a = \widehat{\mathbb{P}}_{\mathbf{X}_t}(a^{1/\alpha}\mathbf{u}). \tag{4.9}$$

From equation 4.9 , equation 4.1 on page 35 and equation 2.4 on page 19 it follows that for any $k, t > 0$

$$\begin{aligned}
&\widehat{\mathbb{P}}_{k^{-1/\alpha}\mathbf{X}_{kt}}(\mathbf{u}) \\
&= \widehat{\mathbb{P}}_{\mathbf{X}_{kt}}(k^{-1/\alpha}\mathbf{u}) \\
&= \left[\widehat{\mathbb{P}}_{\mathbf{X}_t}(k^{-1/\alpha}\mathbf{u})\right]^k \\
&= \widehat{\mathbb{P}}_{\mathbf{X}_t}(k^{-1/\alpha}\mathbf{u}k^{1/\alpha}) \\
&= \widehat{\mathbb{P}}_{\mathbf{X}_t}(\mathbf{u}).
\end{aligned}$$

From the uniqueness item ii on page 19 it follows that $\mathbf{X}_{kt}$ has the same law as $\mathbf{X}_t$, i.e. $\mathbf{X}_t$ is a self-similar stochastic process.

A similar calculation with a stable process using equation 2.4 on page 19 shows that stable processes possess the broad-sense similar property. The converse is also true, as stated in the following proposition.

Proposition 4.2.2

A Lévy-process on $\mathbb{R}^d$ is self-similar or broad-sense similar if it is respectively strictly stable or stable.

Proof: A proof is given on page 71 in Sato (1999).

The following theorem further elucidates the relation between stable and strictly stable processes:

Theorem 4.2.3

(i) If $\{\mathbf{X}_t : t \geq 0\}$ is α-stable with $0 < \alpha < 1$ or $1 < \alpha \leq 2$, then for some $\mathbf{k} \in \mathbb{R}^d$, $\{\mathbf{X}_t - t\mathbf{k}\}$ is strictly α-stable.

(ii) If $\{\mathbf{X}_t : t \geq 0\}$ is 1-stable, then for any choice of a function $\mathbf{k}(t)$, $\{\mathbf{X}_t - k(t)\}$ is not strictly 1-stable.

Proof: A proof can be found on page 82 in Sato (1999).

Due to theorem theorem 4.2.3 we shall henceforth refer to α-stable processes with $\alpha \neq 1$ which are not strictly stable processes, as α-stable processes *with drift*. In particular we shall call a strictly 2-stable process a Wiener process and a 2-stable process that is not strictly stable a Wiener process with drift.

4.2.2 Path properties of Lévy processes

A function $f : [0, \infty) \to \mathbb{R}^d$ is said to be *piecewise constant* if there exist $0 = t_0 < t_1 < \cdots < t_n = \infty$ or $0 = t_0 < t_1 < \cdots < \lim_{j \to \infty} t_j = \infty$, such that $f(t)$ is constant on each interval $[t_{j-1}, t_j)$.

We say that a Lévy process is a zero process if it has the value $\mathbf{0}$ with probability one for every $t \geq 0$.

For Lévy processes with piecewise constant trajectories we have the following result:

Theorem 4.2.4

A Lévy process has piecewise constant trajectories a.s. if and only if it is a compound Poisson process or a zero process. A proof can be found on page 136 in Sato (1999).

Given a function $f : \mathbb{R} \to \mathbb{R}^d$ we define the *total variation function* $\mathrm{T}_f([a, b])$ of f on $[a, b]$ as

$$T_f([a,b]) := \sup\left\{\sum_1^n |f(t_j) - f(t_{j-1})| : n \in 1, 2, \ldots, a = t_0 < \cdots < t_n = b\right\}.$$

We say that f is of *finite variation* on $[a,b]$ if $T_f([a,b]) < \infty$.

We say that a Lévy process is of finite variation if, as functions of t, its trajectory is a.s. of finite variation on every compact interval.

For Lévy processes of finite variation the following can be showed to be true:

Proposition 4.2.5

A Lévy process is of finite variation if and only if its Gaussian coefficient $Q = 0$ and $\int_{\mathbb{R}^d} (\min(1, |\mathbf{x}|))\nu(d\mathbf{x}) < \infty$.

As a corollary of proposition 4.2.5 we have the so-called Lévy-Itô decomposition. In the case of finite variation the Lévy-Khintchine formula (equation 4.3 on page 37) can be simplified as stated below.

Corollary 4.2.6

Let $\mathbf{X}_t$ be a Lévy process of finite variation with its generating triplet given by $(0, \nu, \boldsymbol{\gamma})$ and let $J_{\mathbf{X}}$ have the same definition as in definition 2.3.6 on page 18. Then $\mathbf{X}_t$ can be expressed as a linear drift term plus the sum of its jumps between 0 and t:

$$\mathbf{X}_t = \boldsymbol{\gamma_0} t + \int_{[0,t]\times\mathbb{R}^d} \mathbf{x} J_{\mathbf{X}}(ds \times d\mathbf{x}) = \boldsymbol{\gamma_0} t + \sum_{s\in[0,t]} \Delta \mathbf{X}_s,$$

where $\boldsymbol{\gamma_0}$ is the *drift* defined in equation 3.7 on page 25 as

$$\gamma_0 := \boldsymbol{\gamma} - \int_{\mathbb{R}^d} \mathbf{x} 1_{|\mathbf{x}|\leq 1} \nu(\mathrm{d}\mathbf{x}). \tag{4.10}$$

The characteristic function of $\mathbf{X}_t$ can be expressed as:

$$\mathbb{E}\left[e^{i\langle \mathbf{u}, \mathbf{X}_t\rangle}\right] = \exp t \left\{ i\langle \boldsymbol{\gamma_0}, \mathbf{u}\rangle + \int_{\mathbb{R}^d} (e^{i\langle \mathbf{u}, \mathbf{x}\rangle} - 1)\nu(d\mathbf{x}) \right\}. \tag{4.11}$$

A proof of proposition 4.2.5 and corrolary 4.2.6 can be found on the pages 86-87 in Cont and Tankov (2004).

A Lévy process $\{\mathbf{X}_t : t \geq 0\}$ is called a compound Poisson process if $\mathbf{X}_1$ (the distribution of $\mathbf{X}_t$ at $t = 1$) is compound Poisson distributed.

It can be shown (see page 16 in Bertoin (1996)) that a Lévy process with finite variation is a compound Poisson process if and only if its drift coefficient $\boldsymbol{\gamma_0}$ is zero and its Lévy measure ν is a finite measure.

4.3 Subordination

For a one-dimensional Lévy process X_t we have the following equivalences :

Proposition 4.3.1

(i) $X_t \geq 0$ a.s. for some $t > 0$.

(ii) $X_t \geq 0$ a.s. for every $t > 0$.

(iii) X_t is nondecreasing a.s. as a function of t.

(iv) The generating triplet of X_t satisfies $Q = 0, \nu((-\infty, 0]) = 0$, $\int_0^\infty \min(x, 1)\nu(dx) < \infty$ and $\gamma_0 \geq 0$, where γ_0 is the *drift* defined in equation 4.10 on the previous page.

(A proof can be found on page 88 in Cont and Tankov (2004).)

A Lévy process that satisfies one of the equivalent conditions above is called a *subordinator*.

Since a subordinator S_t is a non-negative random variable for all t it is convenient to describe S_t using the Laplace transform rather than the Fourier transform. If S_t has the generating triplet (Q, ν, γ_0) then its moment generating function $\mathbb{E}\left[e^{uS_t}\right]$ for u is of the form

$$\mathbb{E}\left[e^{uS_t}\right] = e^{tl(u)} \text{ for all } u \leq 0, \text{ where } l(u) = \gamma_0 u + \int_0^\infty (e^{ux} - 1)\nu(\mathrm{d}x). \quad (4.12)$$

We shall call $l(u)$ the *Laplace exponent* of S.

A subordinator can be used to build a new Lévy process by "time-changing" another. Subordination or "time-changing" a Lévy process with a subordinator is a powerful technique which is very popular in financial modelling (see section 4.4 in Cont and Tankov (2004)). As shown in the theorem below the subordinated stochastic process Y_{X_t} is again a Lévy process.

Theorem 4.3.2

Subordination of a Lévy processes:
Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $(\mathbf{X}_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$ with characteristic exponent ϕ (see equation 4.3 on page 37) and triplet $(Q, \nu, \boldsymbol{\gamma})$. Let $(S_t)_{t\geq 0}$ be a subordinator with Laplace exponent $l(u)$ and triplet $(0, \rho, \gamma_0)$. Then the process $(\mathbf{Y}_t)_{t\geq 0}$ defined for each $\omega \in \Omega$ by $\mathbf{Y}(t, \omega) = \mathbf{X}(S(t, \omega), \omega)$ is

a Lévy process. Its characteristic function is

$$\mathbb{E}\left[e^{i\langle \mathbf{u},\mathbf{Y}_t\rangle}\right] = e^{tl(\phi(\mathbf{u}))}. \tag{4.13}$$

In the above the characteristic exponent of $\mathbf{Y}_t$ is obtained by composition of the Laplace exponent of S_t with the characteristic exponent of $\mathbf{X}_t$. The generating triplet $(Q^{\mathbf{Y}}, \nu^{\mathbf{Y}}, \boldsymbol{\gamma}^{\mathbf{Y}})$ is then given by the following:

$$\begin{aligned}
Q^{\mathbf{Y}} &= \gamma_0 Q, \\
\nu^{\mathbf{Y}}(B) &= \gamma_0 \nu(B) + \int_0^\infty p_s^{\mathbf{X}}(B)\rho(ds), \quad \text{for all } \mathscr{B}(\mathbb{R}^d), \\
\boldsymbol{\gamma}^{\mathbf{Y}} &= \gamma_0 \boldsymbol{\gamma} + \int_0^\infty \rho(ds) \int_{|\mathbf{x}|\le 1} \mathbf{x} p_s^{\mathbf{X}}(d\mathbf{x}),
\end{aligned}$$

where $p_t^{\mathbf{X}}$ is the probability distribution of $\mathbf{X}_t$.

A proof can be found on the pages 108-109 in Cont and Tankov (2004).

5

Modeling the dependence structure of multivariate Lévy processes

As described in the introduction of Nelsen (1998), from one point of view, “copulas are functions that join or couple multivariate distributions to their one-dimensional marginal distribution functions. Alternatively, copulas are multivariate distribution functions whose one-dimensional margins are uniform on the interval $(0, 1)$.” They are ”of interest to statisticians for two main reasons. Firstly, as a way of studying scale-free measures of dependence; and secondly, as a starting point for constructing families of bivariate distribution, sometimes with a view to simulation,” (pages 1-2 in Nelsen (1998)). Moreover, copulas have grown in popularity in recent years since copulas can handle various non-linear as well as linear forms of dependency between random variables.

In section 5.1 on the next page we first give a description of the notion of copulas. In section 5.2 on page 49 we explain how Lévy copulas extend the notion of copulas to Lévy processes by linking different Lévy measures together. This is analogous to the way that copulas link probability measures.

In the rest of the chapter we develop the theory of Lévy copulas, first for Lévy processes with positive jumps (section 5.3 on page 51) and finally for general Lévy processes (section 5.4 on page 55). We will apply this theory to implement a simulation algorithm of a stable process in chapter 6 and to estimate a ruin probability in chapter 7. Below [1] is a graphical depiction of some of the concepts involved.

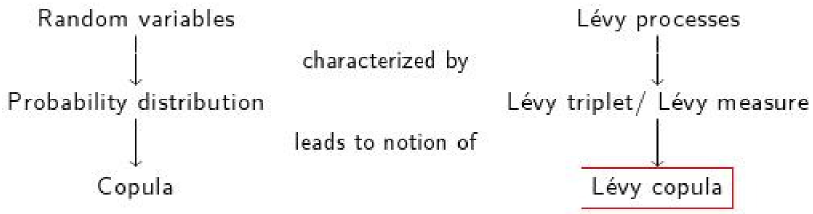

Figure 5.1: Graphical illustration of relation between copulas and Lévy copulas.

5.1 Copulas

5.1.1 About the notation

In this chapter we will follow the notation suggested by Cont and Tankov in Cont and Tankov (2004). We therefore start by giving definitions of d-boxes, quasi-inverse, pseudo-inverse, increasing function, F-volume, d-increasing functions, and margins.

We write the set of non-negative real numbers as $\mathbb{R}_+$.

We let the symbol $\overline{\mathbb{R}} := \mathbb{R} \bigcup \{-\infty, \infty\}$ denote the extended set of real numbers. We let $[\mathbf{a}, \mathbf{b}]$ denote a closed d-box of $\overline{\mathbb{R}}^d$:

$$[\mathbf{a}, \mathbf{b}] := [a_1, b_1] \times \ldots \times [a_d, b_d]$$

such that $a_k < b_k$ for all $k = 1, 2, \ldots, d$.

The vertices of a d-box are the points $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_j, \ldots, \mathbf{c}_{2^d}$ where each c_j has d vector components. Each vector component is equal to either a_k or b_k for some $k \in 1, \ldots, d$.

[1] Figure 5.1 is taken from Packham (2006).

Definition 5.1.1

A one-dimensional *abstract distribution function* is a function F with domain $\overline{\mathbb{R}}$ such that

(1) F is non-decreasing and

(2) $F(-\infty) = 0$ and $F(\infty) = 1$.

Definition 5.1.2

Let F be a a one-dimensional abstract distribution function. Then a *quasi-inverse* of F is any function $F^{(-1)}$ with domain $[0, 1]$ such that

(1) If $u \in \operatorname{Ran} F$, then $F^{(-1)}(u)$ is any number x in $\overline{\mathbb{R}}$ such that $F(x) = u$.

(2) If u is not in $\operatorname{Ran} F$, then

$$F^{(-1)}(u) = \inf\{x : F(x) \geq u\} = \sup\{x : F(x) \leq u\}.$$

We define the pseudo-inverse as follows:

Definition 5.1.3

Let $\phi : [0, 1] \to [0, \infty]$ be a continuous, strictly decreasing function such that $\phi(1) = 0$. The *pseudo-inverse* of ϕ is the function $\phi^{[-1]}$ with $\operatorname{Dom} \phi^{[-1]} = [0, \infty]$ and $\operatorname{Ran} \phi^{[-1]} = [0, 1]$ given by

$$\phi^{[-1]}(u) = \begin{cases} \phi^{-1}(u), & 0 \leq u \leq \phi(0), \\ 0, & \phi(0) < u \leq \infty. \end{cases} \tag{5.1}$$

Note that $\phi^{[-1]}$ is continuous and nonincreasing on $[0, \infty]$, and strictly decreasing on $[0, \phi(0)]$. If $\phi(0) = \infty$, then $\phi^{[-1]} = \phi^{-1}$.

Definition 5.1.4: *F*-volume

Let $S_1 \ldots S_n$ be nonempty subsets of of $\overline{\mathbb{R}}$. Let F be a real-valued function of n variables such that $\operatorname{Dom} F = S_1 \times \ldots \times S_n$ for every n-box $B = [\mathbf{a}, \mathbf{b}]$ whose vertices are in Dom F. Then the *F-volume* is defined by

$$V_F(B) := \sum_{j=1^{2^n}} \text{sign}(\mathbf{c}_j)F(\mathbf{c}_j). \tag{5.2}$$

Here the sum is taken over all vertices $\mathbf{c}_1, \dots \mathbf{c}_{2^n}$ of B, and sign($\mathbf{c}_k$) is as follows:

$$\text{sign}(\mathbf{c}_j) := \begin{cases} 1 & \text{if } \mathbf{c}_j^{(k)} = a_k \text{ for an even number of k's} \\ -1 & \text{otherwise.} \end{cases}$$

In the 2-dimensional case the F-volume of a rectangle $B = [x_1, x_2] \times [y_1, y_2] \subset S_1 \times S_2$, reduces to

$$V_F(B) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1). \tag{5.3}$$

Definition 5.1.5: *n*-increasing function, grounded function and margins

- A real-valued function F of n variables is called *n-increasing* if $V_F(B) \geq 0$ for all n-boxes whose vertices lie in Dom F.

- Suppose that the domain of F is $S_1 \times \cdots \times S_n$, where each S_k has a smallest element a_k. Consider the set of vertices c for the n-box defined by Dom F. Let W be the subset of these vertices where, for at least one value of k, a vector component c_k takes its value from the corresponding least value a_k of the interval S_k. If, for all the vertices in W, $F(\mathbf{c}) = 0$, then the function F is said to be *grounded*.

- If each S_k is nonempty and has a greatest element b_k, then (one-dimensional) *abstract margins* of F are functions F_k with Dom $F_k = S_k$, defined by $F_k(x) = F(b_1, \dots, b_{k-1}, x, b_{k+1}, \dots, b_n)$ for all x in S_k.

Definition 5.1.6

An *n-dimensional copula* is a function $C : [\mathbf{0}, \mathbf{1}] \to [0,1]$ such that

(1) C is grounded and n-increasing, and

(2) for all $k = 1, 2, \ldots, n$, C has abstract margins C_k, which satisfy $C_k(u) = u$ for all $u \in [0,1]$.

We now introduce an n-dimensional *abstract distribution function*.

Definition 5.1.7

An n-dimensional abstract distribution function is a function $F : \overline{\mathbb{R}}^d \to [0,1]$ which is grounded, n-increasing and satisfies $F(\infty, \infty, \ldots, \infty) = 1$.

Theorem 5.1.8: Sklar's theorem

Let F be a n-dimensional abstract distribution function with margins $F_1, \ldots, F_n$. Then there exists an n-dimensional copula C such that for all $\mathbf{x} \in \mathbb{R}^n$,

$$F(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)). \tag{5.4}$$

If $F_1, \ldots, F_n$ are all continuous then C is unique. Otherwise C is uniquely determined on $\operatorname{Ran} F_1 \times \cdots \times \operatorname{Ran} F_n$. Conversely, if C is an n-copula and $F_1, \ldots, F_n$ are one-dimensional abstract distribution functions, then the function F defined by equation 5.4 is a n-dimensional abstract distribution function with abstract margins $F_1, \ldots, F_n$.

Proof: A proof of the two-dimensional case is found on page 18 in Nelsen (1998). A reference to a proof of the general n-dimensional case is given on page 41 in Nelsen (1998).

The class of *Archimedean 2-copulas* is constructed by means of the following theorem:

Theorem 5.1.9

Let $\phi : [0,1] \to [0, \infty]$ be a continuous, strictly decreasing function such that $\phi(1) = 0$, and let $\phi^{[-1]}$ be the pseudo-inverse of ϕ defined by equation 5.1 on page 46. Let the function $C : [0,1] \times [0,1] \to [0,1]$ be given by

$$C(u,v) = \phi^{[-1]}(\phi(u) + \phi(v)). \tag{5.5}$$

Then C is a 2-copula if and only if ϕ is convex.

Proof: A proof can be found on page 91 in Nelsen (1998).

If we take $\phi(r) = \left(r^{-\theta} - 1\right)^{-\frac{1}{\theta}} \quad \theta$ we get the Clayton family of copulas,

$$C_\theta(x,y) = \left(x^{-\theta} + y^{-\theta} - 1\right)^{-\frac{1}{\theta}} \quad \theta > 0. \tag{5.6}$$

5.2 Using the Lévy measure to model dependence structure

Since the law of a Lévy process $\{\mathbf{X}_t : t \geq 0\}$ is completely determined by the law of $\mathbf{X}_s$ for a fixed time $t = r, \quad r > 0$, the dependence structure of a two-dimensional Lévy process (X_t, Y_t) can be parameterized by the copula C_r of X_r and Y_r for some $r > 0$. In Cont and Tankov (2004) the following drawbacks in this approach are noted:

- **Copulas may be time-dependent.**

  "The copula C_r may depend on r (an example is given in Tankov (2003a)). C_s for some $s \neq r$ cannot in general be computed from C_r because C_s also depends on the marginal distributions at time r and at time s."

- **Copulas are invariant for strictly increasing transformations.**

  On page 22 in Nelsen (1998) the following theorem is stated and proved:

  Theorem 5.2.1

  Let X and Y be continuous random variables with copula C_{XY}. If $\alpha(X)$ and $\beta(Y)$ are strictly increasing on Ran X and Ran Y respectively, then $C_{\alpha(X)\beta(Y)} = C_{XY}$. Thus C_{XY} is invariant under strictly increasing transformations of X and Y.

  As noted on page 143 in Cont and Tankov (2004), the property of infinite divisibility of a random variable is destroyed under strictly increasing transformations . We therefore have that "For given infinitely divisible marginal laws $\mathbb{P}^{X_t}$ and $\mathbb{P}^{Y_t}$, it is not clear which copulas C_t will yield a two-dimensional infinitely divisible law." (page 143 in Cont and Tankov (2004)).

