Technische Universität München
Lehrstuhl für Mathematische Statistik

Parameter estimation of Multivariate Lévy processes

Habib Esmaeili

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. Matthias Scherer
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Claudia Klüppelberg
	2.	Univ.-Prof. Dr. Alexander M. Lindner Technische Universität Braunschweig
	3.	Univ.-Prof. Dr. Friedrich Hubalek Technische Universität Wien/ Österreich

Die Dissertation wurde am 27.01.2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 10.05.2011 angenommen.

Zusammenfassung

In dieser Arbeit benutzen wir Lévy-Copulas, um die Abhängigkeitsstruktur multivariater Lévy-Prozesse zu beschreiben und konstruieren mehrere Modelle, die auf Lévy-Copulas basieren. Die Parameterschätzung dieser Modelle ist der Hauptteil dieser Arbeit. Das Schätzverfahren basiert auf dem Maximum-Likelihood-Prinzip.

Für zusammengesetzte Poisson-Prozesse, die endliches Lévymaß haben, zerlegen wir die Träger der Maße in den Teil auf den Achsen und den Teil außerhalb der Achsen. Für einen bivariaten zusammengesetzten Poisson-Prozess erzeugt diese Zerlegung drei unabhängige Komponenten; zwei zeigen nur die Sprünge in einer Komponente, der dritte Teil betrachtet die bivariaten Sprünge in beiden Komponenten. Die Likelihood-Funktion kann mit Hilfe dieser unabhängigen Teile hergeleitet werden. Wir stellen überdies einen neuen Simulationsalgorithmus für einen bivariaten zusammengesetzten Poisson-Prozess vor. Wir wenden unsere Methode an, um Schadendaten einer dänischen Feuerversicherungen zu modellieren und die Parameter zu schätzen.

Die Erweiterung dieser Methode für Lévy-Prozesse mit unendlichem Lévymaß wird im zweiten Teil dieser Arbeit diskutiert. Genauer gesagt betrachten wir einen bivariaten stabilen Lévy-Prozess und schneiden alle kleinen Sprünge ab. Die statistische Analyse basiert nun auf dem daraus hervorgegangenen zusammengesetzten Poisson-Prozess. Die Fisher-Informationsmatrix wird ebenfalls analytisch berechnet und die asymptotische Normalität der Schätzer bewiesen, wenn die Anzahl der Sprünge gegen unendlich strebt. In diesem Modell kann dies geschehen entweder, wenn die Beobachtungsperiode unendlich groß wird,

oder wenn der Stutzungspunkt der kleinen Sprünge gegen 0 geht. Eine Simulationsstudie untersucht den Effizienzverlust durch das Abschneiden der kleine Sprünge.

Schließlich wird im letzten Kapitel ein neue Schätzmethode eingeführt. Die Hauptidee dieses Ansatzes, der zwei-Schritt-Methode, ist ähnlich zur IFM (inference functions for margins) für multivariate Verteilungsfunktionen. Wir schätzen die Parameter der Randprozesse zunächst getrennt. Gegeben die Schätzwerte aus dem ersten Schritt, transformieren wir die Randprozesse und schätzen im zweiten Schritt die Parameter der Abhängigkeitsstruktur. Die Godambe-Informationsmatrix wird ebenfalls analytisch berechnet und die asymptotische Normalität der Schätzer bewiesen, wenn die Anzahl der Sprünge gegen unendlich strebt. Eine Simulationsstudie vergleicht die Effizienz der vorgestellten drei Methoden.

Abstract

In this thesis, we apply Lévy copulas to describe the dependence structure of multivariate Lévy processes and build some Lévy copula-based models. Parameter estimation of the models is the main part of this work. The estimation procedure is based on maximum likelihood principles.

For compound Poisson processes (CPP) which have finite Lévy measure, we decompose the mass on the axes and outside of the axes. This decomposition for a bivariate CPP generates three independent components and shows either the jumps only in one component, or the bivariate jumps in both components. The likelihood function can be derived based on these independent parts. We also suggest a new simulation algorithm for a bivariate CPP. We apply our method to model Danish fire insurance data and estimate the parameters of the model.

The extension of the method for Lévy processes with infinite Lévy measure is discussed in the second part. More precisely we take a bivariate stable Lévy Process and truncate all the small jumps. We base the statistical analysis on the resulting CPP. The Fisher information matrix is also calculated and the asymptotic normality of the estimators is proved as the number of jumps tends to infinity. In this model this may happen either for the observation period going to infinity, or the truncation point going to 0 for a fixed observation period. A simulation study investigates the loss of efficiency because of the truncation.

Finally, a new estimation procedure is introduced in the last chapter. The main idea of this approach, which we call two-step method, is similar to IFM (inference functions for margins) for multivariate distribution functions. First, the parameters of the marginal processes are estimated. Then, given the estimates from the first step, we estimate in a second step only the dependence structure parameters. This method is applied to a bivariate α-stable Clayton subordinator with different or common marginal parameters. For the latter, the Godambe information matrix and asymptotic covariance matrix are analytically calculated. Moreover, the asymptotic normality of the estimators is proved as the time span goes to infinity or the truncation point goes to zero. A simulation study compares the quality of all three estimation methods: the two-step estimates, the MLEs of a full model and the MLEs based on joint jumps only.