Instead of using copulas, Cont and Tankov want to model dependence between two Lévy processes X_t and Y_t in a way that, in their words, "preserves the Lévy property and reflects the dynamic structure of Lévy processes".

To clarify this point they provide the example below:

Example 5.2.2: Dynamic complete dependence for Lévy processes

Let X_t be a pure jump Lévy process. Let Y_t be a Lévy process, constructed from the jumps of X_t: $Y_t = \sum_{s\leq t} \Delta X_s^3$. From the dynamic point of view X_t and Y_t are completely dependent in the following way: The trajectory of one of them can be reconstructered from the trajectory of the other. However, the copula of X_t and Y_t is not that of complete dependence, because Y_t is not a deterministic function of X_t.

Cont and Tankov use the above example to argue that the important dependence concept for Lévy processes is "the dependence of jumps that should be studied using the Lévy measure", since "knowledge of the jump dependence (...) allows (one) to characterize the dynamic structure of a Lévy process (...) which is very important for risk management and other financial applications".

We cite two results from Cont and Tankov (2004) that, in their words, "show how independence of Lévy processes can be expressed in terms of the Lévy measure."

First, as a consequence of proposition 4.1.10 on page 38, we have that the abstract margins of a Lévy measure can be computed in the same way as the margins of a probability measure on $\mathbb{R}^d$, as in the following proposition:

Proposition 5.2.3: Abstract margins of Lévy measure

Let $\mathbf{X}_t = (X_t, Y_t)$ be a two-dimensional Lévy process with generating triplet $(Q, \nu, \boldsymbol{\gamma})$. Then the component X_t of $\mathbf{X}_t$ has generating triplet (Q_X, ν_X, γ_X) where

$$
\begin{aligned}
&Q_X = Q_{11} \\
&\nu(B) = \nu(B \times (-\infty, \infty)), \text{for all } B \in \mathscr{B}(\mathbb{R}) \\
&\gamma_X = \gamma_1 + \int_{\mathbb{R}^2} x\left(1_{x^2\leq 1} - 1_{x^2+y^2\leq 1}\right)\nu(dx \times dy).
\end{aligned}
$$

As a second result, to express that components of Lévy processes are independent, we have the following:

Proposition 5.2.4: Independence of Lévy processes

Let (X_t, Y_t) be a Lévy process with a Lévy measure and without a Gaussian coefficient (the positive nonnegative-definite matrix Q in theorem 4.1.9 on page 37). The components of the Lévy process are independent if and only if the support of its measure ν is contained in the set $\{(x, y) : xy = 0\}$. That is, if and only if the components never jump together. With this restriction

$$\nu(A) = \nu_X(A_X) + \nu_Y(A_Y),$$

where $A_X = \{x : (x, 0) \in A\}$, $A_Y = \{y : (0, y) \in A\}$, and ν_X and ν_Y are Lévy measures of X_t and Y_t.

Proof: A proof can be found on page 144 in Cont and Tankov (2004).

5.3 Lévy copulas for Lévy processes with positive jumps

Definition 5.3.1

A d-dimensional abstract *tail integral* is a function $U : [0, \infty]^d \to [0, \infty]$ such that

(1) $(-1)^d U$ is a d-increasing function.

(2) U is equal to zero if one of its arguments is equal to ∞,
U is finite everywhere except at zero,
and $U(0, \ldots, 0) = \infty$.

The *margins* of a Lévy measure are defined similarly to the margins of a distribution function:

$$U\left(0, \ldots, x_k, 0 \ldots, 0\right) = U_k(x_k).$$

For every Lévy measure ν on $(0,\infty] \times (0,\infty]$ one can define its tail integral as follows:

$$U(x_1,x_2) = 0 \qquad \text{if } x_1 = \infty \text{ or } x_2 = \infty. \tag{5.7}$$

$$U(x_1,x_2) = \nu\left(\left[x_1,\infty\right) \times \left[x_2,\infty\right)\right) \quad \text{for } \left(x_1,x_2\right) \in [0,\infty) \times [0,\infty) \setminus \{(0,0)\}. \tag{5.8}$$

$$U(0,0) = \infty. \tag{5.9}$$

Now let us consider going the other way. That means using an abstract tail integral, as defined in definition 5.3.1 on the preceding page, to define a Lévy measure. We know from theorem 4.1.9 on page 37, that any Lévy measure must satisfy the following integrability requirement:

$$\int_{[0,1]\times[0,1]} |\mathbf{x}|^2\, \nu\,(\mathrm{d}\mathbf{x}) = \int_{[0,1]\times[0,1]} |\mathbf{x}|^2\, \nu\,(\mathrm{d}U) < \infty. \tag{5.10}$$

The exact requirements for a two-dimensional tail integral to define a Lévy measure are specified in the lemma below, taken from chapter 5 in Cont and Tankov (2004).

Lemma 5.3.2

Let U be a two-dimensional tail integral with margins U_1 and U_2. U defines a Lévy measure on $[0,\infty) \times [0,\infty) \setminus (0,0)$ (i.e. the integrability condition equation 5.10 is satisfied) if and only if the following condition is met: The margins of U correspond to Lévy measures on $[0,\infty)$. That is, for $k = 1,2$,

$$\int_0^1 x^2\, \mathrm{d}U_k(x) < \infty.$$

Proof: A proof is on page 147 in Cont and Tankov (2004).

Definition 5.3.3

An n-dimensional *positive Lévy copula* is an n-increasing grounded function $F : [0,\infty]^n \to [0,\infty]$, with margins F_k for $k = 1,\ldots.n$, which satisfy $F_k(u) = u$ for all $u \in [0,\infty]$.

The following is the general theorem equivalent to Sklar's theorem for copulas (theorem 5.1.8 on page 48):

Theorem 5.3.4

Let U be the tail integral of an n-dimensional Lévy process with positive jumps and let $U_1, \ldots, U_n$ be the tail integrals of its components. Then there exists an n-dimensional positive Lévy copula F such that, for all vectors $(x_1, \ldots, x_n)$ in $\mathbb{R}^n_+$,

$$U(x_1, x_2, \ldots, x_n) = F(U_1(x_1), \ldots, U_n(x_n)).$$

If the $U_1, \ldots U_n$ are continuous then F is unique, otherwise it is unique on $\operatorname{Ran} U_1 \times \cdots \times \operatorname{Ran} U_n$.
Conversely, if F is an n-dimensional positive Lévy copula and $U_1, \ldots, U_n$ are tail integrals on $(0, \infty)$, then the function U defined above is the tail integral of an n-dimensional Lévy process with positive jumps having marginal tail integrals $U_1, \ldots, U_n$.

Proof: A proof of the two-dimensional case is found on page 148 in Cont and Tankov (2004). A general guideline on how to prove the n-dimensional case is given on page 155 in Cont and Tankov (2004).

Proposition 5.3.5

Let C be a 2-copula (not a Lévy copula). Let $f(x)$ be an increasing convex function from $[0, 1]$ to $[0, \infty]$. Then

$$F(x, y) = f(C(f^{-1}(x), f^{-1}(y)))$$

defines a two-dimensional positive Lévy copula.

Proof: A proof of the above result is given on page 153 in Cont and Tankov (2004).

Proposition 5.3.6

Let ϕ be a strictly decreasing function from $[0, \infty]$ to $[0, \infty]$ such that $\phi(0) = \infty$ and $\phi(\infty) = 0$. Let the quasi-inverse $\phi^{(-1)}$ have derivatives up to the order n on $(0, \infty)$ with alternating signs. That is $(-1)^k \frac{d^k \phi^{(-1)}(r)}{\mathrm{d}r^k} > 0$. Then

$$F(x_1, \ldots, x_n) = \phi^{(-1)}(\phi(x_1) + \ldots \phi(x_n))$$

defines an n-dimensional positive Lévy copula.

Proof: A proof of the two-dimensional case is given on page 153 in Cont and Tankov (2004).

A general guideline on how to prove the n-dimensional case is given on page 155 in Cont and Tankov (2004).

For example, with $\phi(u) = u^{-\theta}$ for $\theta > 0$ we get a family of positive Lévy copulas of the form

$$F_\theta(u, v) = (u^{-\theta} + v^{-\theta})^{-1/\theta}. \tag{5.11}$$

We shall refer to this as the family of *positive Clayton Lévy copulas*, since it resembles the Clayton copulas defined in equation 5.6 on page 49. The limiting case $\theta \to \infty$ corresponds to complete dependence and $\theta \to 0$ corresponds to independence.

Positive Lévy copulas also have a probabilistic interpretation, as stated in the following results from Cont and Tankov (2004):

Lemma 5.3.7

Let F be a two-dimensional positive Lévy copula. Then for almost all $x \in [0, \infty]$, the function

$$F_x(y) = \frac{\partial}{\partial x} F(x, y)$$

exists and is continuous for all $y \in [0, \infty]$. Moreover, it is a distribution function of a positive random variable, that is, it is increasing and satisfies $F_x(0) = 0$ and $F_x(\infty) = 1$.

Proof: A proof of the above result is given on page 154 in Cont and Tankov (2004).

Theorem 5.3.8

Let (X_t, Y_t) be a two-dimensional Lévy process with positive jumps, having marginal tail integrals U_1, U_2 and Lévy copula F. Let ΔX_t and ΔY_t be the jump sizes of the two components at time t. Then, if U_1 has a non-zero density at x, $F_{U_1(x)}$ is the distribution function of $U_2(\Delta Y_t)$ conditionally on $\Delta X_t = x$:

$$F_{U_1(x)}(y) = P\left\{U_2(\Delta Y_t) \le y | \Delta X_t = x\right\}.$$

Proof: A reference to a proof is given on page 155 in Cont and Tankov (2004).

5.4 Lévy 2-copulas for general Lévy processes

Definition 5.4.1

A function $F : \overline{\mathbb{R}}^2 \to \overline{\mathbb{R}}$ is called a general *Lévy 2-copula* if

- F is 2-increasing,
- $F(0,x) = F(x,0) = 0$ for all x ,
- $F(x,\infty) - F(x,-\infty) = F(\infty,x) - F(-\infty,x) = x$.

An example of a general Lévy 2-copula is given below

Example 5.4.2: General Clayton Lévy 2-copula

$$F_\theta(u,v) = \begin{cases} (|u|^{-\theta} + |v|^{-\theta})^{-1/\theta} 1_{xy \geq 0} & \text{for } \theta > 0. \\ -\left(|u|^{-\theta} + |v|^{-\theta}\right)^{-1/\theta} 1_{xy \leq 0} & \text{for } \theta < 0. \end{cases}$$

Here $\theta \to -\infty$ corresponds to complete negative dependence, $\theta \to 0$ corresponds to independence, and $\theta \to \infty$ corresponds to complete positive dependence.

A tail integral of a general Lévy measure on $\mathbb{R}$ is defined as follows:

Definition 5.4.3

Let ν be a Lévy measure on $\mathbb{R}$. The *tail integral* of ν is a function $U : \overline{\mathbb{R}} \setminus \{0\} \to [-\infty,\infty]$ defined by

$$\begin{aligned} &U(x) = \nu([x,\infty]) \quad \text{for } x \in (0,\infty). \\ &U(x) = -\nu((-\infty,-x]) \quad \text{for } x \in (-\infty,0). \\ &U(\infty) = U(-\infty) = 0. \end{aligned}$$

Having introduced Lévy 2-copulas for general Lévy processes, Cont and Tankov (on page 157 in Cont and Tankov (2004)) go on to show that sufficiently smooth general Lévy copulas can be used to construct two-dimensional Lévy densities from one-dimensional ones.

Proposition 5.4.4

Let F be a two-dimensional Lévy copula, continuous on $[-\infty,\infty]^2$, such that $\dfrac{\partial^2 F(u,v)}{\partial u \partial v}$ exists on $(-\infty,\infty)^2$ and let U_1 and U_2 be one-dimensional tail integrals with densities ν_1 and ν_2. Then

$$\left.\frac{\partial^2 F(u,v)}{\partial u \partial v}\right|_{u=U_1(x),v=U_2(y)} \nu_1(x)\nu_2(y)$$

is the Lévy density of a Lévy measure, with marginal Lévy densities ν_1 and ν_2.

Proof: This result is stated on page 157 in Cont and Tankov (2004).

To construct Lévy copulas with both positive and negative jumps Cont and Tankov suggest treating each corner of the Lévy measure separately, as in the definition cited below from Cont and Tankov (2004).

Definition 5.4.5

A method for constructing general tail integrals:

Consider the 1-dimensional case. Let ν be a Lévy measure on $\mathbb{R}$. This measure has two tail integrals, $U^+ : [0,\infty] \to [0,\infty]$ for the positive part and $U^- : [-\infty,0] \to [-\infty,0]$ for the negative part, defined as follows:

$$\begin{aligned}
&U^+(x) = \nu([x,\infty)) \quad \text{for } x \in (0,\infty), \quad U^+(0) = \infty, \quad U^+(\infty) = 0;\\
&U^-(x) = \nu((-\infty,x]) \text{ for } x \in (-\infty,0),\ U^-(0) = -\infty,\ U^+(-\infty) = 0.
\end{aligned}$$

Now consider the 2-dimensional case. Let ν be a Lévy measure on $\mathbb{R}^2$ with marginal tail integrals U_1^+, U_1^-, U_2^+ and U_2^-. This measure has four tail integrals: U^{++}, U^{+-}, U^{-+} and U^{--}, where each tail integral is defined on its respective quadrant, including the coordinate axis, as follows:

$$\begin{aligned}
U^{++}(x,y) &= \nu([x,\infty) \times [y,\infty)), && \text{if } x \in (0,\infty) \text{ and } y \in (0,\infty)\\
U^{+-}(x,y) &= -\nu([x,\infty) \times (-\infty,y]), && \text{if } x \in (0,\infty) \text{ and } y \in (-\infty,0)\\
U^{-+}(x,y) &= -\nu((-\infty,x]) \times [y,\infty)), && \text{if } x \in (-\infty,0) \text{ and } y \in (0,\infty)\\
U^{--}(x,y) &= \nu((-\infty,x]) \times (-\infty,y]), && \text{if } x \in (-\infty,0) \text{ and } y \in (-\infty,0).
\end{aligned}$$

If x or y is equal to $+\infty$ or $-\infty$, the corresponding tail integral is zero. If x or y is equal to zero, the tail integrals satisfy the following "margin" conditions:

$$
\begin{aligned}
U^{++}(x,0) - U^{+-}(x,0) &= U_1^{+}(x)\\
U^{-+}(x,0) - U^{--}(x,0) &= U_1^{-}(x)\\
U^{++}(0,y) - U^{-+}(0,y) &= U_2^{+}(y)\\
U^{+-}(0,y) - U^{--}(0,y) &= U_2^{-}(y)
\end{aligned}
$$

With the two-dimensional Lévy measure taken separately for each quadrant, a theorem analogous to theorem 5.1.8 on page 48 can be stated for general Lévy measures, as follows:

Theorem 5.4.6

Let ν be a Lévy measure on $\mathbb{R}^2$ with marginal tail integrals U_1^+, U_1^-, U_2^+, and U_2^-. Then there exists a Lévy copula F such that U^{++}, U^{+-}, U^{-+} and U^{--} are tail integrals of ν, as follows:

$$
\begin{aligned}
U^{++}(x,y) &= F(U_1^{+}(x), U_2^{+}(y)) && \text{if } x \geq 0 \text{ and } y \geq 0.\\
U^{+-}(x,y) &= F(U_1^{+}(x), U_2^{-}(y)) && \text{if } x \geq 0 \text{ and } y \leq 0.\\
U^{-+}(x,y) &= F(U_1^{-}(x), U_2^{+}(y)) && \text{if } x \leq 0 \text{ and } y \geq 0.\\
U^{--}(x,y) &= F(U_1^{-}(x), U_2^{-}(y)) && \text{if } x \leq 0 \text{ and } y \leq 0.
\end{aligned}
$$

If the marginal tail integrals are absolutely continuous [2] and ν does not change the coordinate axes, the Lévy copula is unique. Conversely, if F is a Lévy copula and $U_1^+, U_1^-, U_2^+, U_2^-$ are tail integrals of one-dimensional Lévy measures, then the above formulas define a set of tail integrals of a two-dimensional Lévy measure.

Proof: A proof is given on page 158 in Cont and Tankov (2004).

As a technique to construct general Lévy copulas, Cont and Tankov suggest getting them "from positive ones by gluing them together", which

[2] A function $F : \mathbb{R} \to \mathbb{R}$ is called *absolutely continuous* if for every $\epsilon > 0$ there exists $\delta > 0$ such that for any finite set of disjoint intervals $(a_1, b_1), (a_2, b_2) \dots (a_N, b_N)$, $\sum_{j=1}^{N}(b_j - a_j) < \delta$ implies that $\sum_{j=1}^{N} |F(b_j) - F(a_j)| < \epsilon$.

"amounts to specifying the dependence of different signs separately"
Cont and Tankov (2004).

By this procedure, letting $F^{++}, F^{--}, F^{-+}, F^{+-}$ be positive Lévy copulas, it can be shown that (see page 160 in Cont and Tankov (2004))

$$F(x,y) = F^{++}(c_1|x|, c_2|y|)1_{x\geq 0, y\geq 0} + F^{--}(c_3|x|, c_4|y|)1_{x\leq 0, y\leq 0} \\ - F^{+-}((1-c_1)|x|, (1-c_4)|y|)1_{x\geq 0, y\leq 0} - F^{-+}((1-c_3)|x|, (1-c_2)|y|)1_{x\leq 0, y\geq 0} \tag{5.12}$$

defines a Lévy copula if $c_1, \dots, c_4$ are constants between 0 and 1.

In Cont and Tankov (2004) Lévy copulas constructed as defined by equation 5.12 are referred to as *constant proportion Lévy copulas*.

For completeness we also mention the definition of a general n-dimensional Lévy copula, as follows:

Definition 5.4.7

An *n-dimensional Lévy copula* is a function $F : \overline{\mathbb{R}}^n \to \overline{\mathbb{R}}$ with the following three properties:

- F is n-increasing,
- F is equal to zero if at least one of its arguments is zero and,
- $F(x, \infty, \dots, \infty) - F(x, -\infty, \dots, -\infty) = x$,
  $F(\infty, x, \infty, \dots, \infty) - F(-\infty, x, -\infty, \dots, -\infty) = x$, etc.

Using a special type of interval,

$$\mathscr{I}(x) = \begin{cases} [x, \infty), & \text{if } x > 0 \\ (-\infty, x], & \text{if } x < 0, \end{cases}$$

tail integrals $U_1, \dots, U_n$ of the Lévy measure can be computed everywhere except on the axes, as follows:

$$\nu(\mathscr{I}(x_1) \times, \dots, \times \mathscr{I}(x_n)) \\ = (-1)^{\text{sign}(x_1)\dots\text{sign}(x_n)} F(U_1^{\text{sign}(x_1)}(x_1), \dots, U_n^{\text{sign}(x_n)}(x_n)).$$

Here $x_k \in \mathbb{R} \setminus \{0\}$.

As pointed out in Cont and Tankov (2004), constructing constant proportion Lévy copulas in higher dimensions is not very practical. In the general case 2^n positive Lévy copulas and a large number of constants must be specified. In such cases Cont and Tankov suggest using simplified constructions, such as the one in example 5.4.2 on page 55.

6

Simulation and estimation of multi-dimensional Lévy processes

6.1 Simulation of multidimensional subordinators

As shown in chapter 6 in Cont and Tankov (2004), when the dependence of components of a multidimensional Lévy process is specified via a Lévy copula, series representations of the Lévy process can be constructed using the theorem below and the probabilistic interpretation of Lévy copulas.

Theorem 6.1.1

- Let $\{V_i\}_{i\geq 1}$ be an i.i.d. sequence of random elements in a measurable space S. Assume that $\{V_i\}_{i\geq 1}$ is independent of the sequence $\{\Gamma_i\}_{i\geq 1}$ of jumping times of a standard Poisson process.

- Let $\{U_i\}_{i\geq 1}$ be a sequence of independent random variables, uniformly distributed on $[0,1]$ and independent of $\{V_i\}_{i\geq 1}$ and $\{\Gamma_i\}_{i\geq 1}$. The U_i's have the interpretation jump times.

- Let P be a probability measure on $\mathbb{R}^d$ and

$$H : (0,\infty) \times S \to \mathbb{R}^d$$

be a P-measureable function.

- Let $\boldsymbol{\sigma}$ and ν be measures on $\mathbb{R}^d$ defined by

$$\boldsymbol{\sigma}(r,B) := P(H(r,V_i) \in B), \qquad \text{for } r > 0 \text{ and } B \in \mathscr{B}(\mathbb{R}^d),$$
$$\boldsymbol{\nu}(B) := \int_0^\infty \sigma(r,B)\,\mathrm{d}r.$$

- Let

$$A(s) = \int_0^s \int_{|x|\le 1} x\sigma(r,\mathrm{d}x)\,\mathrm{d}r, \quad s \ge 0.$$

- Let ν be a Lévy measure on $\mathbb{R}^d$, that is let

$$\int_{\mathbb{R}^d} \left(\max(|x|^2, 1)\right) \boldsymbol{\nu}(\mathrm{d}\mathbf{x}) < \infty.$$

(i) **Uniform convergence to Lévy process 1:**
If the limit $\gamma = \lim_{s\to\infty} A(s)$ exists in $\mathbb{R}^d$, then the series $\sum_{i=1}^{\infty} H(\Gamma_i, V_i) 1_{U_i \le t}$ converges almost surely and uniformly on $t \in [0,1]$ to a Lévy process $\mathbf{X}_t$ with generating triplet $(0,\gamma,\nu)$. Here $\mathbf{X}_t$ has the characteristic function

$$\mathbb{E}\exp\left(i\langle u, \mathbf{X}_t\rangle\right)$$
$$= \exp\left\{ t \left[i\langle u,\gamma\rangle + \int_{\mathbb{R}^d} \left(e^{i\langle \mathbf{u},\mathbf{x}\rangle} - 1 - i\langle \mathbf{u},\mathbf{x}\rangle 1_{|\mathbf{x}|\le 1}\right) \nu(\mathrm{d}\mathbf{x}) \right] \right\}.$$

(ii) **Uniform convergence to Lévy process 2:**
If $\boldsymbol{\nu}$ is a Lévy measure on $\mathbb{R}^d$ and for each $v \in S$ the function

$$r \mapsto |H(r,v)| \quad \text{is nonincreasing as a function of } r,$$

then as $N \to \infty$ $\sum_{i=1}^{N} \left(H\left(\Gamma_i, V_i\right) 1_{U_i \le t} - t\mathbf{c_i}\right)$ converges almost surely and uniformly on $t \in [0,1]$ to a Lévy process with characteristic triplet $(\mathbf{0},\mathbf{0},\boldsymbol{\nu})$. Here the $\mathbf{c}_i$ are deterministic constant vectors given by $\mathbf{c_i} = A(i) - A(i-1)$.

Proof: This theorem is found on page 195 in Cont and Tankov (2004).

Theorem 6.1.2: Series representation of two-dimensional subordinator

Let $(\mathbf{Z}_t)$ be a two-dimensional Lévy process with positive jumps, marginal tail integrals U_i and U_2, and Lévy copula $F(x,y)$. If F is continuous on $[0,\infty]^2$ then the process $\mathbf{Z}$ is representable in law, on the time interval $[0,1]$, as

$$\{\mathbf{Z}_s, 0 \leq s \leq 1\} \stackrel{d}{=} \left\{\tilde{\mathbf{Z}}_s, 0 \leq s \leq 1\right\}$$

where

$$\tilde{Z}_s^{(1)} = \sum_{i=1}^{\infty} U_1^{(-1)}\left(\Gamma_i^{(1)}\right) 1_{[0,s]}(V_i)$$
$$\tilde{Z}_s^{(2)} = \sum_{i=1}^{\infty} U_2^{(-1)}\left(\Gamma_i^{(2)}\right) 1_{[0,s]}(V_i). \tag{6.1}$$

Here

- The (V_i) are independent and uniformly distributed on $[0,1]$.
- $\left(\Gamma_i^{(1)}\right)$ is an independent sequence of jump times for a standard Poisson process.
- For every i, $\Gamma_i^{(2)}$ conditionally on $\Gamma_i^{(1)}$ is independent of all other variables.
- Viewed as a function of y, $\Gamma_i^{(2)}$ has the distribution function $\frac{\partial}{\partial x} F(x,y)\Big|_{x=\Gamma_i^{(1)}}$.
- All the series in 6.1 converge almost surely and uniformly on $s \in [0,1]$.

This theorem is stated on page 200 in Cont and Tankov (2004). It depends on proposition 6.3 in Cont and Tankov (2004).

The results given above and the probabilistic interpretation of positive Lévy copulas make it possible (see remark 6.7 in Cont and Tankov (2004)) to simulate a two-dimensional Lévy process $\mathbf{X}_t = (X_t, Y_t)$. The dependence structure of X_t and Y_t is specified by a positive Lévy copula F.

The simulating algorithm $\mathbf{X}_t$, found on page 202 in Cont and Tankov (2004), is based on enumerating the jumps of X_t in descending order and simulating the jumps in Y_t conditionally on the size of the jumps in the first component. Let U_1 be the tail integral of X_t and U_2 be the tail integral of Y_t.

Algorithm 1: Simulation of a two-dimensional subordinator with dependent components by series representations

Fix a number τ^* dsepending on the required precision and computational capacity. This number is equal to the average number of terms in the series and determines the truncation level: Jumps in X_t smaller than $U_1^{(-1)}(\tau^*)$ are truncated.

- Initialize $k = 0$, $\Gamma_0^{(1)} = 0$.
- REPEAT WHILE $\Gamma_k^{(1)} < \tau^*$
- Set $k = k + 1$
- Simulate T_k: standard exponential
- Set $\Gamma_k^{(1)} = \Gamma_{k-1}^{(1)} + T_k$ Obtaining the transformed jump in the first component
- Simulate $\Gamma_k^{(2)}$ from distribution function $F_1(y) = \left.\frac{\partial F(x,y)}{\partial x}\right|_{x=\Gamma_k^{(1)}}$ (obtaining the transformed jump in the second component)
- Simulate V_k : uniformly distributed variable on $[0, 1]$ (obtaining the time when the jump occurs)

The trajectory is then given by

$$X_t = \sum_{i=1}^{k} 1_{V_i \le t} U_1^{(-1)} \left(\Gamma_i^{(1)}\right),$$
$$Y_t = \sum_{i=1}^{k} 1_{V_i \le t} U_2^{(-1)} \left(\Gamma_i^{(2)}\right).$$

Example 6.1.3: Linking via a positive Clayton Lévy copula

If X_t and Y_t are linked with a positive Clayton Lévy copula C_θ, that is, a positive Lévy copula of the form

$$C_\theta(x,y) = \left(x^{-\theta} + y^{-\theta} - 1\right) \quad \text{for } \theta > 0, \tag{6.2}$$

then the conditional distribution function F of $\Gamma_k^{(2)}$ given $\Gamma_k^{(1)}$ takes the form

$$F(\Gamma_k^{(2)}|\Gamma_k^{(1)}) = \frac{\partial C_\theta(\Gamma_k^{(1)}, \Gamma_k^{(2)})}{\partial \Gamma_k^{(2)}} = \left\{ 1 + \left(\frac{\Gamma_k^{(1)}}{\Gamma_k^{(2)}} \right)^\theta \right\}^{-1-1/\theta}. \tag{6.3}$$

Since $F(\Gamma_k^{(2)}|\Gamma_k^{(1)})$ is uniformly distributed on $[0,1]$, Cont and Tankov suggest simulating this distribution by inverting equation 6.3.

This produces the following inverse:

$$F^{-1}(W_k|\Gamma_k^{(1)}) = \Gamma_k^{(1)} \left(W_k^{-\frac{\theta}{1+\theta}} - 1 \right)^{-1/\theta}.$$

The two-dimensional subordinator then has the representation

$$X_s = \sum_{i=1}^{\infty} U_1^{(-1)} \left(\Gamma_i^{(1)} \right) 1_{[0,s]}(V_i)$$

$$Y_s = \sum_{i=1}^{\infty} U_2^{(-1)} \left(F^{-1}(W_i|\Gamma_i^{(1)}) \right) 1_{[0,s]}(V_i),$$

where (W_i) and (V_i) are independent sequences of independent random variables, uniformly distributed on $[0,1]$, and $(\Gamma_i^{(1)})$ is an independent sequence of jump times of a standard Poisson process.

6.2 Implementations of algorithm 1

The source code for two implementations of algorithm 1 with marginal positive 1/2-stable processes, the files "Levy.R" and "GPD.R", can be downloaded from `http://www.student.uib.no/~mhu080/master`.

6.2.1 1/2 stable processes

Recall that the Lévy distribution (i.e. a positive stable distribution with index $\frac{1}{2}$) has a Lévy density of the form $\nu(x) = \frac{\sigma}{2\sqrt{\pi}} \frac{1}{x^{3/2}} 1_{x>0}$.

The general one-dimensional tail integral is defined as

$$U(x) := \begin{cases} \infty & x = 0 \\ \nu([x, \infty)), & x > 0). \end{cases}$$

Working out the implied integration we have the following:

$$U(x) = \begin{cases} \infty & \text{for} x = 0 \\ 2\left(\frac{\sigma}{2\sqrt{\pi}}\right) x^{-1/2} & \text{for } x > 0. \end{cases}$$

The quasi-inverse is thus

$$U^{(-1)}(y) = \left(\frac{2\left(\frac{\sigma}{2\sqrt{\pi}}\right)}{y}\right)^2 = \frac{\sigma^2}{\pi}\frac{1}{y^2} \quad \text{for } 0 < y < \infty.$$

In the code we have assumed parameter values $\mu = 0$ and $\sigma = 1$ in the Lévy distribution (see equation 3.20 on page 33) for both marginals.

If figure 6.1 on the following page below are some simulated trajectories for 4 different values of the Clayton Lévy copula.

6.2.2 Compound Poisson marginals

In another implementation of of algorithm 1 the marginal subordinators are two equal compound Poisson processes, where the jump distribution is a generalized Pareto distribution.

The distribution function of the three-parameter generalized Pareto distribution is as follows (see page 162 in Embrechts *et al.* (1999)):

$$G_{\xi;\tau;\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{\beta}(x - \tau)\right)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - e^{-\frac{(x-\tau)}{\beta}} & \text{if } \xi = 0, \end{cases} \tag{6.4}$$

where $\beta > 0$ and

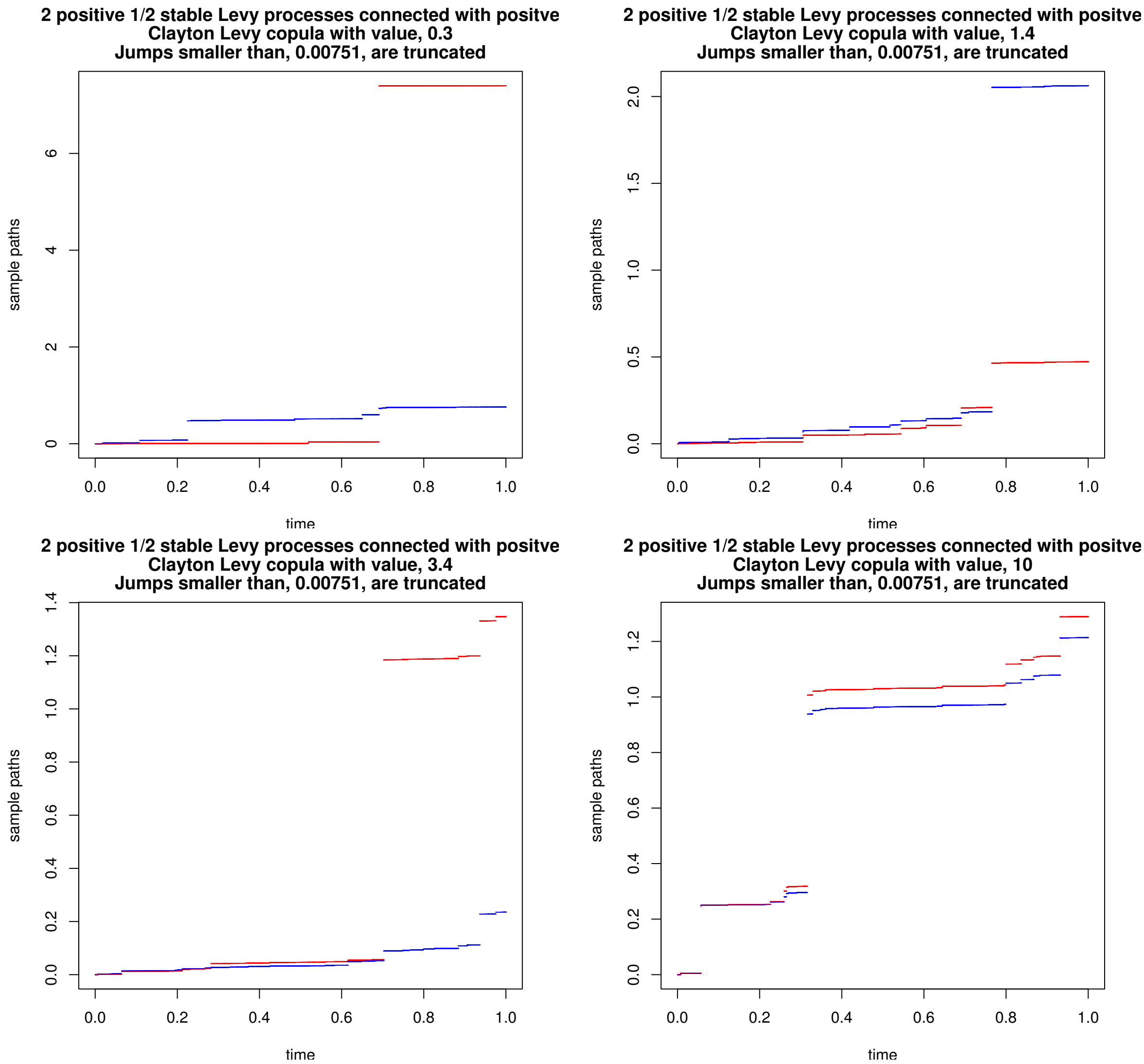

Figure 6.1: 4 simulations of a two-dimensional Lévy process $\mathbf{X}_t = (X_t, Y_t)$. The components X_t and Y_t are both $\frac{1}{2}$-stable processes with positive jumps and are linked together with a positive Clayton Lévy copula. In the upper left corner the value of the parameter θ for the Clayton Lévy copula is set to 0.3, which corresponds to a weak dependence. In the upper right corner and lower left corner the dependence is stronger and the components tend to jump simultaneously, but with different jump sizes. In the bottom right corner θ is set to 10, which corresponds to a very strong dependence. Here the components almost follow each other.

$$x \in \begin{cases} [\tau, \infty) & \text{if } \xi \geq 0, \\ [\tau, \tau - \beta/\xi] & \text{if } \xi < 0. \end{cases}$$

The parameter ξ is called the *shape* parameter, the parameter τ is called the *threshold*, and the parameter β is called the *scale* parameter.

A one-dimensional compound Poisson process X_t with intensity parameter λ and jump probability measure F has a characteristic function of the form (see proposition 3.4 in Cont and Tankov (2004))

$$\widehat{\mathbb{P}}_{X_t}(u) = \exp\left\{ t\lambda \int_{-\infty}^{\infty} (e^{iux} - 1)\, \mathrm{d}F(x) \right\}.$$

By comparison with the Lévy-Khintchine formula we see that the Lévy measure here is $\nu(A) = \lambda F(A)$.

The tail integral for the positive random variable X is thus

$$U(x) = \begin{cases} \infty & x = 0 \\ \lambda \bar{F}(x) & x > 0, \text{ where } \bar{F}(x) = \mathbb{P}(X > x). \end{cases} \tag{6.5}$$

In the particular case that the jump distribution is the generalized Pareto distribution (GPD) with $\xi > 0$, defined in equation 6.4 on page 65 above, has $\xi > 0$, we get the following for the tail integral of X :

$$U(x) = \begin{cases} \infty & \text{if } x = 0 \\ \lambda & \text{if } 0 < x < \tau \\ \lambda \left(1 + \frac{\xi}{\beta}(x - \tau)\right)^{-1/\xi} & \text{if } x > \tau. \end{cases} \tag{6.6}$$

Now we want to get $U^{(-1)}(\tau^*)$:

$$U^{(-1)}(y) := \sup\{x : U(x) \geq y\}.$$

By solving the inequality

$$\lambda \left(1 + \frac{\xi}{\beta}(x - \tau)\right)^{-1/\xi} \geq \Gamma$$

we have

$$U^{-1}(\Gamma) = \begin{cases} 0 & \text{if } \Gamma > \lambda \\ \tau + \frac{\beta}{\xi}\left[\left(\frac{\lambda}{\Gamma}\right)^{\xi} - 1\right] & \text{if } \Gamma \leq \lambda. \end{cases}$$

Below are the trajectories of some simulations of compound Poisson processes with generalized Pareto distributed jumps, with parameters $\xi = 0.618$, $\beta = 1$ and different values for the Clayton Lévy copula parameter θ.

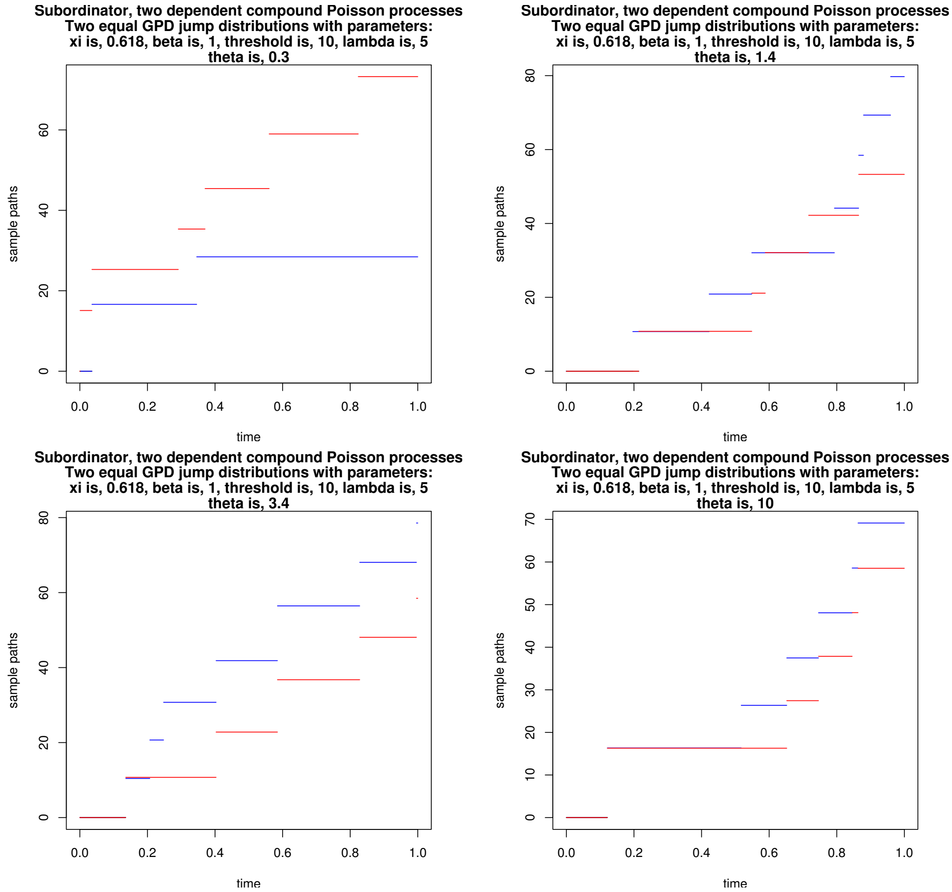

Figure 6.2: Simulation of compound Poisson processes linked with positive Clayton Lévy copulas. In the upper left corner the Clayton Lévy copula parameter θ is set to 0.3, which corresponds to a very weak dependence. The other three corners correspond to increasing dependence. Compared with the simulated trajectories in figure 6.1 on page 66 we see that there are fewer jumps. In particular there are no small jumps.

6.3 Simulation of general stable processes linked with Lévy copulas

Let $F^{++}, F^{--}, F^{+-}, F^{--}$ be positive Lévy copulas and $U^{++}, U^{+-}, U^{-+}, U^{--}$ be tail integrals on $\mathbb{R}^2$, as in definition 5.4.5 on page 56.

It is then shown on page 160 of Cont and Tankov (2004) that

if $c_1, c_2, \dots c_4$ are constants between 0 and 1, a Lévy copula can be defined as follows:

$$F(x,y) = F^{++}\left(c_1|x|, c_2|y|\right) 1_{x\geq 0, y\geq 0} + F^{--}\left(c_3|x|, c_4|y|\right) 1_{x\leq 0, y\leq 0}$$
$$- F^{+-}\left((1-c_1)|x|, (1-c_4)|y|\right) 1_{x\geq 0, y\leq 0} - F^{-+}\left((1-c_3)|x|, (1-c_2)|y|\right) 1_{x\geq 0, y\leq 0}. \tag{6.7}$$

With fixed marginal Lévy measures ν_1 and ν_2, the upper right-hand tail integral is

$$U^{++}(x,y) = F^{++}\left(c_1 U_1^+(x), c_2 U_2^+(y)\right).$$

Thus the upper right-hand quadrant corresponds to a Lévy process with positive jumps, Lévy copula F^{++}, and marginal Lévy measures $c_1\nu_1(\mathrm{d}x)1_{x>0}$ and $c_2\nu(\mathrm{d}y)1_{y>0}$.