Acknowledgments

First of all, I would like to express my sincere gratitude to Prof. Dr. Claudia Klüppelberg for introducing this topic to me and for her excellent guidance and valuable support during our discussions in frequent meetings in Garching.

Special thanks go to my colleagues at the Chair of Mathematical Statistics for their support, collaboration and fruitful suggestions.

I spent six months of my research period at the Wolfgang Pauli institute (WPI) at the University of Vienna and in the research unit of financial and actuarial mathematics (FAM) at the Vienna University of Technology. Their hospitality and financial support are cordially acknowledged.

Last but not least, I would like to thank my family for everything especially their encouragement and, of course, my wife for her understanding and patience during the period of this work.

Contents

Chapter 1

Introduction

Before giving an outline of the thesis, we review some topics that are common in the following chapters. We start with the definition of a Lévy process and present the dependence structure in a Lévy process in terms of the notion of a Lévy copula. For general treatment of Lévy processes we refer to Applebaum [1], Bertoin [7], Sato [32] and Cont and Tankov [13].

1.1 Lévy processes

A càdlàg stochastic process $\mathbf{S} = (\mathbf{S}(t))_{t\geq 0}$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ with values on $\mathbb{R}^d$, for $d \in \mathbb{N}$ is called a *Lévy process*[1] if and only if

(i) $\mathbf{S}(0) = \mathbf{0}$ $a.s.$

(ii) $\mathbf{S}$ has independent increments, i.e. for all $n \in \mathbb{N}$ and $0 \leq t_0 < t_1 < \ldots < t_n$ the random vectors $\mathbf{S}(t_0), \mathbf{S}(t_1) - \mathbf{S}(t_0), \ldots, \mathbf{S}(t_n) - \mathbf{S}(t_{n-1})$ are independent.

(iii) $\mathbf{S}$ has stationary increments, i.e. the distribution of $\mathbf{S}(t+h) - \mathbf{S}(t)$ does not depend on t.

[1] The term "Lévy process" is in honor of the French mathematician Paul Lévy, one of the founding fathers of modern theory of stochastic processes.

(iv) $\mathbf{S}$ is *stochastically continuous*, i.e. for every $t > 0$ and $\varepsilon > 0$

$$\lim_{h\to 0} P(|\mathbf{S}(t+h) - \mathbf{S}(t)| > \varepsilon) = 0.$$

Stochastic continuity does not imply that the sample paths of the process are continuous, but says that discontinuities occur at random times. In other words, the probability of a jump at any fixed time $t > 0$ is zero.

For each $t > 0$ the distribution of $\mathbf{S}(t)$ will be in the class of infinitely divisible distributions, i.e. for each $n \in \mathbb{N}$, there exist n i.i.d. random vectors $\mathbf{Y}_{t,1}, \ldots, \mathbf{Y}_{t,n}$ such that $\mathbf{Y}_{t,1}+\ldots+\mathbf{Y}_{t,n}$ has the same distribution as $\mathbf{S}(t)$. This is obvious by setting $\mathbf{Y}_{t,1} = \mathbf{S}(t/n)$, $\mathbf{Y}_{t,2} = \mathbf{S}(2t/n) - \mathbf{S}(t/n), \ldots, \mathbf{Y}_{t,n} = \mathbf{S}(t) - \mathbf{S}((n-1)t/n)$.

The distribution of a Lévy process $(\mathbf{S}(t))_{t\geq 0}$ is characterized by the *Lévy-Khintchine representation* of the characteristic function

$$\mathbb{E}\big[e^{i(z,\mathbf{S}(t))}\big] = e^{t\Psi(z)},\ t \geq 0,\ z \in \mathbb{R}^d,$$

with

$$\Psi(z) \ = \ i(\gamma, z) - \frac{1}{2} z^\top A z + \int_{\mathbb{R}^d} \big(e^{i(z,x)} - 1 - i(z,x)1_{\{|x|\leq 1\}}\big)\Pi(dx), , \tag{1.1.1}$$

where $(.\,,.)$ denotes the inner product in $\mathbb{R}^d$, $|\cdot|$ is an arbitrary norm in $\mathbb{R}^d$ and 1_A represents the indicator function of set A. The triplet (γ, A, Π) with $\gamma \in \mathbb{R}^d$, the symmetric non-negative definite $d \times d$ matrix A and the measure Π satisfying (1.1.1) uniquely determines the distribution of $\mathbf{S}$ and is called the *characteristic triplet*. The vector γ is not an intrinsic quantity and depends on the truncation function in the Lévy Khintchine representation, which in (1.1.1) is chosen to be 1. Although it does not have in general a clear intuitive meaning, in some cases one can write the representation differently to give such a parameter the meaning of a *drift*. The matrix A in characteristic triplet represents the *covariance matrix* of Gaussian part of the Lévy process. The *Lévy measure* Π is a measure on $\mathbb{R}^d$ satisfying

$$\Pi(\{\mathbf{0}\}) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} (|x|^2 \wedge 1)\,\Pi(dx) \ < \ \infty.$$

The class of Lévy processes is very rich and consists of a wide range of applicable processes. Brownian motion is a well-known Lévy process with a.s. continuous sample paths. The characteristic triplet for a Brownian motion without drift will be $(0, A, 0)$. The Lévy measure Π for a Borel set $B \subset \mathbb{R}^d$ denotes the expected number of jumps per unit of time with size in B. For a d-dimensional Lévy process the Lévy measure Π is either finite or infinity. If $\Pi(\mathbb{R}^d) < \infty$, the process is a compound Poisson process (CPP), and with probability one it has finitely many jumps in each bounded time interval. The Lévy process with infinite Lévy measure, i.e. $\Pi(\mathbb{R}^d) = \infty$, has countably many jumps in every bounded time interval forming a dense subset of $[0, \infty)$, cf. Theorem 21.3. of Sato [32].