Treating the other quadrants in the same manner, Cont and Tankov conclude that a Lévy process with Lévy copula of the form of equation 6.7 is a sum of four independent parts, corresponding to the four quadrants of the Lévy measure. For the first independent part, corresponding to the upper right-hand quadrant and with linking via the positive Lévy copula F^{++}:

- The 1st component jumps upward and has Lévy measure $c_1\nu_1(\mathrm{d}x)1_{x>0}$.
- The 2nd component jumps upward and has Lévy measure $c_2\nu_2(\mathrm{d}y)1_{y>0}$.

For the third independent part, corresponding to the lower right-hand quadrant, and with linking via the positive Lévy copula F^{+-}:

- The 1st component jumps upward and has Lévy measure $(1-c_1)\nu(\mathrm{d}x)1_{x>0}$.
- The 2nd component jumps downward and has Lévy measure $(1-c_4)\nu(\mathrm{d}x)1_{x<0}$.

The other independent parts of the Lévy process can be characterized in the same way. For example, an α-stable process with $\alpha < 2$ has a Lévy measure of the following form:

$$\nu(x) = \frac{A}{x^{\alpha+1}} 1_{x>0} + \frac{B}{|x|^{\alpha+1}} 1_{x<0}.$$

Therefore it is clear that Lévy copulas of all α-stable processess with $\alpha < 2$ can be represented in the form of equation 6.7 on the preceding page. That equation can thus be used to model bivariate Lévy processes with stable margins.

We have used this approach to implement a simulation algorithm for a truncated bivariate Lévy process, with α-stable margins linked via a Clayton Lévy copula in each quadrant. The blue component is marginally α-stable with index 1.75 and the red component is marginally α-stable with index 1.9. The other specifications are $c_1 = c_2 = c_4 = c_4 = 0.5$ and $A_X = B_X = A_Y = B_Y = 1$. The blue component is truncated at 0.0025 (no jumps of the blue component smaller than 0.0025 are being plotted).

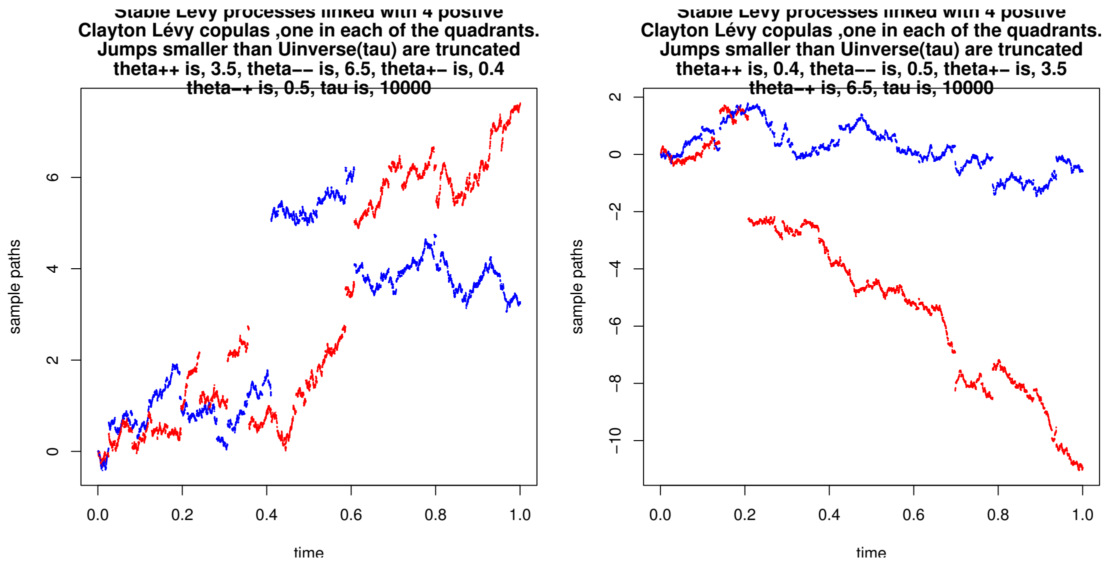

Figure 6.3: Two simulations of a Lévy process on $\mathbb{R}^2$ with stable marginals. The components are linked together with 4 Clayton Lévy copulas, one in each quadrant. In the figure to the left the components are positively correlated and tend to have jumps of the same sign. In the figure to the right the components are negatively correlated and tend to have jumps of opposite signs.

6.4 Estimation of a positive Lévy copula

Let X_t and Y_t be two compound Poisson subordinators, each having a known intensity parameter and each having a known jump distribution on $[0, \infty)$.

Let $(x_1, y_1), \ldots, (x_{n_1}, y_{n_1})$ be pairs of observed simultaneous jumps of X_t and Y_t.

Let $x_1^*, \ldots, x_{n_2}^*$ be the observed jumps of X_t not corresponding to any observed jumps of Y_t. Similarly, let $y_1^*, \ldots, y_{n_3}^*$ be the observed jumps of Y_t not corresponding to any observed jumps of X_t.

Recall that in the simulation algorithm 1 the trajectory was given by

$$X_t = \sum_{i=1}^{k} 1_{V_i \le t} U_1^{(-1)} \left(\Gamma_i^{(1)}\right) \tag{6.8}$$

$$Y_t = \sum_{i=1}^{k} 1_{V_i \le t} U_2^{(-1)} \left(\Gamma_i^{(2)}\right), \tag{6.9}$$

$$\tag{6.10}$$

where $U_1^{(-1)}$ and $U_2^{(-1)}$ were the quasi-inverses of the tail integrals. We estimate the $\Gamma_i^{(j)}$'s $\quad j = 1, 2$ below.

We make the assumption that the observations were made with the parameters of X_t and Y_t all known. In particular, X_t is assumed to have a known intensity parameter λ_X and Y_t is assumed to have a known intensity parameter λ_Y. , $k = 1, 2$

Under this assumption we try to restore the latent variables $\Gamma_i^{(j)}$'s, from equations 6.8 to 6.9 by making the following transformations:

$\Gamma_1^{(1)} = U_1(x_1), \ldots, \Gamma_{n_1}^{(1)} = U_1\left(x_{n_1}\right),$

$\Gamma_1^{(2)} = U_2(y_1), \ldots, \Gamma_{n_1}^{(2)} = U_2\left(y_{n_1}\right),$

$\Gamma_1^{(*1)} = U_1\left(x_1^*\right), \ldots, \Gamma_{n_2}^{(*1)} = U_1\left(x_{n_2}^*\right)$

and $\Gamma_1^{(*2)} = U_2\left(y_1^*\right), \ldots, \Gamma_{n_3}^{(*1)} = U_2\left(x_{n_3}^*\right)$.

In algorithm 1, with dependence given by a Clayton Lévy copula $\Gamma_i^{(2)}$ conditioned on $\Gamma_i^{(1)}$, the following distribution function was simulated:

$$F_{\Gamma_i^{(2)}|\Gamma_i^{(1)}=\gamma_1}(\gamma_2) = \left\{1 + \left(\frac{\gamma_1}{\gamma_2}\right)^{\theta}\right\}^{-1-1/\theta}. \tag{6.11}$$

(See equation 6.3 on page 64.)

The idea is to estimate the value of the Clayton Lévy copula using the method of maximum likelihood on this distribution function and using the restored latent variables $\Gamma_i^{(j)}$'s. By differentiating this distribution function we obtain the following probability density:

$$f_{\Gamma_i^{(2)}|\Gamma_i^{(1)}=\gamma_1}(\gamma_2) = (1+\theta)\left(1 + \left(\frac{\gamma_1}{\gamma_2}\right)^{\theta}\right)^{-(2+1/\theta)}\left(\left(\frac{\gamma_1}{\gamma_2}\right)^{\theta}\gamma_2^{-1}\right). \tag{6.12}$$

In the case of simultaneous jumps we use this density as the likelihood.

Recall that the tail integral is

$$U(x) = \begin{cases} \infty & x = 0 \\ \lambda \bar{F}(x) & x > 0. \end{cases} \tag{6.13}$$

For the case that X_t has an observed jump x_i^* while Y_t does not have a jump we therefore have the following probability:

$$\mathbb{P}\left(\Gamma_i^{(2)} \in \left\{\gamma_2 : U_2^{(-1)}(\gamma_2) = 0\right\} \middle| \Gamma_i^{(*1)} = U_1(x_i^*)\right) \tag{6.14}$$

$$= \mathbb{P}\left(\Gamma_i^{(2)} > \lambda_Y | \Gamma_i^{(*1)} = U_1(x_i^*)\right) \tag{6.15}$$

$$= 1 - \left(1 + \left(\frac{\Gamma_i^{(*1)}}{\lambda_Y}\right)^{\theta}\right)^{-1-1/\theta}. \tag{6.16}$$

Correspondingly, when Y_t has an observed jump y_i^* while X_t does not have a jump, we interchange X_t and Y_t and have the probability

$$1 - \left(1 + \left(\frac{\Gamma_i^{(*2)}}{\lambda_X}\right)^{\theta}\right)^{-1-1/\theta}. \tag{6.17}$$

Using equation 6.12 on the preceding page, equation 6.14 on the facing page and equation 6.17 on the preceding page we obtain the following log-likelihood function:

$$
\begin{aligned}
& l\left(\theta | x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_1}, x_1^*, \ldots, x_{n_2}^*, y_1^*, \ldots, y_{n_3}^*\right) \\
&= \sum_{i=1}^{n_1} \left\{ \log(1+\theta) - \left(2+\frac{1}{\theta}\right) \log\left(1 + \left(\frac{\Gamma_i^{(1)}}{\Gamma_i^{(2)}}\right)^{\theta}\right) + \theta\left(\log\Gamma_i^{(1)} - \log\Gamma_i^{(2)}\right) - \log\Gamma_i^{(2)} \right\} \\
&\quad + \sum_{i=1}^{n_2} \log\left\{1 - \left(1 + \left(\frac{\Gamma_i^{(*1)}}{\lambda_Y}\right)^{\theta}\right)^{-1-1/\theta}\right\} + \sum_{i=1}^{n_3} \log\left\{1 - \left(1 + \left(\frac{\Gamma_i^{(*2)}}{\lambda_X}\right)^{\theta}\right)^{-1-1/\theta}\right\}. \\
&= n_1 \log(1+\theta) - \left(2+\frac{1}{\theta}\right) \sum_{i=1}^{n_1} \log\left(\left(\Gamma_i^{(1)}\right)^{\theta} + \left(\Gamma_i^{(2)}\right)^{\theta}\right) + \theta \sum_{i=1}^{n_1} \left(\log\Gamma_i^{(1)} + \Gamma_i^{(2)}\right) \\
&\quad + \sum_{i=1}^{n_2} \log\left\{1 - \left(1 + \left(\frac{\Gamma_i^{(*1)}}{\lambda_Y}\right)^{\theta}\right)^{-1-1/\theta}\right\} + \sum_{i=1}^{n_3} \log\left\{1 - \left(1 + \left(\frac{\Gamma_i^{(*2)}}{\lambda_X}\right)^{\theta}\right)^{-1-1/\theta}\right\}.
\end{aligned}
\tag{6.18}
$$

Optimization of this function must be done numerically. In our implementation we use the R function “optimize” from the package “stats” to maximize equation 6.18.

Newton’s binomial theorem says that if $-1 < x < 1$ and $\alpha \in \mathbb{R}$ then (see page 201 in Rudin (1976))

$$
(1+x)^{\alpha} = 1 + \sum_{n=1}^{\infty} \frac{\alpha(\alpha-1)\ldots(\alpha-n+1)}{n!} x^n. \tag{6.19}
$$

Let $c = 1 + \left(\frac{\Gamma_i^{(*2)}}{\lambda_X}\right)^{\theta}$ and $\alpha = -1 - 1/\theta$.

In the event that Γ_i^{*2} a is close to zero we are faced with the task of calculating $\log(1-c^{\alpha})$ with c^{α} almost equal to one.

To avoid numerical problems we then use a truncated version of equation 6.19 and calculate $\log(1-c^{\alpha})$ as

$$\begin{aligned}
&\log(1-c^{\alpha}) \\
&= \log\left(1-\left(1+\sum_{n=1}^{N}\frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!}c^{n}\right)\right) \\
&= \log\left(\sum_{n=1}^{N}\frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!}c^{n}\right),
\end{aligned}$$

where the N is chosen high enough that the residual is small.

In the event that Γ_i^{*1} is close to zero we make the same calculation, but with $c = 1 + \left(\frac{\Gamma_i^{(*1)}}{\lambda_Y}\right)^{\theta}$.

To get some idea how this estimatator performs we tried the estimator on simulated observations from two compound Poisson processes on $\mathbb{R}$, X_t and Y_t, linked with a Clayton Lévy copula. We let X_t and Y have generalized Pareto distributed jumps with $\xi = 0.618$, $\beta = 1$, threshold 10, intensity parameter $\lambda = 25$, and tried different values of the Clayton Lévy copula parameter θ. Under the assumption of known values of all the parameters except θ, we then compared the obtained point estimates of θ, with the real value. We let the time period be 1, so letting $\lambda = 25$ caused the expected number of jumps from each component was 25. We repeated this experiment 10000 times for each value of θ. In our evalutation of the estimator we considered the bias and the root mean square error (RMSE) defined as $\sqrt{n^{-1}\sum_{i=1}^{n}\left(\theta-\hat{\theta}_i\right)^2}$, where $\hat{\theta}$ is the ith estimate of θ.

θ	0.3	0.7	3.3	10
Bias	-0.014	0.016	0.106	0.330
RMSE	0.108	0.171	0.706	1.971
2.5 percentile	0.014	0.434	2.272	7.189
97.5 percentile	0.469	1.099	5.0225	14.716

Table 6.1: Bias, RMSE, 2.5 percentile and 97.5 percentile for $\theta = 0.3$, $\theta = 0.7$, $\theta = 3.3$ and $\theta = 10$.

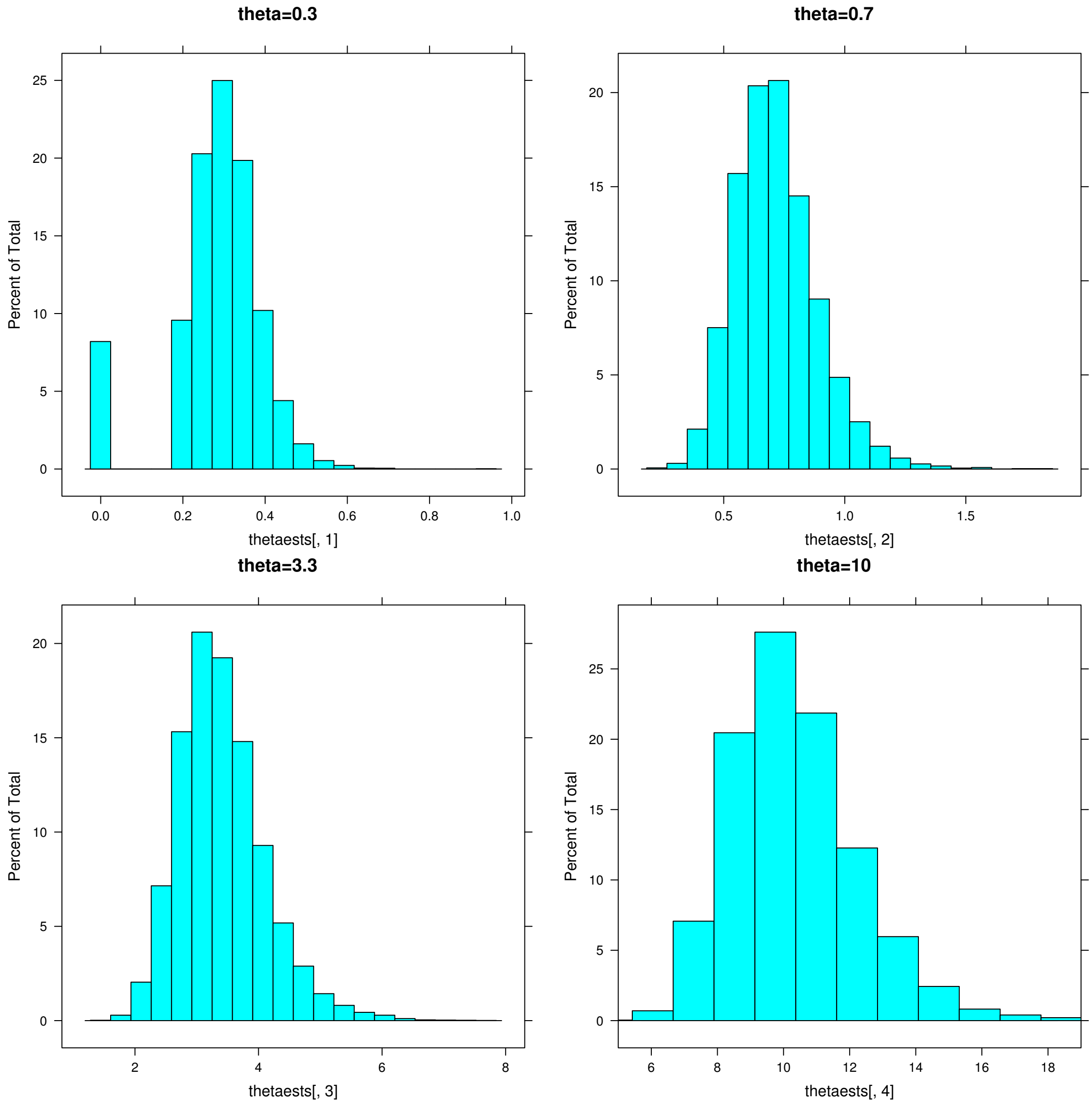

Figure 6.4: Histograms from the simulation experiment: In the upper left hand corner: $\theta = 0.3$, upper right hand corner: $\theta = 0.7$, lower left hand corner: $\theta = 3.3$: lower right hand corner: $\theta = 10$. We see that the estimates are less precise for large values of the Clayton Lévy copula parameter θ.

7 Application to Danish fire insurance data

7.1 Application of positive Lévy copulas to ruin theory

7.1.1 Some classical Ruin theory

We begin this discussion by giving a few basic definitions as well as classical ruin theory results taken from Bregman and Klüppelberg (2005).

In this discussion the net risk portfolio of an insurance company is modelled as a multivariate Lévy process $R_t = (R_t^{(1)}, R_t^{(2)}, \dots R_t^{(d)}), \quad t \geq 0$. The corresponding net risk reserve of the insurance company is given by the stochastic process $R^+ = \left(R_t^+\right)_{t \geq 0}$, where

$$R_t^+ = R_t^{(1)} + R_t^{(2)} + \dots + R_t^{(d)}, \quad t \geq 0.$$

Here each component can be taken as a risk process $R_t^{(i)} = x_i + c_i t - C_t^{(i)}, t \geq 0$, for initial risk reserves $x_i \geq 0$, premium rates $c_i > 0$ and for all $i = 1, \dots, d$. Then

$$R_t^+ = x + ct - C_t^+, \quad t \geq 0$$

for $x = \sum_{i=1}^d x_i$, $c = \sum_{i=1}^d c_i$ and $C_t^+ = \sum_{i=1}^d C_t^{(i)}$.

For the initial risk reserve $x \geq 0$ the *ruin probability* is defined as

$$\Psi(x) := \mathbb{P}\left(R_t^+ < 0 \quad \text{for some } t \geq 0\right).$$

Before further discussion of ruin probability we need to introduce a few concepts regarding the tails of probability distributions.

Definition 7.1.1

(a) A Lebesgue-measurable function $h : [0, \infty) \to (0, \infty)$ is *regularly varying* with index $\gamma \in \mathbb{R}$ (written as $h \in \mathscr{R}_\gamma$) if

$$\lim_{x\to\infty} \frac{h(rx)}{h(x)} = r^\gamma, \quad \text{for all } r > 0.$$

(b) Let F be the distribution function of a positive random variable X. We define the survival function $\bar{F}(x)$ of X as $\bar{F}(x) := 1 - F(x)$.