An important class of Lévy processes are the *spectrally one-sided Lévy processes.* These are the processes, which have a.s. only positive jumps or only negative jumps. For the class of spectrally positive Lévy processes, there is a subclass called *increasing Lévy processes* or *subordinators* which are very important ingredients for building Lévy-based models in finance, cf. [13], Section 3.5. A subordinator is a spectrally positive Lévy process with a.s. non-decreasing sample paths, i.e. it has with probability one positive jumps and a positive drift. This implies immediately that it has sample paths of finite variation. According to Theorem 21.5 in Sato [32] or Proposition 3.10 in Cont and Tankov [13] a subordinator does not have a Gaussian part, i.e. $A = 0$, and its Lévy measure Π satisfies

$$\Pi(\mathbb{R}^d \setminus \mathbb{R}^d_+) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d_+} (|x| \wedge 1) \, \Pi(dx) < \infty,$$

where $\mathbb{R}^d_+ = [0, \infty)^d \setminus \{\mathbf{0}\}$ and $\mathbf{0}$ is the zero in $\mathbb{R}^d$.

1.2 Dependence of random variables

Let $\mathbf{X} = (X_1, \dots, X_d)$ be a d-dimensional random variable in $\mathbb{R}^d$. Its distribution function (d.f.) F is usually defined by

$$F(x_1, \dots, x_d) := P(X_1 \le x_1, \dots, X_d \le x_d), \quad (x_1, \dots, x_d) \in \overline{\mathbb{R}}^d,$$

with one-dimensional marginal d.f.s $F_i(x_i) := P(X_i \le x_i)$ for $i = 1, \dots, d$. The dependence between the components of $\mathbf{X}$ is usually modelled by a so-called (distributional) copula.

Before giving a definition for a copula function, we start with some definitions from Kallsen and Tankov [25], Section 2.

Definition 1.2.1. *Let* $A \subset \overline{\mathbb{R}}^d$ *and define* $F : A \to \overline{\mathbb{R}}$. *For* $\mathbf{a}, \mathbf{b} \in A$ *with* $\mathbf{a} \leq \mathbf{b}$ *(componentwise) and* $\overline{(\mathbf{a}, \mathbf{b}]} \subset A$*, the* F-volume of $(\mathbf{a}, \mathbf{b}]$ *is defined by*

$$V_F\big((\mathbf{a}, \mathbf{b}]\big) = \sum_{u \in \{a_1, b_1\} \times \ldots \times \{a_d, b_d\}} (-1)^{N(u)} F(u)$$

where $N(u) = \#\{k \,|\, u_k = a_k\}$.

In particular, $V_F((a, b]) = F(b) - F(a)$ for $d = 1$ and $V_F((\mathbf{a}, \mathbf{b}]) = F(b_1, b_2) + F(a_1, a_2) - F(a_1, b_2) - F(b_1, a_2)$ for $d = 2$. The F-volume of any interval is equal to its Lebesgue measure provided $F(u_1, \ldots, u_d) = \prod_{i=1}^d u_i$.

Definition 1.2.2. *Let* $A \subset \overline{\mathbb{R}}^d$ *and define* $F : A \to \overline{\mathbb{R}}$. F *is called* d-increasing *if* $V_F\big((\mathbf{a}, \mathbf{b}]\big) \geq 0$ *for all* $\mathbf{a}$, $\mathbf{b} \in A$ *with* $\mathbf{a} \leq \mathbf{b}$ *and* $\overline{(\mathbf{a}, \mathbf{b}]} \subset A$.

Definition 1.2.3. *A* d*-increasing function* $C : [0, 1]^d \to [0, 1]$ *is called a* (distributional) copula *if*

(i) $C(u_1, \ldots, u_d) = 0$ *if* $u_i = 0$ *for at least one* $i \in \{1, 2, \ldots, d\}$,

(ii) $C(\underbrace{1, \ldots, 1}_{i-1}, u_i, 1, \ldots, 1) = u_i$ *for every* $i \in \{1, 2, \ldots, d\}$ *and* $u_i \in [0, 1]$.