Denote by $F^{2*} = F * F$ the convolution of F with itself and by $\bar{F}^{2*}$ the survival function $1 - F^{2*}$. F or X is called *subexponential*, if

$$\bar{F}^{2*} \sim 2\bar{F}(x), \quad \text{as } x \to \infty.$$

As noted in Bregman and Klüppelberg (2005) all distribution functions with regularly varying tails are included in the class of subexponential distributions.

For subexponential distributions we have the following ruin theory result (see Theorem 2.11 in Bregman and Klüppelberg (2005) and theorem 1.3.6 in Embrechts *et al.* (1999)):

Theorem 7.1.2

Let C be a compound Poisson process with Poisson rate $\lambda > 0$.
Then the corresponding risk process is

$R_t = x + ct - C_t = x + ct - \sum_{i=1}^{N_t} Y_i, \quad t \geq 0$, where N_t is a Poisson process with rate λ and the $Y_i's$ are i.i.d. random variables.

If the claims Y_i obey a subexponential distribution F and have a finite expectation $\mathbb{E}Y$, then, under the net profit condition $c - \lambda\mathbb{E}Y > 0$, we obtain the ruin probability

$$\Psi(x) \sim \frac{\lambda}{c - \lambda\mathbb{E}Y} \int_x^\infty \bar{F}(y)\,dy, \quad \text{as } x \to \infty.$$

Here $\bar{F}$ is the survival function of the claim distribution F.

Combined with Karamata's theorem (see page 28 in Bingham *et al.* (1987)) we get that if $\bar{F} \in \mathscr{R}_{-b}$ and $b > 1$, then

$$\Psi(x) \sim \frac{\lambda}{c - \lambda \mathbb{E} Y} \frac{x}{b-1} \bar{F}(x) \quad \text{as } x \to \infty. \tag{7.1}$$

7.2 The Clayton risk process

In this section we will be employing the positive Clayton Lévy copula and tail integrals (see definition 5.3.1 on page 51) in a ruin theory context, with heavy-tail claims obeying a generalized Pareto distribution.

Definition 7.2.1

$C = (C^{(1)}, C^{(2)})$ denote a bivariate subordinator. Define $C^{+} := C^{(1)} + C^{(2)}$. Then C^{+} has tail integral

$$U^{+}(z) = \nu\left(\left\{(x,y) \in [0,\infty)^2 : x + y \geq z\right\}\right),$$

where $z \geq 0$ and ν is the Lévy measure.

As in Bregman and Klüppelberg (2005), we define $U^{+}(0)$ as the limit from the right, $\lim_{x \to 0^{+}} U(x)$.

Let $C^{(1)}, C^{(2)}$ denote compound Poisson processes with rates λ_1 and $\lambda_2 > 0$ and claim size distribution functions F_1 and F_2. Let the dependence between C^1, C^2 be given by a positive Clayton Lévy copula with parameter $\theta \in [0, \infty)$. Then the following can be shown:

Proposition 7.2.2

(a) The process C^+ defined in definition 7.2.1 on the preceding page is a compound Poisson process with tail integral given by

$$U^+(z) = I_1(z) + I_2(z) + I_3(z). \tag{7.2}$$

Here for $z > 0$,

$$\begin{aligned} I_1(z) &= \lambda_1 \lambda_2^{\theta+1} \int_{(0,z)} \left(\frac{\bar{F}_2^{\theta}(z-x)}{\lambda_1^{\theta} \bar{F}_1^{\theta}(x) + \lambda_2^{\theta} \bar{F}_2^{\theta}(z-x)} \right)^{\frac{\theta+1}{\theta}} F_1(\mathrm{d}x), \\ I_2(z) &= \lambda_1 \lambda_2 \bar{F}_1(z) \left(\lambda_1^{\theta} \bar{F}_1^{\theta}(z) + \lambda_2^{\theta} \right)^{-1/\theta}, \\ I_3(z) &= \lambda_1 \bar{F}_1(z) + \lambda_2 \bar{F}_2^{\theta}(z) - \left(\lambda_1^{-\theta} \bar{F}_1^{-\theta}(z) + \lambda_2^{-\theta} \right)^{-1/\theta} \\ &\quad - \left(\lambda_1^{-\theta} + \lambda_2^{-\theta} \bar{F}_2^{-\theta}(z) \right)^{-1/\theta}. \end{aligned}$$

Moreover,

$$I_2(z) \sim \lambda_1 \bar{F}_1(z), \quad z \to \infty. \tag{7.3}$$

(b) Assume now that $F = F_1 = F_2$ and $\lambda = \lambda_1 = \lambda_2$. Then

$$U^+(z) = \lambda \left(I_1'(z) + I_2'(z) + 2I_3'(z) \right)$$

where

$$\begin{aligned} I_1'(z) &= \int_{(0,z)} \left(\frac{\bar{F}(z-x)}{\bar{F}^{\theta}(z-x) + \bar{F}^{\theta}(x)} \right)^{\frac{\theta+1}{\theta}} F(\mathrm{d}x), \\ I_2'(z) &= \bar{F}(z) \left(\bar{F}^{\theta}(z) + 1 \right)^{-1/\theta} \sim \bar{F}(z), \quad z \to \infty \text{ and} \\ I_3'(z) &= \bar{F}(z) \left(1 - (1 + \bar{F}^{\theta}(z))^{-1/\theta} \right) = o(\bar{F}(z)), \quad z \to \infty. \end{aligned}$$

For a proof see page 11 in Bregman and Klüppelberg (2005).

With Pareto claim sizes (as opposed to generalized Pareto) the following theorem and corollary is proven on page 14 in Bregman and Klüppelberg (2005):

Theorem 7.2.3

Let $C^{(1)}, C^{(2)}$ be compound Poisson processes, both with rate $\lambda > 0$ and with Pareto claim sizes having survival functions as follows:

$$\bar{F}(x) = \left(\frac{a}{a+x}\right)^b, \quad x > 0, \tag{7.4}$$

for some $a > 0$ and $b > 1$. Assume that the dependence between C^1, C^2 is given by the positive Clayton copula S_θ for $\theta \in (0, \infty)$.

Then

$$U^+(x) = \lambda\left(K_* x^{-b} + \bar{F}(x) + o\left(\bar{F}(x)\right)\right), \quad x \to \infty,$$

for some constant $K_* > 0$ not dependent of x. In particular $U^+(\cdot) \in \mathscr{R}_{-b}$.

For convenience we define $\bar{K_*}(b, \theta) = a^{-b} K_*$.

Corollary 7.2.4

Suppose that the conditions of theorem 7.2.3 hold. Assume also that the net profit condition holds. Note that since the process C^+ is the sum of two compound process, each with Poisson rate λ, the net profit condition is in this case $c - 2\lambda\,\mathbb{E}\,Y > 0$. Then the ruin probability viewed as a function of the initial reserve x is

$$\Psi(x) \sim \frac{\lambda}{c - 2\lambda\,\mathbb{E}\,Y} \frac{a^b}{(b-1)} \left(\bar{K_*}(b, \theta) + 1\right) x^{-(b-1)}, \quad x \to \infty.$$

Proof: A proof is given on page 15 in Bregman and Klüppelberg (2005).

In this connection note that the tail of the generalized Pareto distribution (GPD) is of the same form, under the following conditions: Positive shape parameter (ξ) and positive scaling parameter (β). Then in $\bar{F} \sim a^b x^{-b}$, $\frac{\beta}{\xi}$ corresponds to a and $\frac{1}{\xi}$ corresponds to b. Thus we would expect a similar asymptotic result to be true also for the generalized Pareto distribution, with the above-mentioned substitutions. This is indeed true, with a proof very similar to the proof for theorem 7.2.3 given in Bregman and Klüppelberg (2005). We get the following result:

Theorem 7.2.5

Let $C^{(1)}, C^{(2)}$ be compound Poisson processes, both with rate $\lambda > 0$ and with generalized Pareto distributed claim sizes having the following survival functions:

$$\bar{F}(x) = \left(1 + \frac{\xi}{\beta}(x - \tau)\right)^{-1/\xi}, \quad x > \tau, \tag{7.5}$$

for $\beta, \tau > 0$ and $\xi > 0$. Assume that the dependence between C^1, C^2 is given by the positive Clayton copula S_θ, with $\theta \in (0, \infty)$,

and that $0 < \xi < 1$.

Then the ruin probability $\Psi(x)$ viewed as a function of the initial reserve x is

$$\Psi(x) \sim \frac{\lambda}{c - 2\lambda \mathbb{E}Y} \frac{\left(\frac{\beta}{\xi}\right)^{1/\xi}}{\left(\frac{1}{\xi} - 1\right)} \left(\bar{K}(\xi, \theta) + 1\right) x^{-(1/\xi - 1)}, \quad \text{as } x \to \infty \tag{7.6}$$

for some constant $\bar{K}(\xi, \theta) > 0$.

Proof: A proof is given in section A.1 on page 103.

7.3 About the data and the model

For his article ((McNeil (1997)) Alexander McNeil choose to split the Danish fire insurance claims data into three categories: damage on building, damage to furniture and personal property, and a loss of profits category. This dataset can be downloaded either from his website
`http://www.ma.hw.ac.uk/~mcneil/data.html`
or from
`http://www.student.uib.no/~mhu080/master`.
The claim sizes are adjusted for inflation as explained in McNeil (1997).

In our experiment we used the categories "damage to building" and "damage to furniture and property" from McNeil's
multivariate Danish fire insurance claims dataset. In our model the
excess of claims above a certain threshold τ is assumed to obey a generalized Pareto distribution. The waiting times between claims of size greater than τ are assumed to obey an exponential distribution. This amounts to a model with two one-dimensional compound Poisson processes, (X_t, Y_t), where X_t is the sum of claims due to damage to buildings up to time t and Y_t is the corresponding sum of claims due to damage to furniture and personal property. The experiment was conducted under the assumption that all the parameters of X_t and Y_t were equal.

We first estimated the marginals, assuming that the two processes were compound Poisson processes with equal intensities. Then we estimated a value for the Clayton Lévy copula and used some theorems from Bregman and Klüppelberg (2005) to try to make some inferences about the probability of ruin due to claims larger in size than the threshold τ.

7.4 Exploratory data analysis

The mean excess function is defined as $e(u) := \mathbb{E}(X - u|X > u)$. Let $X_1, \dots X_n$ be i.i.d. random variables with the distribution function F. Let F_n denote the empirical distribution function and let $\delta_n(u) = \{i : i = 1, \dots, n \quad X_i > u\}$. Then the *empirical mean excess function* $e_n(u)$ is defined as

$$e_n(u) = \frac{1}{\bar{F}_n(u)} \int_u^\infty \bar{F}_n(y)\,\mathrm{d}y = \frac{1}{\operatorname{card}\delta_n(u)} \sum_{i \in \delta_n(u)} (X_i - u)\,, u \geq 0,$$

with the convention that $0/0 = 0$. A mean excess plot consists of the graph

$$\left\{ \left(X_{k,n,e_n} \left(X_{k,n} \right) \right) : k = 1, \dots, n \right\}$$

The mean excess function for the generalized Pareto distribution is linear. Therefore if the generalized Pareto distribution fits the data well, then the mean excess plot should be approximately linear. Below are the mean excess plots for claims of the building category and the furniture category. To plot the mean excess functions we used the function “meplot” from the R package “evir”. The description of this function is given below:

```
"Description
An upward trend in plot shows heavy-tailed behaviour.
In particular, a straight line with positive gradient
 above some threshold is a sign of Pareto behaviour in tail.
 A downward trend shows thintailed behaviour whereas a line
with zero gradient shows an exponential tail. Because
upper plotting points are the average of a handful of
 extreme excesses, these may be omitted for a prettier plot."
```

The data seems to fit a line quite nicely for both categories except, as warned, the maximum observations of each. The generalized Pareto distribution has also been found to fit the univarate version of this dataset quite well in both Embrechts *et al.* (1999) and McNeil (1997).

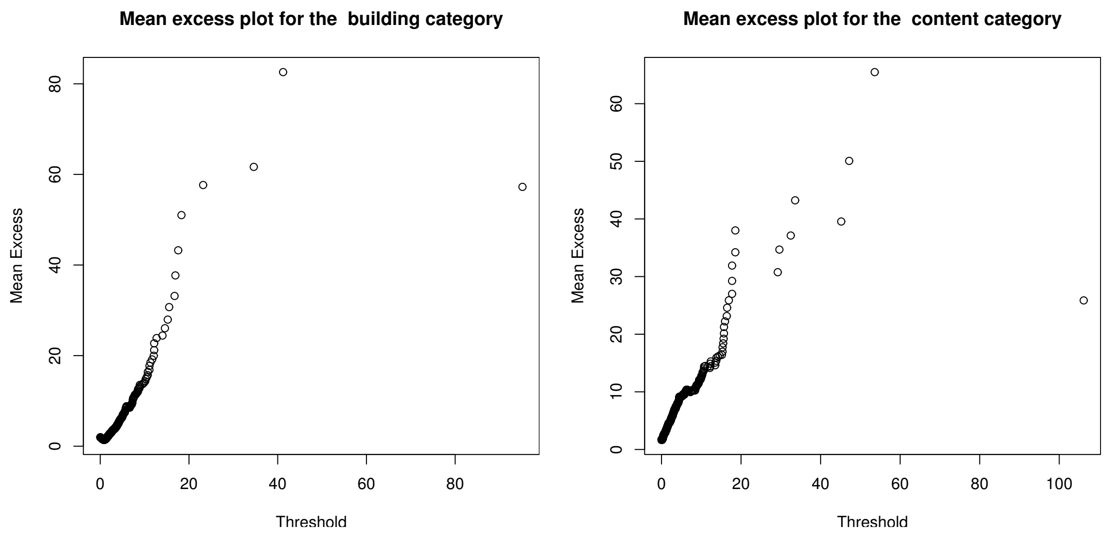

Figure 7.1: Mean excess plot of claims from the building category on the left and from the furniture category on the right. Both plots are fairly linear, which is a sign of Pareto behavior.

We also used the function “qplot” from “evir” to make QQ plots of the building category and the furniture category vs the standard exponential distribution. A description of the “qplot” function is given below:

```
"Details
If xi is zero the reference distribution is the exponential;
if xi is non-zero the reference  distribution is the
 generalized Pareto with that value of xi. In the case of
the exponential, the plot is interpreted as follows.
Concave departures from a straight line are a sign of
heavy-tailed behaviour. Convex departures show
 thin-tailed behaviour."
```

In a QQ plot against a standard exponential distribution we see that both categories display a tail behavior which is heavier than the exponential.

7.5 Some inference theory

Definition 7.5.1

A statistic $T(\mathbf{X})$ is a *sufficient statistic* for θ if the conditional distribution of the sample $\mathbf{X}$ given the value of $T(\mathbf{X})$ does not depend on θ.

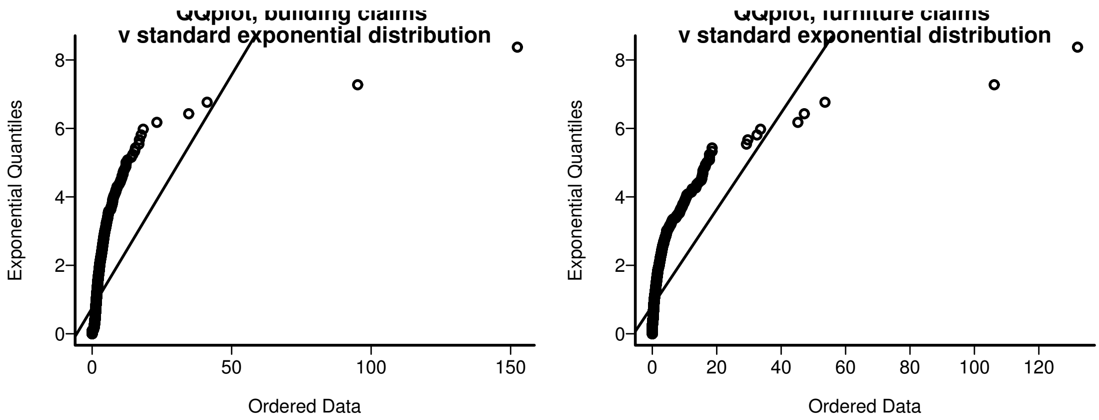

Figure 7.2: QQ plot of claims from the building category on the left and furniture category on the right vs a standard exponential distribution. Both plots show a concave deviation from a straight line, which is a sign of heavy-tailed behavior.

Definition 7.5.2

Let $\mathbf{X}$ be a random vector with a probability distribution belonging to a known family of probability distributions $\boldsymbol{F}_\theta$ parameterized by θ. Let $T(\mathbf{X})$ be any statistic based on $\mathbf{X}$. Then $T(\mathbf{X})$ is *complete* if for all measurable functions $g(T(\mathbf{X}))$, $\mathbb{E}\left(g(T(\mathbf{X}))\middle|\,\theta\right) = 0$ for all θ implies $\mathbb{P}\left(g(T(\mathbf{X})) = 0\,\middle|\,\theta\right) = 1$ for all θ. Equivalently, $T(\mathbf{X})$ is called a *complete statistic*.

Theorem 7.5.3: Factorization theorem

Let $f(\mathbf{X}|\theta)$ denote the joint distribution function or density of a sample $\mathbf{X}$. A statistic $T(\mathbf{X})$ is a sufficient statistic for θ if and only if there exist measurable functions $g(r|\theta)$ and $h(\mathbf{x})$ such that for all sample points $\mathbf{x}$ and all parameter points θ,

$$f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)\,h(\mathbf{x}).$$

A proof can be found on page 276 in Casella and Berger (2001).

Theorem 7.5.4

Let $X_1, \ldots, X_n$ be i.i.d. observations from an exponential family of distribution functions or densities of the form

$$f(x|\boldsymbol{\theta}) = h(x)c(\boldsymbol{\theta})\exp\left(\sum_{j=1}^{k} w_j(\boldsymbol{\theta})r_j(x)\right),$$

where $\boldsymbol{\theta} = (\theta_1, \ldots \theta_k)$. Then the statistic

$$T(\mathbf{X}) = \left(\sum_{i=1}^{n} r_1(X_i), \sum_{i=1}^{n} r_2(X_i), \ldots, \sum_{i=1}^{n} r_k(X_i)\right)$$

is complete if the range of $\{(w_1(\boldsymbol{\theta}), \ldots, w_k(\boldsymbol{\theta}))\}$ contains an open set in $\mathbb{R}^k$.
This result is stated on page 288 in Casella and Berger (2001).

Theorem 7.5.5

Let T be a complete sufficient statistic for a parameter θ and let $\phi(T)$ be any estimator based only on T. Then $\phi(T)$ is the unique minimum variance unbiased estimator(UMVUE) of its expected value. This theorem is stated on page 347 in Casella and Berger (2001).

7.5.1 Estimation of intensity/rate

In the compound Poisson model with intensity rate λ the waiting times between jumps obey the exponential distribution with probability density

$$f(w|\lambda) = \lambda e^{-\lambda w} 1_{w>0}.$$

This is readily seen to be a member of the exponential family.
Consider the case $k = 1$: Let $\mathbf{W} = W_1, \ldots, W_n$ be the observed waiting times. In theorem 7.5.4 let $r_1(W_i) = W_i$. We see from that theorem that
$T(\mathbf{W}) = \sum_{i=1}^{n} r_1(W_i)$ is a complete statistic for λ.

Consider the joint probability density of $\mathbf{W} = W_1, \ldots, W_n$. Since $W_1, \ldots, W_n$ are independent observations their joint probability density is
$f(\mathbf{w}|\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} w_i} 1_{\{\min(w_1, \ldots w_n) > 0\}}$. In theorem 7.5.3 on the facing page let
$g(T(\mathbf{W})|\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} w_i}$ and $h(\mathbf{w}) = 1_{\{\min(w_1, \ldots w_n) > 0\}}$. We then see that
$T(\mathbf{W}) = \sum_{i=1}^{n} r_1(W_i) = \sum_{i=1}^{n} W_i$ is also a sufficient statistic.

The calculation below shows that if Z is a Gamma-distributed random variable with parameters $\alpha > 1$ and β, then $\mathbb{E}\left(\frac{1}{Z}\right) = \frac{\beta}{\alpha-1}$.

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{Z}\right) &= \int_0^\infty \frac{1}{z}\frac{\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} e^{-\beta z}\,\mathrm{d}z \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha-1)}{\beta^{\alpha-1}} s \underbrace{\int_0^\infty \frac{\beta^{\alpha-1}}{\Gamma(\alpha-1)} z^{\alpha-1-1} e^{-\beta z}\,\mathrm{d}z}_{=1} \\
&= \frac{\beta}{\alpha-1}.
\end{aligned}$$

We also have that $2\beta Z$ is χ^2 distributed with 2α degrees of freedom.