The property (i) in Def. 1.2.3 is called *groundedness* and a function with this property called a *grounded function.* From a probabilistic point of view Definition 1.2.3 means that a copula is a distribution function on $[0, 1]^d$ with uniform margins. For the theory of copulas the fundamental theorem is *Sklar's theorem* in [33] which represents a copula as dependence structure of a random vector or of a multivariate distribution function. In fact, for a multivariate d-dimensional distribution function F with marginals $F_1, \ldots, F_d$ and for all $\boldsymbol{x} = (x_1, \ldots, x_d) \in \overline{\mathbb{R}}^d$,

$$F(x_1, \ldots, x_d) = C\Big(F_1(x_1), \ldots, F_d(x_d)\Big). \tag{1.2.1}$$

The copula C is unique, if the marginal distribution functions are all continuous; otherwise, C is uniquely determined on $Ran\, F_1 \times \ldots \times Ran\, F_d$. Conversely, the function F in (1.2.1) is a d-dimensional distribution function if C is a copula and $F_1, \ldots, F_d$ are distribution functions. That is, the copula C determines the characteristics of a distribution that do not depend on the margins, but together with the margins, allow to reconstruct the entire distribution. This is what is called the "dependence structure".

The *survival function* $\overline{F}$ corresponding to the d.f. F is defined by

$$\overline{F}(x_1, \ldots, x_d) := P(X_1 > x_1, \ldots, X_d > x)$$

with univariate margins $\overline{F}_i(x_i) := P(X_i > x_i)$ for $i = 1, \ldots, d$. Equation (1.2.1) can be reformulated for a d-dimensional survival function $\overline{F}$ in terms of a copula $\overline{C}$ and the univariate survival functions $\overline{F}_1, \overline{F}_2, \ldots, \overline{F}_d$, i.e.

$$\overline{F}(x_1, \ldots, x_d) = \overline{C}\Big(\overline{F}_1(x_1), \ldots, \overline{F}_d(x_d)\Big).$$

The function $\overline{C}$ is called the *survival copula*.

1.3 Dependence of Lévy processes

In principle, the whole distribution of a d-dimensional Lévy process $\mathbf{S}$ is determined by the law of $\mathbf{S}(t)$ for some $t > 0$. Therefore, one can model the dependence structure among the components of $\mathbf{S}$ by the distributional copula C_t of the random vector $\mathbf{S}(t)$. However, this approach as discussed in Kallsen and Tankov [25] has two drawbacks:

- For given infinitely divisible one-dimensional laws the copulas that can yield an infinitely divisible d-dimensional law depend strongly on the margins and can not be calculated in general.

- The distributional copula C_t of $\mathbf{S}(t)$ depends on t and for some $h \neq t$ the copula C_h of $\mathbf{S}(h)$ cannot in general be computed from C_t. One also needs to know the marginal distributions at times h and t. Furthermore, even if C_h can be calculated from C_t and the margins, the numerical computation will be very demanding.

We need therefore to redefine the notion of dependence structure, independent of the margins and with preserving the Lévy property and the dynamic structure of the Lévy process. The dependence of a Lévy process in general has two parts. First, the dependence of the Gaussian part which is entirely determined by its covariance matrix A in the characteristic triplet. Second, the dependence structure of jump part of the process which is completely characterized by the Lévy measure Π. As a result of the Lévy-Itô decomposition the Gaussian part of every Lévy process is independent from its jump part, cf. Theorem 19.2 of Sato [32]. Therefore, these two dependence structures can be treated separately from each other. Since the covariance matrix of a Lévy process is well-studied, the processes in this thesis are supposed to have no Gaussian part and their dependence is studied in terms of the Lévy measure.

Since the Lévy measure is a measure on $\mathbb{R}^d$ and not on $[0,1]^d$, we need to define a suitable notion of a copula. Moreover, one has to take care of the fact that the Lévy measure is possibly infinite with a singularity at the origin. Because the singularity is in the center of the domain of interest, each corner of the Lévy measure must be treated separately. Therefore, we need a special interval associated with every $x \in \mathbb{R}$.

Definition 1.3.1. *(Tail integral, Def. 3.3 in [25])Let* $\mathbf{S} = (\mathbf{S}(t))_{t\geq 0}$ *be a* $\mathbb{R}^d$*-valued Lévy process with Lévy measure* Π*. The tail integral of* $\mathbf{S}$ *is the function* $\overline{\Pi} : (\mathbb{R} \setminus \{\mathbf{0}\})^d \to \mathbb{R}$ *defined by*

$$\overline{\Pi}(x_1, \ldots, x_d) := \prod_{i=1}^{d} sgn(x_i)\Pi\Big(\prod_{i=1}^{d} \mathcal{I}(x_i)\Big).$$

where $sgn(x) = 1_{\{x\geq 0\}} - 1_{\{x<0\}}$ *and*

$$\mathcal{I}(x) = \begin{cases} [x, \infty), & x \geq 0, \\ (-\infty, x), & x < 0. \end{cases}$$