Let us now consider the sum of n i.i.d exponentially distributed Poisson variables W_i, each with intensity λ. This sum is Gamma-distributed with $\alpha = n$ and $\beta = \lambda$. From the above we have that $\mathbb{E}\left(\frac{1}{\sum_{i=1}^n W_i}\right) = \frac{\lambda}{n-1}$. Since $\hat{\lambda} := \frac{n-1}{\sum_{i=1}^n W_i}$ is a statistic based only on $T(\mathbf{W}) = \sum_{i=1}^n W_i$, it follows from the sufficiency and completeness of $T(\mathbf{W})$, and theorem 7.5.5 on the previous page, that $\hat{\lambda}$ is a UMVUE for $\mathbb{E}(\hat{\lambda}) = \lambda$. Since $\sum_{i=1}^n W_i$ is Gamma-distributed with parameters $\alpha = n$ and $\beta = \lambda$ we have that $2\lambda \sum_{i=1}^n W_i$ is χ^2-distributed with $2n$ degrees of freedom. We can therefore construct the following $1-\varrho$ confidence interval for the parameter λ:

$$\left(\frac{\chi^2_{n,\frac{\varrho}{2}}}{2\sum_{i=1}^n W_i} < \lambda < \frac{\chi^2_{n,1-\frac{\varrho}{2}}}{2\sum_{i=1}^n W_i} \right).$$

Here $\chi^2_{n,\frac{\varrho}{2}}$ and $\chi^2_{n,1-\frac{\varrho}{2}}$ are the $\frac{\varrho}{2}$ and $1-\frac{\varrho}{2}$ quantiles of the χ^2 distribution with $2n$ degrees of freedom.

7.6 Estimation of shape and scale

7.6.1 Choice of estimation method

We initially considered four different estimators for the shape parameter ξ and the scaling parameter β: the maximum likelihood estimatator (MLE), the method of moment estimator (MOM), the probability-weighted moments (PWM) and the elemental percentile method (EPM).

The procedure for finding the MLE is based on solving the following equation for b (see Zhang (2007)):

$$n^{-1}\sum_{i=1}(1-bX_i)^{-1}-\left(1+n^{-1}\sum_{i=1}\log(1-bX_i)^{-1}\right)=0. \tag{7.7}$$

Numerical methods for solving this equation do not always converge. Furthermore, for ξ in the range $\left(-\frac{1}{2},\frac{1}{2}\right)$ it was shown in Hosking (1987) that the MLE does not display asymptotic efficiency even in samples as large as 500.

The method of moment estimators $\hat{\xi}_{\text{MOM}}$ and $\hat{\beta}_{\text{MOM}}$ is given by

$$\hat{\xi}_{\text{MOM}}=\left(\bar{x}^2/s^2-1\right)/2$$

and

$$\hat{\beta}=\bar{x}\left(\bar{x}^2/s^2+1\right)/2,$$

where $\bar{x}$ and s^2 are the sample mean and the sample variance. Both of these estimators involve the second moment of the distribution. However, the second moment of the generalized Pareto distribution does not exist if the shape parameter $\xi>0.5$. Intuitively this would make this method much less viable for $\xi>0.5$ and indeed, in the small-sample simulation study conducted in Castillo and Hadi (1997) this method was found to be severely biased for $\xi=1$ and $\xi=2$.

The probability-weighted moment method (PWM) was introduced in Greenwood *et al.* (1979) and is based on the quantities

$$M_{p,r,s}=\mathbb{E}\left[X^p\{F(X)\}^r\{(1-F(X))\}^s\right].$$

Let n be the number of observations and $i=1,2,\ldots,n$.
Let $\gamma\in(0,1)$ and $\delta>0$ be positive constants and $p_{i:n}=\frac{i-\gamma}{n+\delta}$.
The estimators $\hat{\xi}_{PWM}$ and $\hat{\beta}_{PWM}$ are then given by

$$\hat{\xi}_{PWM}=2-\frac{\bar{x}}{\bar{x}-2m}$$

and

$$\hat{\beta}_{PWM} = \frac{2\bar{x}}{\bar{x} - 2m},$$

where $m = n^{-1} \sum_{i=1}^{n} \left(1 - p_{i:n}\right) x_{i:n}$
(see Castillo and Hadi (1997).

This estimator also has its problems, since the PWM estimates do not exist for $\xi > 1$. In Zhang (2007) this estimator was found to be "nearly efficient if $\xi \approx \frac{1}{4}$ ", but behaving poorly otherwise.

The last estimator we considered was the elemental percentile method (EPM) described in Castillo and Hadi (1997). The algorithm for the elemental percentile method is somewhat elaborate and is given in the appendix.

In the simulation study conducted in Castillo and Hadi (1997) this estimator was found to outperform the PWM and the MOM for $\xi = 1$ and $\xi = 2$. Castillo and Hadi did not consider the maximum likelihood method.

In Embrechts *et al.* (1999) the generalized Pareto distribution (GPD) was fitted to a univariate version of the Danish fire insurance claims dataset. The estimate for ξ for excesses above 10 was 0.618 (see page 332 in Embrechts *et al.* (1999)). For the ruin probability the shape parameter is much more important than the scale parameter. Taking 0.618 as an initial guess we conducted a simulation test to see how each one of the methods described above performed. We simulated 50 variables from a GPD distribution with the shape parameter $\xi = 0.618$ and the scale parameter $\beta = 1$, and then estimated the shape parameter with each of the four methods. This was repeated 100,000 times for each estimator. We found that the MOM in particular, but also the PWM method had a substantial negative bias. Using a negatively biased estimator of ξ is particularly bad in ruin estimation since underestimation of the shape parameter ξ corresponds to underestimating the "heaviness" of the tail.

In our evalutation of the estimators we considered the bias and the root mean square error (RMSE), defined as $\sqrt{n^{-1} \sum_{i=1}^{n} \left(\xi - \hat{\xi}_i\right)^2}$, where $\hat{\xi}_i$ is the i-th estimate of ξ.

Although the root mean square error (RMSE) of the MOM and PWM estimators were on par with that of the other two, because of the substantial negative bias and the bias reported in studies like Castillo and Hadi (1997) we decided not to use them. In our test the maximum likelihood method and the elemental percentile

Method	MLE	MOM	PWM	EPM
Bias	-0.046	-0.297	-0.122	0.046
RMSE	0.245	0.316	0.217	0.357

Table 7.1: Bias and RMSE for the MLE, MOM, PWM and EPM estimators of the shape parameter

method had almost the same bias, but with different signs. The RSME was a little lower for the MLE than for the EPM. Moreover, it turned out that the Danish data set contained a number of repeated values, as seen in the table below, which made the EPM estimator not viable, so we settled for the maximum likelihood method.

Positions	building	contents
08/31/1985	5.0000000	0.850000
09/29/1985	5.0000000	1.200000
02/21/1988	3.1055901	4.880213
06/24/1988	0.8873115	4.880213
04/03/1989	0.0000000	5.927180
09/29/1989	3.3869602	5.927180
04/26/1989	1.6934801	8.467401
08/04/1989	5.0804403	8.467401
09/25/1989	0.0000000	8.467401
03/25/1988	7.0984916	17.746230
05/17/1988	6.0425910	17.746230
06/05/1988	6.6548359	17.746230

Table 7.2: Repeated values in Danish multivariate dataset

We did ask $McNeil$ why there were repeated values, but the response we got was that he himself had asked the provider of the data about this, but never got a reply.

Below are the histograms from the simulation experiment where we estimated the shape parameter with sample size 50. The simulated variables had $\xi = 0.618, \beta = 1$ and threshold $\tau = 10$. We repeated this experiment 100,000 times for each estimator.

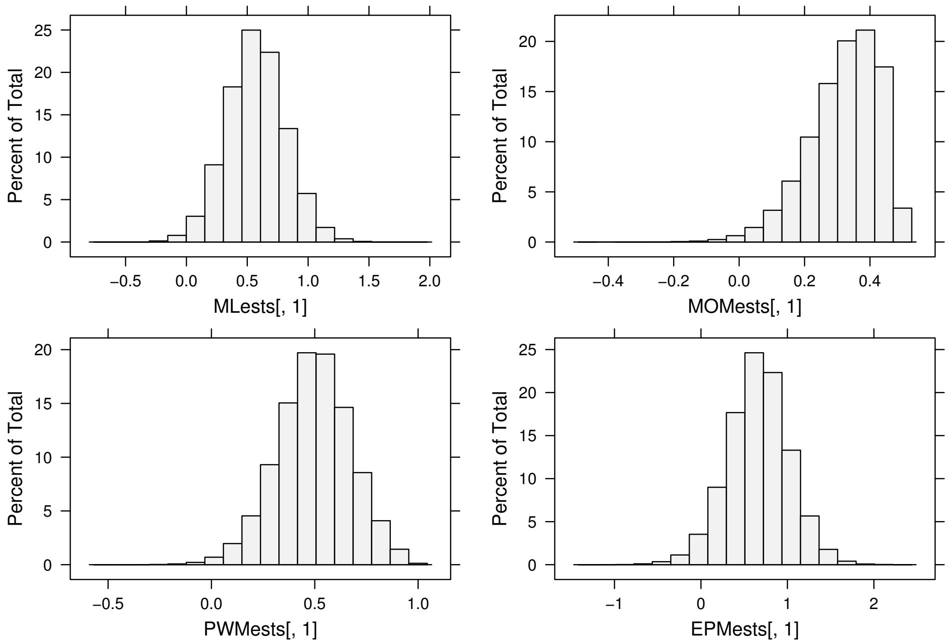

Figure 7.3: Histograms of the estimated value of the GPD shape parameters, from the simulation experiment: In the upper left hand corner:Maximum Likelihood estimations, upper right hand corner: Method of Moment (MOM) estimations, lower left hand corner: probability-weighted moment (PWM) estimations, lower right hand corner: Elemental Percentile Method estimations. From the histograms of the shape parameter we see that the median of both the MOM and the PWM are significantly below the real ξ-value 0.618. A calculation shows that the average of the MOM and PWM estimates are also substantially below 0.618.

7.6.2 Estimation

Using the maximum likelihood method as implemented in the evir R package we obtained estimates of the shape parameters for damage to building and damage to furniture. For 5 or fewer exceedances the MLE algorithm failed to converge in both cases. A description from `http://www.r-project.org/` of the function "gpd" in the "evir" package is given below:

```
Description
Returns an object of class "gpd" representing the fit of a
 generalized Pareto model to excesses over a high threshold.
Usage
gpd(data, threshold = NA, nextremes = NA, method = c("ml", "pwm"),
information = c("observed", "expected"), ...).
```

The choice of threshold is a trade off between bias (such a low threshold that the GPD-distribution is not appropriate) and uncertainty of the estimate (estimating the shape parameter based on a very small sample leads to a high degree of uncertainty of the estimate).

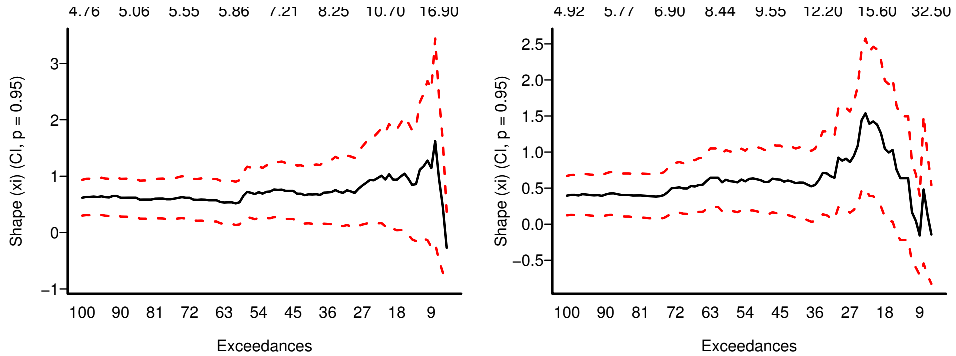

Figure 7.4: MLE Shape parameter plot for the building category on the left and damage to furniture on the right. The dotted lines are one standard error above and one standard error below the estimates. We see that for both the building category and the furniture category the shape estimates are reasonably stable for exceedances corresponding to thresholds less than 11.

In the plots of estimates of ξ for the building and furniture categories we see that for small up to medium high thresholds the shape parameter estimates are smaller than 1 for both categories. This was a necessary, but far from sufficient condition for equation 7.6 on page 81 to be applicable. For thresholds around 10 the estimates are fairly stable while we still have enough observations. Choosing 10 as threshold we obtain results as shown in the following table:

	Building	Furniture	Combined estimate
Exceedances	26	45	
Shape parameter ξ	0.93	0.62	0.73
Standard error for ξ	0.32	0.23	
Scale parameter β	3.39	5.41	4.68
Standard error for β	1.27	1.41	
Intensity λ	2.32	4.02	3.17
95% conf. int. for λ	$(1.58, 3.42)$	$(3.00, 5.40)$	
Clayton Lévy copula θ			0.36

Table 7.3: Estimates, standard errors and 95% confindence intervals, choosing threshold 10. The combined estimate of ξ is the weighted average $\frac{(26*0.93+45*0.62)}{26+45}$. The combined estimate of β is a similarly weighted average, while the combined estimate of λ is the plain average. The values for the intensities and their confidence intervals were obtained using a year as the time unit and using the estimator described in section 7.5.1 on page 85.

For the weighted averages of the shape and scale parameters, we see that with one exception the average is within one standard error of the individual esimtates. The exception is the scale parameter of the building category. That is 1.02 standard errors from the corresponding overall average. The assumption of common jump distribution does therefore not seem unreasonable.

The above values for the intensities and their confidence intervals were obtained using a year as the time unit and using the estimator described in section 7.5.1 on page 85. This suggests that the assumption of equal intensities is not unreasonable.

In the above table the value 0.36 for the Clayton Lévy copula was obtained using the other estimates and the algorithm described in section 6.4 on page 71 . We see that the average intensity is well within the 95% confidence intervals for both the building category and the furniture category. It is also just barely inside the 90% confidence intervals $(1.69, 3.23)$ for the the building category and $(3.16, 5.17)$ for the furniture category. This suggests that the assumption of equal intensities is not unreasonable.

In the above table the value 0.36 for the copula was obtained using the estimates of all the other parameters and using the estimator described in section 6.4 on page 71. This was an unexpectedly low estimate, but looking at the data we see that only 4 times in the time period 1980-1990 covered by the dataset one and the same fire led to damages over 10 million DKK in both the building category and the furniture category.

Recall that in theorem 7.2.5 on page 80 we had a constant $\bar{K}$ such that asymptotically the ruin probability $\Psi(x)$ as a function of the initial reserve x was:

$$\Psi(x) \sim \frac{\lambda}{c - 2\lambda\,\mathbb{E}\,Y} \frac{\left(\frac{\beta}{\xi}\right)^{1/\xi}}{\left(\frac{1}{\xi} - 1\right)} \left(\bar{K}\,(\xi,\theta) + 1\right) x^{-(1/\xi-1)}, \quad x \to \infty \tag{7.8}$$

Here c is the premium rate, assumed to be greater than $2\lambda\,\mathbb{E}\,Y$.

As one of the steps in the derivation of equation 7.8 (see equation A.3 on page 105) we have that, asymptotically,

$$L(w) := \frac{1}{\xi} \int_1^{w-1} \frac{y^{\frac{\theta}{\xi}-1}}{\left(y^{\frac{\theta}{\xi}} + (w-y)^{\frac{\theta}{\xi}}\right)^{\frac{\theta+1}{\theta}}}\, \mathrm{d}y \quad \sim \bar{K}(\xi,\theta) w^{\frac{-1}{\xi}}. \tag{7.9}$$

Here $\bar{K}$ is the same $\bar{K}$ as in equation 7.8.

We can find an approximate value for the constant $\bar{K}(\xi,\theta)$ by inserting the estimated ξ and θ, multiplying both sides of equation 7.9 by $w^{\frac{1}{\xi}}$ and plotting $w^{\frac{1}{\xi}} L(w)$ for large w, as in the graph below.

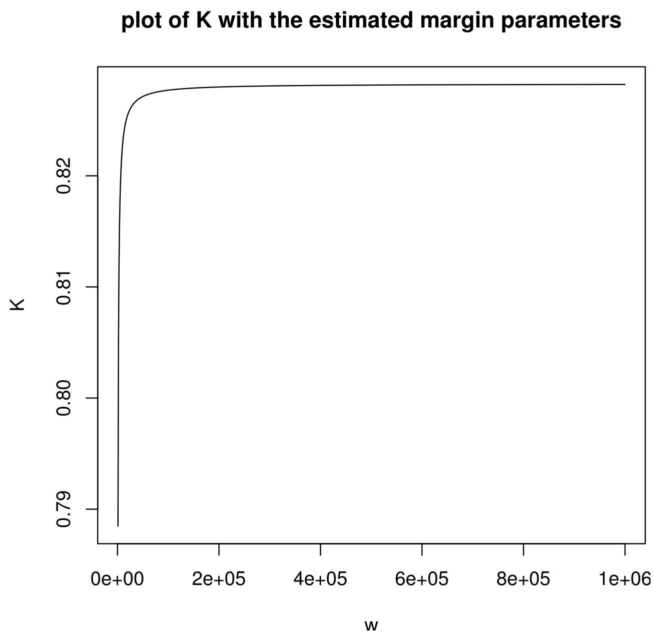

Figure 7.5: Plot of $w^{\frac{1}{\xi}}L(w)$ for increasing w with estimated ξ and θ. This plot can be used to find the value of $\bar{K}$, which is the asymptotic value of $w^{\frac{1}{\xi}}L(w)$ (see equation A.3 on page 105).

We see that $w^{\frac{1}{\xi}}L(w)$ converges to about 0.83.

ξ	β	λ	θ	$\bar{K}$
0.73	4.68	3.17	0.36	0.83

Table 7.4: Final estimates of the parameters

For $\xi \in (0,1)$ the expectation of the generalized Pareto distribution defined in equation 7.5 on page 81 can be calculated to be $\frac{\beta}{1-\xi}$.

Inserting all the estimated parameters as well as the value 0.83 for the $\bar{K}$ we finally obtain the asymptotic ruin probability

$$\Psi(x) \sim \frac{\lambda}{c - 2\lambda\frac{\beta}{1-\xi}} \frac{\left(\frac{\beta}{\xi}\right)^{1/\xi}}{\frac{1}{\xi} - 1} \left(\tilde{K}(\xi, \theta) + 1\right) x^{-(1/\xi - 1)},$$

$$= \frac{3.17}{c - 2 * 3.17 * \frac{4.68}{1-0.73}} \frac{\left(\frac{4.68}{0.73}\right)^{1/0.73}}{\frac{1}{0.73} - 1} (0.83 + 1) x^{-(1/0.73-1)},$$

$$\approx \frac{199.9}{c - 109.9} x^{-0.37} \quad (7.10)$$

for ruin due to the exceedances over 10 from either the building category or the furniture category.

Here x is the initial reserve and c is the premium rate, assumed to be larger than $2\lambda \mathbb{E} Y = 109.9$.

Now let us investigate the sensitivity of $\bar{K}$ and hence the ruin probability $\Psi(x)$ to changes in the copula parameter θ. Recall that $w^{\frac{1}{\xi}} L(w)$ converges to $\bar{K}$ as $w \to \infty$. Figure 7.6 below is a plot of $w^{\frac{1}{\xi}} L(w)$ vs θ for $\xi = 0.73$ and two very large values of w. We see that $\bar{K}$, and hence $\Psi(x)$, is quite sensitive to θ if θ is between 0 and 0.9. On the other hand $\bar{K}$ is quite close to one in the entire range one to ten. Thus except when the two risk processes are almost independent the ruin probability $\Psi(x)$ is quite robust for small changes in θ.

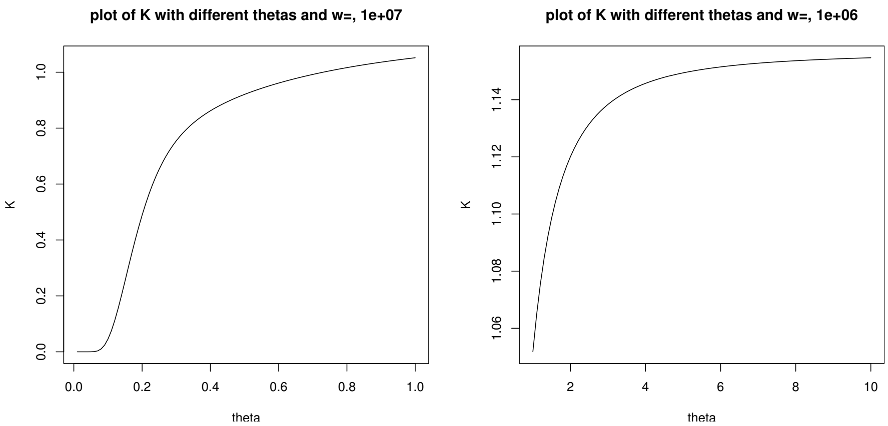

Figure 7.6: On the left, plot of $\bar{K}$ for $\theta \in (0, 1)$. On the right, plot of $\bar{K}$ for $\theta \in (1, 10)$

From these plots and the formula equation 7.8 on page 93 we conclude that with a medium strong or strong dependency the ruin probability is about twice as large as the ruin probability with indepencence or extremely weak dependence ($0 \leq \theta \leq 0.1$). In general the plots suggest that for ruin estimation using the model described in section 7.3 on page 81 the most crucial question is whether the risk processes are independent or not. Once dependence is established the ruin probability increases only slowly with the Clayton Lévy parameter θ.

7.6.3 Goodness of fit

To assess the goodness of fit we used the Anderson-Darling statistic A^2 and the Cramer-von Mises statistic W^2.

Both of these statistics are based on first finding the order statistics. Let n be the number of observations.

Let $x_{(1)},\ldots,x_{(n)}$ be the order statistics and let

$H(\xi,\tau,\beta) = 1-(1+\frac{\xi}{\beta}(x-\tau))^{-\frac{1}{\xi}}$ be the distribution function of the generalized Pareto distribution.

Then, letting $z_{(i)} = H\left(x_{(i)},\xi,\tau,\beta\right)$, the Anderson-Darling statistic A^2 is defined as follows:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\ln\left\{z_{(i)}\right\} + \ln\left\{1 - z_{n+1-i}\right\}\right]$$

and the Cramer-von Mises statistic W^2 is defined as

$$W^2 = \frac{1}{12n} + \sum_{i=1}^{n}\left\{z_{(i)} - \frac{(2i-1)}{2n}\right\}^2$$

(see Choulakian and Stephens (2001)).

Asymptotic critical values for A^2 and W^2 are found in the table below (taken from Choulakian and Stephens (2001)), where k corresponds to $-\xi$ in our parameterization.

Having chosen threshold 10 we have 26 observations from the building category and 45 observations from the furniture category.

The Monte Carlo simulation study conducted in Choulakian and Stephens (2001) indicated that the critical values have good accuracy for $n \geq 25$, so these statistics should be viable in our case. To find critical values at significance level 0.05 for $\xi = 0.73$ we used linear interpolation, as suggested in Choulakian and Stephens (2001), and obtained the critical values 0.805 for A^2 and 0.12 for W^2.

Table 2. Case 3: Both k and a Unknown: Upper-Tail Asymptotic Percentage Points for W^2 (normal type) and for A^2 (bold); p is $Pr(W^2 \geq z)$, or $Pr(A^2 \geq z)$, Where z Is the Table Entry

$k \backslash p$	*0.500*	*0.250*	*0.100*	*0.050*	*0.025*	*0.010*	*0.005*	*0.001*
−0.90	0.046	0.067	0.094	0.115	0.136	0.[illegible]65	0.187	0.239
−0.90	**0.339**	**0.471**	**0.641**	**0.771**	**0.905**	**1.086**	**1.226**	**1.559**
−0.50	0.049	0.072	0.101	0.124	0.147	0.179	0.204	0.264
−0.50	**0.356**	**0.499**	**0.685**	**0.830**	**0.978**	**1.180**	**1.336**	**1.707**
−0.20	0.053	0.078	0.111	0.137	0.164	0.200	0.228	0.294
−0.20	**0.376**	**0.534**	**0.741**	**0.903**	**1.069**	**1.296**	**1.471**	**1.893**
−0.10	0.055	0.081	0.116	0.144	0.172	0.210	0.240	0.310
−0.10	**0.386**	**0.550**	**0.766**	**0.935**	**1.110**	**1.348**	**1.532**	**1.966**
0.00	0.057	0.086	0.124	0.153	0.183	0.224	0.255	0.330
0.00	**0.397**	**0.569**	**0.796**	**0.974**	**1.158**	**1.409**	**1.603**	**2.064**
0.10	0.059	0.089	0.129	0.160	0.192	0.236	0.270	0.351
0.10	**0.410**	**0.591**	**0.831**	**1.020**	**1.215**	**1.481**	**1.687**	**2.176**
0.20	0.062	0.094	0.137	0.171	0.206	0.254	0.291	0.380
0.20	**0.426**	**0.617**	**0.873**	**1.074**	**1.283**	**1.567**	**1.788**	**2.314**
0.30	0.065	0.100	0.147	0.184	0.223	0.276	0.317	0.415
0.30	**0.445**	**0.649**	**0.924**	**1.140**	**1.365**	**1.672**	**1.909**	**2.475**
0.40	0.069	0.107	0.159	0.201	0.244	0.303	0.349	0.458
0.40	**0.468**	**0.688**	**0.985**	**1.221**	**1.465**	**1.799**	**2.058**	**2.674**
0.50	0.074	0.116	0.174	0.222	0.271	0.338	0.390	0.513
0.50	**0.496**	**0.735**	**1.061**	**1.321**	**1.590**	**1.958**	**2.243**	**2.922**

Table 7.5: Asymptotic critical values for A^2 and W^2

With threshold 10 and estimated values 0.73 for ξ and 4.68 for β, as found in table 7.3 on page 92, we obtain the following values for A^2 and W^2 for the building category and the furniture category.

Category	A^2	W^2
critical (maximum) value	0.805	0.12
building	0.453	0.166
furniture	0.718	0.233

Table 7.6: A^2 (Anderson-Darling) and W^2 (Cramer-von Mises) goodness of fit statistics for the building and furniture category

We find that the fit is good enough to pass the Anderson-Darling test, but unfortunately fails the Cramer-von Mises test.

8

Conclusion, final remarks and topics for future research

"Oft expectation fails, and most oft there Where most it promises; and oft it hits Where hope is coldest, and despair most fits."

From *"All's Well That Ends Well"* by William Shakespeare

8.1 Lévy copula

8.1.1 Motivation

In Cont and Tankov (2004) the following motivation is given for introducing Lévy copulas:

- "The construction allows one to choose any one-dimensional spectrally positive Lévy process for each of the components. In particular, it is possible to couple a compound Poisson process with a process which has almost surely an infinite number of jumps in every bounded interval."

- "The range of possible dependence structures includes complete dependence and independence with a smooth transition between these two extremes."
- "The dependence can be modeled in a parametric fashion."

8.1.2 Shortcomings of Lévy copula

As warned in Baüerle (2007), the Lévy copula is not sufficient to characterize all types of dependence. In particular the properties "conditionally increasing in sequence (CIS) " and "Multivariate total positivity of order 2" (MTP_2) can not be characterized. To see what this means we first need to introduce some concepts and a little notation.

Definition 8.1.1: MTP_2

Let $\mathbf{X}$ be a random vector $\mathbf{X} = (X_1, \ldots, X_n)$ on $\mathbb{R}^n$ with distribution $\mathbb{P}$ and a probability density f with respect to a σ-finite [1]product measure (if $\mathbb{P}$ is a continuous distribution) or a counting measure (if $\mathbb{P}$ is a discrete distribution).
Let $\mathbf{x} \vee \mathbf{y}$ signify $(\max\{x_1, y_1\}, \ldots, \max\{x_n, y_n\})$
and $\mathbf{x} \wedge \mathbf{y}$ signify $(\min\{x_1, y_1\}, \ldots, \min\{x_n, y_n\})$.
Then $\mathbf{X}$ (or f) is said to be MTP_2 if

$$f(\mathbf{x})f(\mathbf{y}) \leq f(\mathbf{x} \wedge \mathbf{y})f(\mathbf{x} \vee \mathbf{y}) \tag{8.1}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Definition 8.1.2: Usual stochastic order

We say that a function $f : \mathbb{R}^n \to \mathbb{R}$ is bounded and increasing if f is bounded and increasing in each component.

Let $\mathbf{X}$ and $\mathbf{Y}$ be random vectors on $\mathbb{R}^n$. $\mathbf{X}$ and $\mathbf{Y}$ are said to be comparable with respect to *usual stochastic order* (written $\mathbf{X} \leq_{st} \mathbf{Y}$) if
$\mathbb{E} f(\mathbf{X}) \leq \mathbb{E} f(\mathbf{Y})$ for all bounded increasing $f : \mathbb{R}^n \to \mathbb{R}$.

We say that $\mathbf{X}$ is *stochastically increasing* in $\mathbf{Y}$, denoted $\mathbf{X} \uparrow_{st} \mathbf{Y}$, if the conditional distribution of $\mathbf{X}$ given $\mathbf{Y} = \mathbf{y}$ is $\leq_{st}$-increasing in $\mathbf{y}$.

As a simple example suppose $\mathbf{X} = (X_1, X_2)$ is a random vector on $\mathbb{R}^2$ and that $-\infty < \tau_1 < \tau_2 < \infty$. Then $X_2 \uparrow_{st} X_1$ means that for all increasing and bounded functions f, $\mathbb{E}\left(f(X_2) \middle| X_1 = \tau_1\right) \leq E\left(f(X_2) \middle| X_1 = \tau_2\right)$.

Definition 8.1.3: Conditionally increasing in sequence

A random vector $\mathbf{X} = (X_1, \ldots X_n)$ is said to be *conditionally increasing in sequence* (CIS) if $X_i \uparrow_{st} (X_1, X_2 \ldots, X_{i-1})$ for all $i = 2, \ldots, n$.

[1] Let $(\Omega, \mathscr{F}, \rho)$ be a measure space. ρ is *σ-finite* if there exists a countable family of measurable subsets of Ω, $A_1, A_2, \ldots$, such that $\rho(A_k) < \infty$ for each $k = 1, 2, \ldots$.
ρ is a *counting measure* if $\rho(E) = \sum_{x \in E} 1$ for every $E \in \mathscr{F}$.

Example 8.1.4: CIS random vector

Let $\mathbf{X} = (X_1, X_2, X_3)$ be a random vector with the CIS property.
Let $x_1, x_2, x_3, y_1, y_2, y_3$ be real constants such that $x_1 < y_1$, $x_2 < y_2$ and $x_3 < y_3$. Then the CIS property gives that $X_3 \uparrow_{st} (X_1, X_2)$. Hence

$$\mathbb{E}\left(X_3 | X_1 = x_1, X_2 = x_2\right) \leq \mathbb{E}\left(X_3 | X_1 = y_1, X_2 = y_2\right).$$

It can be shown that Lévy copulas unfortunately characterize neither CIS nor MTP_2. That is to say, as in the example given in section A.2 on page 107, it is possible for a Lévy process $\mathbf{Y}_t$ with both the MTP_2 property and the CIS property to have the same Lévy copula as another Lévy process $\mathbf{X}_t$ which is neither MTP_2 nor CIS.

8.2 Stable processes

As pointed out in example 5.4 in Cont and Tankov (2004) it is possible to specify dependence between the components of a two-dimensional stable process $\mathbf{X}_t = \left(X_t^{(1)}, X_t^{(2)}\right)$ by means of the spherical measure (see theorem 3.4.2 on page 32). In this case the $\mathbf{X}_t$ will also be a stable process. Unfortunately the class of dependencies that can be modelled this way is quite small (see section 3.3.3 on page 30). Modelling the dependence with a Lévy copula, however, makes it possible to specify a much larger class of dependencies (although even this class is not exhaustive (see section A.2 on page 107)). Lévy copulas also allow model specification and simulation of multi-dimensional Lévy processes having stable components with different indexes.

8.3 Ruin probability

In chapter 7 we used a multivariate fire insurance dataset to estimate the parameters for the ruin probability of ruin caused by large insurance claims. The claims of this dataset are given in millions of danish kroner and adjusted for inflation.

We first estimated the parameters of the assumed common generalized Pareto claim size distribution. To decide which estimators to use we tried 4 different estimators on simulated observations from a generalized Pareto distribution. Since the method of moment estimator and the probability-weighted method showed a large bias and the Elemental Percentile method could not cope with the presence of repeated values in the dataset, we decided on using the maximum likelihood estimator. Choosing threshold 10 we obtained point estimates 0.73 for the shape parameter ξ and 4.68 for the scale parameter β. Choosing one year as the time unit we then obtained the point estimate 3.17 for the intensity λ. Using as input these estimated values as well as the dataset we obtained the estimate 0.36 for the Lévy copula. This was a surprisingly low estimate, since we expected large claims from the damage category and large values of the furniture category to be more strongly correlated. Having estimated all the parameters in equation 7.6 on page 81 we arrived at equation 7.10 on page 95, which give the ruin probability as a function of the premium rate. Experimenting with different values of the Lévy copula we found that the effect on the ruin probability from a shift from independence to weak dependence was much greater than that from a shift from weak dependence to medium or quite strong dependence (up to a value 10 for the Lévy copula).

Afterwards we looked at the goodness of fit of the generalized Pareto distribution using the Anderson-Darling statistic and the Cramer-von Mises statistic. The Anderson-Darling statistic suggested that the fitted model was adequate. The Cramer-von Mises statistic, however, unfortunately does not give support to the model. Indeed, a conservative minded person would certainly reject the ruin probability estimate of equation 7.6 on page 81 for the following reasons:

1. The model fitted poorly, as measured by the Cramer-von-Mises statistic, and

2. The estimated shape parameter ξ was too close to one rule out that $\xi > 1$ (which would violate the assumptions of theorem 7.2.5 on page 80).

If, on the other hand, one is willing to accept as much risk for underestimation as overestimation, we believe equation 7.6 on page 81 can be regarded as a fair estimate of the ruin probability.

8.4 Topics for future research

An obvious area of future research is to find estimators of the Lévy copula for more general Lévy processes, not just compound Poisson processes. The usefulness of Lévy copulas would no doubt be enhanced if more estimation algorithms were developed.

The requirement of equal parameters for the two risk processes, which is the basis of the ruin probability formulas given in chapter 7, seems very restrictive. A natural topic for further research is to see if the requirement of equal parameters can be loosened somewhat.

A
Proofs of some results

A.1 Proof of ruin probability theorem

In this section we give a proof of theorem 7.2.5 on page 80.

Proof: From proposition 7.2.2 on page 79 we have that

$$I_1'(z) = \int_0^z \left(\frac{\bar{F}^\theta(z-x)}{\bar{F}^\theta(z-x) + \bar{F}^\theta(x)} \right)^{\frac{\theta+1}{\theta}} F(\mathrm{d}x),$$

$$\bar{F}(x) = \begin{cases} \left(1 + \frac{\xi}{\beta}(x-\tau)\right)^{-1/\xi} & \text{if } x > \tau, \\ 1 & \text{otherwise, and} \end{cases}$$

$$f(x) = \frac{1}{\beta}\left(1 + \frac{\xi}{\beta}(x-\tau)\right)^{-(1+1/\xi)} 1_{\{x>\tau\}}.$$

If $0 < x < z - \tau$ and $z > 2\tau$ then

$$I_1'(z) = \int_{\tau}^{z-\tau} \left(\frac{\left[1+\frac{\xi}{\beta}(z-x-\tau)\right]^{-\theta/\xi}}{\left(1+\frac{\xi}{\beta}(z-x-\tau)\right)^{-\theta/\xi} + \left(1+\frac{\xi}{\beta}(x-\tau)\right)^{-\theta/\xi}} \right)^{\frac{\theta+1}{\theta}} \times \frac{1}{\beta}\left(1+\frac{\xi}{\beta}(x-\tau)\right)^{-(1+1/\xi)} \mathrm{d}x$$

$$+ \int_{z-\tau}^{z} \left[\frac{1}{1+\left(1+\frac{\xi}{\beta}(x-\tau)\right)^{-\theta/\xi}} \right]^{\frac{\theta+1}{\theta}} \frac{1}{\beta}\left(1+\frac{\xi}{\beta}(x-\tau)\right)^{-(1+1/\xi)} \mathrm{d}x$$

$$= \underbrace{\frac{1}{\beta}\int_{\tau}^{z-\tau} \frac{\left(1+\frac{\xi}{\beta}(x-\tau)\right)^{\frac{\theta}{\xi}-1}}{\left(\left(1+\frac{\xi}{\beta}(x-\tau)\right)^{\theta/\xi} + \left(1+\frac{\xi}{\beta}(z-x-\tau)\right)^{\theta/\xi}\right)^{\frac{\theta+1}{\theta}}} \mathrm{d}x}_{r(z)} \tag{A.1}$$

$$+ \underbrace{\int_{z-\tau}^{z} \frac{\frac{1}{\beta}\left(1+\frac{\xi}{\beta}(x-\tau)\right)^{-(1+1/\xi)}}{\left(1+\left(1+\frac{\xi}{\beta}(x-\tau)\right)^{-\theta/\xi}\right)^{\frac{\theta+1}{\theta}}} \mathrm{d}x}_{s(z)} .$$

We obviously have that

$$s(z) < \int_{z-\tau}^{z} \frac{1}{\beta}\left(1+\frac{\xi}{\beta}(x-\tau)\right)^{-(1+1/\xi)} \mathrm{d}x < \frac{\tau}{\beta}\left(1+\frac{\xi}{\beta}(z-\tau)\right)^{-(1+1/\xi)},$$

so that $s(z) = o\left(z^{-1/\xi}\right)$.

Now let $y := 1+\frac{\xi}{\beta}(x-\tau)$ and $w := 2+\frac{\xi}{\beta}(z-\tau)$.

Then $\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{\beta}{\xi}$ and

$$1+\frac{\xi}{\beta}(z-x-\tau) = 2+\frac{\xi}{\beta}(z-\tau) - y = w - y. \tag{A.2}$$

With a change of integral variable from x to y in equation A.1 we have that

$$r(z) = \frac{1}{\xi}\int_1^{1+\frac{\xi}{\beta}(z-2\tau)} \frac{y^{\frac{\theta}{\xi}-1}}{\left(y^{\frac{\theta}{\xi}} + \left(2+\frac{\xi}{\beta}(z-\tau)-y\right)^{\frac{\theta}{\xi}}\right)^{\frac{\theta+1}{\theta}}}\,\mathrm{d}y.$$

$$= \frac{1}{\xi}\int_1^{w-1} \frac{y^{\frac{\theta}{\xi}-1}}{\left(y^{\frac{\theta}{\xi}} + (w-y)^{\frac{\theta}{\xi}}\right)^{\frac{\theta+1}{\theta}}}\,\mathrm{d}y.$$

Let

$$L(w) := \frac{1}{\xi}\int_1^{w-1} \frac{y^{\frac{\theta}{\xi}-1}}{\left(y^{\frac{\theta}{\xi}} + (w-y)^{\frac{\theta}{\xi}}\right)^{\frac{\theta+1}{\theta}}}\,\mathrm{d}y. \qquad \text{(A.3)}$$

Then for any $t > 0$

$$L(tw) := \frac{1}{\xi}\int_1^{tw-1} \frac{y^{\frac{\theta}{\xi}-1}}{\left(y^{\frac{\theta}{\xi}} + (tw-y)^{\frac{\theta}{\xi}}\right)^{\frac{\theta+1}{\theta}}}\,\mathrm{d}y$$

$$= \frac{1}{\xi}t^{-1/\xi}\left\{\int_{1/t}^{1} + \int_1^{w-1} + \int_{w-1}^{w-1/t}\right\} \frac{x^{\frac{\theta}{\xi}-1}}{\left(x^{\frac{\theta}{\xi}} + (w-x)^{\frac{\theta}{\xi}}\right)^{\frac{\theta+1}{\theta}}}\,\mathrm{d}x$$

$$:= L_1(w) + L_2(w) + L_3(w).$$

We see that $L_2(w) = t^{-1/\xi}L(w)$.

Since

$$|L_1(w)| < \left|\frac{1}{\xi}t^{-1/\xi}\int_{1/t}^{1} \frac{1}{(w-1)^{\frac{\theta}{\xi}+\frac{1}{\xi}}}\,\mathrm{d}x\right|$$

$$< \left|\frac{1}{\xi}t^{-1/\xi}(1-1/t)(w-1)^{-(\frac{\theta+1}{\xi})}\right|$$

and

$$|L_3(w)| < \left| \frac{1}{\xi} t^{-1/\xi} \int_{w-1}^{w-1/\xi} x^{-1/\xi-1}\, \mathrm{d}x \right| < \left| \frac{1}{\xi} t^{-1/xi} (\frac{1}{\xi} - 1)(w-1)^{-\frac{1}{\xi}-1} \right|,$$

both $L_2(w)$ and $L_3(w)$ are $o(w^{-1/\xi})$.

Thus

$$L(tw) = t^{-1/\xi} \left(o \left(w^{-1/\xi} \right) + L(w) \right),$$

so asymptotically $L(w) \backsim \bar{K} w^{-1/\xi}$ for some $\bar{K}(\xi, \theta)$.