As known from the definition of the tail integral, the Lévy measure is always considered on the cones away from the origin because of the singularity at zero, see Fig. 1.1 for an illustration of a bivariate tail integral. From Def. 1.3.1, the tail integral for $\boldsymbol{x} = (x_1, x_2) > \mathbf{0}$ is given by

$$\overline{\Pi}(x_1, x_2) = \Pi([x_1, \infty) \times [x_2, \infty)).$$

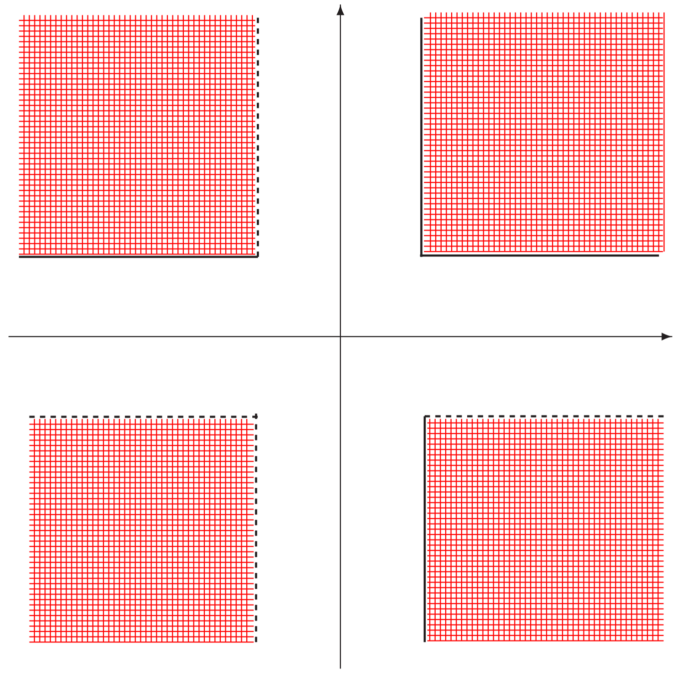

Figure 1.1: Illustration of the support of a bivariate tail integral $\overline{\Pi}$ in Def.1.3.1 for a value $\boldsymbol{x}$ in different quadrants.

This corresponds to the tail integral of a bivariate spectrally positive Lévy process where the process has only positive jumps. This is somehow similar to the survival function and will be discussed later. Now we define the margins of a Lévy process.

Definition 1.3.2. *(I-margins of a Lévy process, Def. 3.4 in [25]) Let $\mathbf{S} = (S_1, \ldots, S_d)$ be a Lévy process with values in $\mathbb{R}^d$ and with Lévy measure Π. Let $I \subset \{1, 2, \ldots, d\}$ be a non-empty index set. The I-margin of $\mathbf{S}$ is the Lévy process $\mathbf{S}_I := (S_i)_{i \in I}$. The Lévy measure and the tail integral of $\mathbf{S}_I$ are denoted by Π_I and $\overline{\Pi}_I$ and given by*

$$\begin{aligned} \Pi_I(B) &= \Pi(\{\boldsymbol{x} \in \mathbb{R}^d \ : \ (x_i)_{i \in I} \in B\}), \quad B \in \mathcal{B}(\mathbb{R}^{|I|} \setminus \{\mathbf{0}\}) \\ \overline{\Pi}_I((x_i)_{i \in I}) &= \prod_{i \in I} sgn(x_i)\, \Pi_I\Big(\prod_{i \in I} \mathcal{I}(x_i)\Big). \end{aligned}$$

where $|I|$ denotes the cardinality of I.

For every $i \in \{1, 2, \ldots, d\}$ we simplify the one-dimensional marginal process, Lévy measure and tail integral by S_i, Π_i and $\overline{\Pi}_i$.Moreover, in this thesis a margin (without clarifying the index set I) always denotes a one-dimensional margin.

Now before starting to define a suitable notion of a copula for the dependence structure of a Lévy measure, we recall from Def. 1.3.1 that the tail integrals are supported by $\mathbb{R}$.

Therefore, the domain of a copula for a Lévy measure should suitably be generalized from $[0,1]^d$ to $\mathbb{R}^d$.

Definition 1.3.3. *(Lévy copula, Def. 3.1 in [25]) A function $\mathfrak{C} : \overline{\mathbb{R}}^d \to \overline{\mathbb{R}}$ is called a Lévy copula if*

(i) $\mathfrak{C}$ is a d-increasing function,

(ii) $\mathfrak{C}$ is grounded, i.e. $\mathfrak{C}(u_1, \ldots, u_d) = 0$, if $u_i = 0$ for at least one $i \in \{1, \ldots, d\}$,

(iii) $\mathfrak{C}(u_1, \ldots, u_d) \neq \infty$ for $(u_1, \ldots, u_d) \neq (\infty, \ldots, \infty)$,

(iv) $\mathfrak{C}$ has Lebesgue margins, i.e. $\mathfrak{C}_i(u) = u$ for $i \in \{1, \ldots, d\}$ and $u \in \mathbb{R}$.

The next theorem plays a fundamental role to link the tail integral of a multidimensional Lévy process to the tail integrals of its margins. It is similar to Sklar's theorem for distribution functions and called Sklar's theorem for tail integrals or Lévy processes.

Theorem 1.3.4. *(Sklar's theorem for Lévy processes, Theorem 3.6. in [25])*

(i) Let $\mathbf{S} = (S_1, \ldots, S_d)$ be a $\mathbb{R}^d$-valued Lévy process. Then there exists a Lévy copula $\mathfrak{C}$ such that the tail integrals of $\mathbf{S}$ satisfy

$$\overline{\Pi}((\boldsymbol{x}_i)_{i\in I}) = \mathfrak{C}_I((\overline{\Pi}_i(x_i))_{i\in I}) \tag{1.3.1}$$

for every non-empty $I \subset \{1, \ldots, d\}$ and every $(x_i)_{i\in I} \in (\mathbb{R} \setminus \{0\})^{|I|}$ with

$$\mathfrak{C}_I((u_i)_{i\in I}) := \lim_{a\to\infty} \sum_{(u_j)_{j\in I^c} \in\{-a,\infty\}} \mathfrak{C}(u_1, \ldots, u_d) \prod_{j\in I^c} sgn(u_i).$$

The Lévy copula $\mathfrak{C}$ is unique on $\prod_{i=1}^d \overline{Ran\overline{\Pi}}$.

(ii) Let $\mathfrak{C}$ be a d-dimensional Lévy copula and $\overline{\Pi}_i$, for $i = 1, \ldots, d$ tail integrals of real-valued Lévy processes. Then there exists a $\mathbb{R}^d$-valued Lévy process $\mathbf{S}$ whose components have tail integrals $\overline{\Pi}_1, \ldots, \overline{\Pi}_d$ and whose marginal tail integrals satisfy equation (1.3.1) for every non-empty $I \subset \{1, \ldots, d\}$ and every $(x_i)_{i\in I} \in (\mathbb{R} \setminus \{0\})^{|I|}$. The Lévy measure of $\mathbf{S}$ is uniquely determined by $\mathfrak{C}$ and $\overline{\Pi}_i$, for $i = 1, \ldots, d$.

In particular, for $I = \{1, \ldots, d\}$

$$\overline{\Pi}(x_1, \ldots, x_d) = \mathfrak{C}(\overline{\Pi}_1(x_1), \ldots, \overline{\Pi}_d(x_d)).$$

1.3.1 Stable Lévy processes

Since chapters 3 and 4 in this thesis deal mainly with the Stable Lévy processes, we review in this section some results and facts about the class of α-stable Lévy processes. For more details we refer to Samorodnitsky and Taqqu [31], Chapter 7, Sato [32], Chapter 4, Bertoin [7], Chapter VIII and Cont and Tankov [13], Section 3.7.

A *stable Lévy process* is a Lévy process $\mathbf{S} = (\mathbf{S}(t))_{t \geq 0}$ in which each $\mathbf{S}(t)$ is a stable random variable in $\mathbb{R}^d$. Its stability or scaling property is characterized by the index α. That is, for every $a > 0$ the processes $(\mathbf{S}(t))_{t \geq 0}$ and $(a^{-\frac{1}{\alpha}}\mathbf{S}(at))_{t \geq 0}$ have the same finite-dimensional distributions. It follows from the Lévy-Khintchine formula that the range of the index α is $(0, 2]$. The cases $\alpha = 1$ and $\alpha = 2$ are special and correspond to Cauchy and Gaussian processes, respectively. We only consider $\alpha \in (0, 2)$. Theorem 14.3 in Sato [32] gives the characteristic triplet of an α-stable Lévy process.

Theorem 1.3.5. *(Theorem 14.3 (ii) in [32]) Let $\mathbf{S}$ be a Lévy process in $\mathbb{R}^d$ with characteristic triplet (γ, A, Π) and $\alpha \in (0, 2)$. The following statements are equivalent:*

(i) $\mathbf{S}(1)$ *is α-stable.*

(ii) $A = 0$ *and Π is homogeneous of order α, i. e. for all $t > 0$, it holds*

$$\Pi(tB) = t^{-\alpha}\Pi(B) \quad \textit{for} \quad B \in \mathcal{B}(\mathbb{R}^d).$$

(iii) $A = 0$ *and there is a finite measure $\widetilde{\rho}$ on the unit sphere $\mathcal{S}^{d-1} := \{\boldsymbol{x} \in \mathbb{R}^d : |\boldsymbol{x}| = 1\}$ such that*

$$\Pi(B) = \int_{\mathcal{S}^{d-1}} \int_0^\infty 1_B(r\xi) r^{-\alpha-1} dr \widetilde{\rho}(d\xi) \quad \textit{for} \quad B \in \mathcal{B}(\mathbb{R}^d).$$

The probability measure $\frac{\tilde{\rho}(\cdot)}{\tilde{\rho}(\mathcal{S}^{d-1})}$ is called the *spectral measure* of Π. If $d = 1$, then $\mathcal{S}^0 = \{1, -1\}$ and the Lévy measure Π is absolutely continuous w.r.t. the Lebesgue measure, with density

$$\Pi(dx) = \begin{cases} c_1 x^{-1-\alpha} dx & \text{if } x > 0, \\ c_2 |x|^{-1-\alpha} dx & \text{if } x < 0, \end{cases}$$

with $c_1 \geq 0$, $c_2 \geq 0$ and $c_1 + c_2 > 0$. The process has no positive jumps if $c_1 = 0$ and no negative jumps if $c_2 = 0$. It is *symmetric* if $c_1 = c_2$.