We thus have that

$$I_1'(z) \backsim L \left(2 + \frac{\xi}{\beta}(z - \tau) \right) \backsim \left(2 + \frac{\xi}{\beta}(z - \tau) \right)^{-1/\xi} \bar{K}(\xi, \theta)$$
$$\backsim \bar{K}(\xi, \theta) \left(\frac{\beta}{\xi} \right)^{1/\xi} z^{-1/\xi}.$$

Furthermore, from proposition 7.2.2 on page 79, we have that $I_2'(z) \sim \bar{F}(z)$ and $I_3'(z) \sim o(I_1(z))$. Since $\bar{F}(z) \sim \left(\frac{\xi}{\beta} \right)^{1/\xi}$ we get that asymptotically $U(z)^+ \sim \left(\frac{\xi}{\beta} \right)^{1/\xi} \left(\bar{K}(\xi, \theta) z^{-1/\xi} + 1 \right)$.

If $1/\xi > 1$ then by equation 7.1 on page 78 (see remark 2.12 in Bregman and Klüppelberg (2005)) we also have the following result for the ruin probability:

$$\Psi(x) \sim \frac{\lambda}{c - 2\lambda \mathbb{E} Y} \frac{\left(\frac{\beta}{\xi} \right)^{1/\xi}}{\frac{1}{\xi} - 1} \left(\bar{K}(\xi, \theta) + 1 \right) x^{-(1/\xi - 1)}, \quad x \to \infty. \qquad \text{(A.4)}$$

Here x is the initial reserve, ξ and β are the parameters of the generalized Pareto claim distribution, θ is the Clayton Lévy copula parameter and $\bar{K}$ is a constant depending on ξ and θ, but not on x.

A.2 Lévy copulas characterize neither MTP2 nor CIS

Example A.2.1: Different properties in spite of same Lévy copula

Let $\mathbf{X}_t = \left(X_t^{(1)}, X_t^{(2)}\right)$ be a 2-dimensional compound Poisson process with

$$\nu_{\mathbf{X}} = \frac{1}{3}t\left(\delta_{(1,0)} + \delta_{(2,1)} + \delta_{(3,3)}\right). \tag{A.5}$$

Let $\mathbf{Y}_t = \left(Y_t^{(1)}, Y_t^{(2)}\right)$ be a 2-dimensional compound Poisson process with

$$\nu_{\mathbf{Y}} = \frac{1}{3}t\left(\delta_{(1,1)} + \delta_{(2,2)} + \delta_{(3,3)}\right). \tag{A.6}$$

Note that the jumps of $X^{(1)}$ are completely dependent on the jumps of $X^{(2)}$ and that the jumps of $Y^{(1)}$ are completely dependent on the jumps of $Y^{(2)}$.

It can be shown (see proposition 5.4 in Cont and Tankov (2004)) that if the jumps of two-dimensional Lévy process are positive and completely dependent then $F(x_1, x_2) := \min(x_1, x_2)$ is a possible Lévy copula. We therefore have that $F(x_1, x_2) := \min(x_1, x_2)$ is a possible Lévy copula for both $\mathbf{X}_t$ and $\mathbf{Y}_t$.

We now go on to show that $\mathbf{Y}_t$ has the MTP_2 and the CIS property, while $\mathbf{X}_t$ is neither MTP_2 nor CIS.

It follows from equation A.6 that jumps of the two components of $\mathbf{Y}_t$ are of the same size a.s.

We hence have that $\mathbb{P}\left(Y_t^{(1)} = u_1, Y_t^{(2)} = u_2\right)\mathbb{P}\left(Y_t^{(1)} = v_1, Y_t^{(2)} = v_2\right) > 0$ implies that $u_1 = u_2$ and $v_1 = v_2$.

With $u_1 = u_2$ and $v_1 = v_2$ we have that

$$\begin{aligned}
&\mathbb{P}\left(Y_t^{(1)} = \min(u_1, v_1), rvY_t^{(2)} = \min(u_2, v_2)\right)\mathbb{P}\left(Y_t^{(1)} = \max(u_1, v_1), Y_t^{(2)} = \max(u_2, v_2)\right) \\
&= \mathbb{P}\left(Y_t^{(1)} = \min(u_1, v_1), Y_t^{(2)} = \min(u_1, v_1)\right)\mathbb{P}\left(Y_t^{(1)} = \max(u_1, v_1), Y_t^{(2)} = \max(u_1, v_1)\right) \\
&= \mathbb{P}\left(Y_t^{(1)} = u_1, Y_t^{(2)} = u_1\right)\mathbb{P}\left(Y_t^{(1)} = v_1, Y_t^{(2)} = v_1\right) \\
&= \mathbb{P}\left(Y_t^{(1)} = u_1, Y_t^{(2)} = u_2\right)\mathbb{P}\left(Y_t^{(1)} = v_1, Y_t^{(2)} = v_2\right).
\end{aligned}$$

Hence $\mathbf{Y}_t$ satisfies equation 8.1 on page 100 and thus has the MTP_2 property.

We also have that, for every bounded and increasing function $f:\mathbb{R}\to\mathbb{R}$, $\mathbb{E}\left(f(Y_t^{(2)}|Y_t^{(1)}=y\right)=y$, so clearly $\mathbf{Y}_t$ also has the CIS property.

We will now show that $\mathbf{X}_t$ is neither MTP_2 nor CIS,
Let $u_1=3, u_2=3, v_1=4$ $v_2=2$ and $t>0$.
We then have that
$\mathbb{P}\left(X_t^{(1)}=3, X_t^{(2)}=3\right)\mathbb{P}\left(X_t^{(1)}=4, X_t^{(2)}=2\right)>0.$
On the other hand

$$
\begin{aligned}
&\mathbb{P}\left(X_t^{(1)}=\min(3,4), X_t^{(2)}=\min(3,2)\right)\mathbb{P}\left(X_t^{(1)}=\max(3,4), X_t^{(2)}=\max(3,2)\right)\\
&=\mathbb{P}\left(X_t^{(1)}=3, X_t^{(2)}=2\right)\mathbb{P}\left(X_t^{(1)}=4, X_t^{(2)}=3\right)\\
&=0*\mathbb{P}\left(X_t^{(1)}=4, X_t^{(2)}=3\right)=0,
\end{aligned}
$$

so $\mathbf{X}_t$ does not have the MTP_2 property.
Now fix $t=1$. From equation A.5 on the previous page and definition 4.1.5 on page 36 wee see that $\mathbf{X}_t$ has the intensity $\lambda=1$.
Let N be the number of jumps occurring before or at $t=1$.
We have that

$$
\begin{aligned}
&\mathbb{P}\left(X_1^{(1)}=3\right)\\
&=\mathbb{P}\left(X_1^{(1)}=3|N=1\right)\mathbb{P}(N=1) \qquad +\mathbb{P}\left(X_1^{(1)}=3|N=2\right)\mathbb{P}(N=2)\\
&+\mathbb{P}\left(X_1^{(1)}=3|N=3\right)\mathbb{P}(N=3)\\
&=\frac{1}{3}\left(e^{-1}\right)+2\left(\frac{1}{3}\right)^2\frac{e^{-1}}{2!}+\left(\frac{1}{3}\right)^2\frac{e^{-1}}{3!}\\
&=e^{-1}\left\{\frac{1}{3}+\left(\frac{1}{3}\right)^2+\frac{\left(\frac{1}{3}\right)^2}{3!}\right\}
\end{aligned}
\tag{A.7}
$$

and

$$
\begin{aligned}
&\mathbb{P}\left(X_1^{(1)}=4\right)\\
&=\mathbb{P}\left(X_1^{(1)}=4|N=2\right)\mathbb{P}(N=2)+\mathbb{P}\left(X_1^{(1)}=4|N=3\right)\mathbb{P}(N=3)\\
&+\mathbb{P}\left(X_1^{(1)}=4|N=4\right)\mathbb{P}(N=4)\\
&=\left(\left(\frac{1}{3}\right)^2+2\left(\frac{1}{3}\right)^2\right)\frac{e^{-1}}{2!}+3\left(\frac{1}{3}\right)^3\frac{e^{-1}}{3!}+\left(\frac{1}{3}\right)^4\frac{e^{-1}}{4!}\\
&=e^{-1}\left\{\frac{1}{3}+\frac{1}{54}+\frac{1}{3^4}\frac{1}{3!}\right\}.
\end{aligned}
\tag{A.8}
$$

Let $f(x) := 1_{x\geq 3}$.
Making use of equations A.7 to A.8 on pages 108–109 we have that

$$
\begin{aligned}
&\mathbb{E}\left(f\left(X_1^{(2)}\right)X_1^{(1)}=3\right)\\
&=\mathbb{E}\left(1_{X_1^{(2)}}|X_1^{(2)}=3\right)\\
&=\mathbb{P}\left(X_1^{(2)}\geq 3|X_1^{(1)}=3\right)\\
&=\frac{\mathbb{P}\left(X_1^{(1)}=3,X_1^{(2)}=3\right)}{\mathbb{P}\left(X_1^{(1)}=3\right)}\\
&=\frac{\mathbb{P}\left(X_1^{(1)}=3,X_1^{(2)}=3|N=1\right)\mathbb{P}(N=1)}{\mathbb{P}\left(X_1^{(1)}=3\right)}\\
&=\frac{\frac{1}{3}e^{-1}}{=e^{-1}\left\{\frac{1}{3}+\left(\frac{1}{3}\right)^2+\frac{\left(\frac{1}{3}\right)^2}{3!}\right\}}=\frac{18}{25}=0.72.
\end{aligned}
$$

We also have that

$$
\begin{aligned}
&\mathbb{E}\left(f\left(X_1^{(2)}\right)|X_1^{(1)}=4\right)\\
&=\mathbb{E}\left(1_{X_1^{(2)}}|X_1^{(2)}=4\right)\\
&=\mathbb{P}\left(X_1^{(2)}\geq 3|X_1^{(1)}=4\right)\\
&=\frac{\mathbb{P}\left(X_1^{(1)}=3,X_1^{(2)}=4\right)}{\mathbb{P}\left(X_1^{(1)}=4\right)}\\
&=\frac{\mathbb{P}\left(X_1^{(1)}=3,X_1^{(2)}=3|N=2\right)\mathbb{P}(N=2)}{\mathbb{P}\left(X_1^{(1)}=4\right)}\\
&=\frac{2\left(\frac{1}{3}\right)^2 e^{-1}}{e^{-1}\left\{\frac{1}{3}+\frac{1}{54}+\frac{1}{3^4}\frac{1}{3!}.\right\}}=\frac{27}{43}\approx 0.63.
\end{aligned}
$$

Since $\mathbb{E}\left(X_1^{(2)}|X_1^{(1)}=3\right)>\mathbb{E}\left(X_1^{(2)}|X_1^{(1)}=4\right)$
$\mathbf{X}_1$ does not have the CIS property.

B
Algorithm of the Elemental Percentile Metod

Taken from Castillo and Hadi (1997).

In Castillo and Hadi (1997) the distribution function of the generalized Pareto distribution with shape parameter $\xi \neq 0$ is written as

$$F(x) = 1 - \left(1 - \frac{x}{\delta}\right)^{1/k},$$

where k corresponds to $-\xi$ in our parameterization in chapter 7.

Let $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ be the ordered observations.

Let $p_{i:n} = \frac{i-\gamma}{n+\eta}$

and $C_i = \ln\left(1 - p_{i:n}\right)$.

Here $\gamma \in (0,1)$ and $\eta > 0$ are positive constants.

In the simulation study conducted in Castillo and Hadi (1997), setting $\gamma = 0$ and $\eta = 1$ were found to give the best results.

Let

$$\ln\left(1-\frac{x_{i:n}}{\delta}\right)=kC_i \tag{B.1}$$

and

$$\ln\left(1-\frac{x_{j:n}}{\delta}\right)=kC_j$$

be a system of two equations and two unknowns.
Eliminating k we obtain

$$C_i\ln\left(1-\frac{x_{j:n}}{\delta}\right)=C_j\ln\left(1-\frac{x_{i:n}}{\delta}\right). \tag{B.2}$$

Eliminating δ we obtain

$$x_{i:n}\left[1-\left(1-p_{j:n}\right)^k\right]=x_{j:n}\left[1-(1-p_{i:n})^k\right]. \tag{B.3}$$

Note that each of equations B.2 and B.3 are functions of only one unknown variable, and as shown in Castillo and Hadi (1997), can be solved numerically with a bisection method. After solving equations B.2 to B.3 we obtain a corresponding estimator $\hat{k}(i,h)$ for k given by

$$\hat{k}(i,j)=\frac{\ln\left(\frac{1-x_{i:n}}{\hat{\delta}(i,j)}\right)}{C_i}. \tag{B.4}$$

An estimator of β can then be computed as

$$\hat{\beta}(i,j)=\hat{k}(i,j)\hat{\delta}(i,j). \tag{B.5}$$

Algorithm 1

1. Select any two distinct order statistics $x_{i:n}<x_{j:n}$, and compute C_i and C_j. Let $d=C_jx_{i:n}-C_ix_{j:n}$.

2. If $d = 0$, then let $\hat{\delta}(i, j) = \pm\infty$ and $\hat{k}(i, j) = 0$.

3. Compute $\delta_0 = x_{i:n} x_{j:n} \left(C_j - C_i \right) / d$. If $\delta_0 > 0$, then $\delta_0 > x_{j:n}$ (a proof can be found in Castillo and Hadi (1997)). Thus use the bisection method on the interval $[x_{j:n}, \delta_0]$ to obtain a solution $\hat{\delta}(i, j)$ of equation B.2 on the facing page and go to Step 5; if $\delta_0 < 0$ go to Step 4.

4. Use the bisection method on the interval $[\delta_0, 0]$ to solve equation B.2 on the preceding page and obtain $\hat{\delta}(i, j)$.

5. Use $\hat{\delta}(i, j)$ to compute $\hat{k}(i, j)$ and $\hat{\beta}(i, j)$ by means of equation B.5 on the facing page.

Algorithm 2

1. Use Algorithm 1 to compute $\hat{k}(i, j)$ and $\hat{\beta}(i, j)$ for all distinct pairs $x_{i:n} < x_{j:n}$.

2. Use the median of each of the foregoing sets of estimators to obtain corresponding overall estimators of k and β; that is

$$\hat{k}_{EPM} = median \left(\hat{k}(1,2), \hat{k}(1,3), \ldots, \hat{k}(n-1,n) \right)$$

and

$$\hat{\beta}_{EPM} = median \left(\hat{\beta}(1,2), \hat{\beta}(1,3), \ldots, \hat{\beta}(n-1,n) \right).$$

Bibliography

Bachelier L. (1900). "Theorie de la spéculation". *Annales de l'Ecole Normale Supérieure*, pages 21–86.
URL: http://www.numdam.org/item?id=ASENS_1900_3_17__21_0. Cited on page 9.

Baüerle N. (2007). "Dependence properties and construction of multivariate stochastic processes". *Technical report*, Oberwolfach Matematisches Forschungsinstitut.
URL: http://www.mathematik.uni-kl.de/~korn/OW/Nicole_OW.pdf. Cited on page 100.

Bertoin J. (1996). *Lévy processes*. Cambridge University Press. Cited on pages 35, 37, and 41.

Billingsley P. (1986). *Probability and Measure*. Wiley, New York, second edition. Cited on page 17.

Bingham N., Goldie C., and Teugels J. (1987). *Regular Variaton*. Cambridge University Press. Cited on page 77.

Black F. and Scholes M. (1973). "The pricing of options and corporate liabilities." *Journal of Political Economy*, **volume 81**, pages 637–654.
URL: http://www.jstor.org/view/00223808/di950954/95p0059k/0. Cited on page 9.

Bregman Y. and Klüppelberg C. (2005). "Ruin estimation in multivariate models with Clayton dependence structure". *Scandinavian Actuarial Journal*, , no. 6, pages 462–480. URL: http://www.mathematik.uni-muenchen.de/~bregman/copula051009.pdf. Cited on pages 8, 76, 77, 78, 79, 80, 82, and 106.

Breiman L. (1968). *Probability*. Addison-Wesley Publishing Company. Cited on pages 23, 25, and 35.

Casella G. and Berger R.L. (2001). *Statistical Inference*. Duxbury Advanced Series. Duxbury. Cited on pages 84 and 85.

Castillo E. and Hadi A.S. (1997). "Fitting the generalized pareto distribution to data". *Journal of the American Statistical Association*, **volume 92**, no. 440, pages 1609–1620.
URL: `http://www.jstor.org/view/01621459/di986014/98p0323m/0`. Cited on pages 87, 88, 111, 112, and 113.

Choulakian V. and Stephens M. (2001). "Goodness of fit tests for the generalized pareto distribution". *Technometrics*, **volume 43**, no. 4.
URL: `http://www.ingentaconnect.com/content/asa/tech/2001/00000043/00000004/art00011`. Cited on page 96.

Cont R. and Tankov P. (2004). *Financial Modelling with Jump Processes*. Chapman & Hall. Cited on pages 8, 9, 10, 22, 32, 38, 41, 42, 43, 45, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 67, 69, 99, 101, and 107.

Embrechts P., Klüppelberg C., and Mikosch T. (1999). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag. Cited on pages 65, 77, 82, and 88.

Feller W. (1971). *An introduction to Probability Theory and its Applications*, volume II. John Wiley & Sons, second edition. Cited on page 27.

Folland G.B. (1999). *Real Analysis*. John Wiley & Sons, inc., second edition. Cited on page 15.

Greenwood J., Landwehr J.M., Matalas N.C., and Wallis J.R. (1979). "Probability weighted moments: Definition and relation to parameters of several distributions". *Water Resources Research*, **volume 15**, pages 1049–1054. Cited on page 87.

Hosking J. (1987). "Parameter and quantile estimation for the generalized pareto distribution". *Technometrics*, **volume 29**, no. 3, pages 339–349.
URL: `http://www.jstor.org/view/00401706/ap040116/04a00150/0`. Cited on page 87.

Jacod J. and Protter P. (2004). *Probability Essentials*. Springer-Verlag Berlin, Heidelberg, second edition. Cited on pages 15 and 16.

McNeil A.J. (1997). "Estimating the tails of loss severity distributions using extreme value theory". *ASTIN Bulletin*, , no. 27, pages 117–132.
URL: `http://www.ma.hw.ac.uk/~mcneil/data.html`. Available at. Cited on pages 81 and 82.

Nelsen R.B. (1998). *An Introduction to Copulas*. Springer-Verlag. Cited on pages 44, 48, and 49.

Packham N. (2006). "Multidimensional models with jumps." URL: http://www.frankfurt-school.de/content/en/research/quantitative_Finance/levy. Seminar on Lévy processes in Finance, HfB Business School of Finance and Management. Cited on page 45.

Rudin W. (1976). *Principles of Mathematical Analysis*. McGraw-Hill, third edition edition. Cited on page 73.

Samorodnitsky G. and Taqqu M.S. (1994). *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall. Cited on pages 26, 27, 30, and 32.

Samuelson P. (1965). "Rational theory of warrant pricing". *Industrial Management Review*, **volume 6**, pages 13–32. Cited on page 9.

Sato K.I. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press. Cited on pages 12, 14, 16, 17, 20, 21, 23, 24, 25, 26, 27, 28, 29, 32, 37, 38, and 40.

Schoutens W. (2003). *Lévy Processes in Finance*. John Wiley & Sons. Cited on page 9.

Tankov P. (2003a). "Dependence structure of Lévy processes with applications in risk management". URL: http://www.cmap.polytechnique.fr/preprint/comment.php?showdetails=1\&paper=502. Raport interne 502. Preprint available at. Cited on page 49.

Tankov P. (2003b). "Dependence Structure of spectrally positive multi-dimensional Lévy processes". URL: http://www.math.jussieu.fr/~tankov. Manuscript, 25 pages. Cited on page 8.

Uchaikin V.V. and Zolotarev V.M. (1999). *Chance and Stability Stable Distributions and their Applications*. VSP BV. Cited on pages 10 and 11.

Woyczyńscki W.A. (2001). *In "Lévy processes Theory and Applications"*, chapter : "Lévy processes in the Physical Sciences", pages 242–263. Birkhäuser. Cited on pages 10 and 11.

Zhang J. (2007). "Likelihood moment estimation for the generalized pareto distribution". *Australian & New Zealand Journal of Statistics*, , no. 49, pages 69–77. DOI: 10.1111/j.1467-842X.2006.00464.x. URL: http://www.

blackwell-synergy.com/doi/abs/10.1111/j.1467-842X.2006.00464.x.
DOI:. Cited on pages 87 and 88.