1.4 Outline of the thesis

This thesis is based on three papers written jointly with Professor Claudia Klüppelberg. Besides developing a general MLE theory for Lévy processes, we exemplify this theory for compound Poisson and stable Lévy processes with dependence structure given by a Clayton Lévy copula. We also introduce a new parameter estimation method for a multidimensional Lévy process. In this approach which we call the two-step estimation method, we first estimate the parameters of marginal processes, and then estimate in a second step the dependence parameter. In the following we present an overview to the thesis, summarized from Chapter 2 to Chapter 4.

Chapter 2 is devoted to the parameter estimation of a bivariate compound Poisson process. This chapter is organized as follows. After a short introduction in Section 2.1, multivariate compound Poisson processes and their possible dependence structure are discussed in Section 2.2. The dependence structure of a bivariate model and the decomposition of a process into three independent processes is also shown in detail. Section 2.3 is devoted to the definition of a tail integral, a Lévy copula and Sklar's theorem for Lévy processes. Section 2.4 presents the likelihood function of a bivariate compound Poisson process by means of the aforementioned decomposition of a process. The likelihood functions of an exponential Clayton model and a Weibull Clayton model are also calculated in this section. These likelihood functions are needed for the maximum likelihood estimation of the parameters in the following section of this chapter. The distributional copula of the joint jumps for a bivariate compound Poisson process is also calculated in this section. It is proven that the copula of joint jumps for a CPP with a Clayton Lévy copula regardless of jump size distributions is a distributional Clayton copula. A new simulation method for a bivariate compound Poisson process is proposed in Section 2.5, where the idea of the decomposition of a process in Section 2.2 is used. Moreover, the parameters of a bivariate exponential Clayton model are estimated by a maximum likelihood approach. The parameters are estimated either for different values of the dependence parameter or for all parameters of a full model. Section 2.6 presents a real data analysis. The Danish fire insurance data are analysed in detail and, based on this analysis, a Weibull-Clayton model is fitted to the data.

Chapter 3 is devoted to the parameter estimation of a bivariate stable Lévy process. The dependence of the process is modelled by a Clayton Lévy copula, a homogeneous Lévy copula of order one. Following an introduction into the problem in Section 3.1, some definitions on Lévy processes and Lévy copulas are presented in Section 3.2. In Section 3.3 the maximum likelihood estimation of the parameters of a one-dimensional Lévy process is explained in detail. The characteristic function of a jump-truncated process is calculated to show the intensity and the jump size distribution function of the resulting CPP. These, of course, are the base for the likelihood function of a one-dimensional CPP. As observation scheme for one-dimensional Lévy processes all the jumps larger than some $\varepsilon > 0$ are taken into account. The asymptotic behaviour of the estimates based on the Fisher information matrix is discussed in Section 3.3.2. Asymptotic normality is proved when the number of jumps tends to infinity regardless of whether the observed time interval tends to ∞ for a fixed truncation point $\varepsilon > 0$ or the truncation point tends to 0 over a fixed observed time interval. Section 3.4 extends this ML theory to a bivariate stable Lévy process where the small jumps are truncated and the likelihood function is based on bivariate jumps larger than ε in both components. The consequence of jump truncation on Lévy copula and the asymptotic behaviour of estimates are also discussed in this section. Section 3.5 contains the simulation study for a bivariate α-stable Clayton subordinator. As expected the more precise estimates are obtained by taking a smaller value of the jump truncation point. Moreover, the asymptotic covariance matrix of estimators for a bivariate α-stable subordinator is calculated by a numerical integration and by a Monte Carlo simulation.

In Chapter 4 a new estimation method for multivariate Lévy processes is introduced. In this parametric approach which we call a two-step method, the parameters of the marginal processes are estimated first, and then in a second step we estimate only the dependence parameter. In Section 4.2 we explain the truncation scheme of the observed jumps and present the bivariate α-stable Clayton subordinator. Section 4.3 is dedicated to the two-step estimation procedure. We introduce this method for Lévy processes and compare the likelihood equations in a two-step method with the equations for the maximum likelihood estimation. We apply the two-step estimation procedure to a bivariate α-stable Clayton subordinator. The parameters estimation of the model is explained for different marginal parameters and its reduction for common marginal parameters. In Section 4.4 we calculate the Godambe information matrix for a bivariate α-stable Clayton subordinator analytically. The consistency of the two-step estimators as well as their asymptotic normality are proved in this section. The asymptotic covariance matrix is calculated as the observation interval tends to ∞ for a fixed truncation point ε, or as ε tends to zero for a fixed observation interval. The log-likelihood and the score functions of the full model are calculated in Section 4.5. These can be compared with the log-likelihood of a bivariate compound Poisson process based only on the joint jumps of a bivariate α-stable process as discussed in Section 3.4. Finally, in Section 4.6, a small simulation study is performed to compare the quality of all three estimation methods: the maximum likelihood estimation based only on joint jumps of the process larger than some $\varepsilon > 0$ in both components, the maximum likelihood estimation of a full model based on single and joint jumps larger than ε, and the two-step estimation method. These three procedures are discussed in Sections 3.4, 4.3 and 4.5, respectively.

Chapter 2

Parameter Estimation of a Bivariate Compound Poisson Process

SUMMARY

In this article, we review the concept of a Lévy copula to describe the dependence structure of a bivariate compound Poisson process. In this first statistical approach we consider a parametric model for the Lévy copula and estimate the parameters of the full dependent model based on a maximum likelihood approach. This approach ensures that the estimated model remains in the class of multivariate compound Poisson processes. A simulation study investigates the small sample behaviour of the MLEs, where we also suggest a new simulation algorithm. Finally, we apply our method to the Danish fire insurance data.

2.1 Introduction

Copulas open a convenient way to represent the dependence of a probability distribution. In fact they provide a complete characterization of possible dependence structures of a random vector with fixed margins. Moreover, using copulas, one can construct multivariate distributions with a pre-specified dependence structure from a collection of univariate laws. Modern results about copulas originate more than fifty years back when Sklar [33] defined and derived the fundamental properties of a copula. Further important references are Nelson [27] and Joe [23]. Financial applications of copulas have been numerous in recent years; cf. Cherubini, Luciano and Vecchiato [12] for examples and further references.

We are considering multivariate Lévy processes, whose dependence can be modelled by a "copula" on the components of the Lévy measure. This has been suggested in Tankov [34] for subordinators, the case of general Lévy processes was treated in Kallsen and Tankov [25]; Lévy copulas can also be found in the monograph of Cont and Tankov [13].

Modelling dependence in multivariate Lévy processes by Lévy copulas offers the same flexibility for modelling the marginal Lévy processes independently of their dependence structure as we know from distributional copulas. Statistical methods, which have existed for distributional copulas for a long time, still have to be developed for Lévy copulas. The present paper is a first step.

The Lévy copula concept has been applied to insurance risk problems; more precisely, Bregman and Klüppelberg [8] have used this approach for ruin estimation in multivariate models. Eder and Klüppelberg [14] extended this work to derive the so-called quintuple law for sums of dependent Lévy processes. This describes the ruin event by stating not only the ruin probability, but also quantities like ruin time, overshoot, undershoot; i.e. they present a ladder process analysis. The notion of multivariate regular variation can also be linked to Lévy copulas, which is investigated and presented in Eder and Klüppelberg [15].

In a series of papers, Böcker and Klüppelberg [9, 10, 11] used a multivariate compound Poisson process to model operational risk in different business lines and risk types. Again dependence is modelled by a Lévy copula. Analytic approximations for the operational Value-at-Risk explain the influence of dependence on the institution's total operational risk.

In view of these economic problems, which are well recognised in academia and among practitioners, the present paper is concerned with statistical inference for bivariate compound Poisson processes. Our method is based on Sklar's theorem for Lévy copulas, which guarantees that the estimated model is again multivariate compound Poisson.

This approach, whose importance is already manifested by the above mentioned economic applications as well as in a data analysis at the end of our paper, will have far reaching implications for the estimation of multivariate Lévy processes with infinite activity sample paths as is relevant in finance. This has been worked out in Esmaeili and Klüppelberg [18].

Our paper is organized as follows. Section 2.2 presents the definition of a multivariate compound Poisson process (CPP) and explains the dependence structure in three possible ways. This prepares the ground for a new simulation algorithm for multivariate compound Poisson processes and for the maximum likelihood estimation. Then we define the concept of a tail integral and a Lévy copula for such processes in Section 2.3. In Section 2.4 we derive the likelihood function for the process parameters, where we assume that we observe the continuous-time sample path. In Section 2.5 we suggest a new simulation algorithm for multivariate compound Poisson processes and show it at work by simulating a bivariate CPP, whose dependence structure is modelled by a Clayton Lévy copula. Finally, in Section 2.6 we fit a compound Poisson process to the bivariate Danish fire insurance data, and present some conclusions in Section 2.7.

2.2 The multivariate compound Poisson process

A d-dimensional *compound Poisson process* (CPP) is a Lévy process $\mathbf{S} = (\mathbf{S}(t))_{t\geq 0}$, i.e. a process with independent and stationary increments, defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$, with values in $\mathbb{R}^d$. It is stochastically continuous, i.e. for all $a > 0$,

$$\lim_{t\to h} P(|\mathbf{S}(t) - \mathbf{S}(h)| > a) = 0\,, \quad h \geq 0\,,$$

and as is well-known (see e.g. Sato [32], Def. 1.6), a càdlàg version exists, and we assume this property throughout. For each $t > 0$ the characteristic function has the so-called Lévy-Khintchine representation:

$$\mathbb{E}[e^{i(\mathbf{z},\mathbf{S}(t))}] \;=\; \exp\left\{t\int_{\mathbb{R}^d}\left(e^{i(\mathbf{z},\boldsymbol{x})}-1\right)\Pi(d\boldsymbol{x})\right\}, \quad \mathbf{z}\in\mathbb{R}^d\,, \tag{2.2.1}$$

where $(\cdot,\cdot)$ denotes the inner product in $\mathbb{R}^d$. The so-called Lévy measure Π is a measure on $\mathbb{R}^d$ satisfying $\Pi(\{\mathbf{0}\}) = 0$ and $\int_{\mathbb{R}^d}\Pi(d\boldsymbol{x}) < \infty$. Moreover, CPPs are the only Lévy processes with finite Lévy measure.

According to Sato [32], Theorem 4.3, a compound Poisson process is a stochastic process

$$\mathbf{S}(t) = \sum_{i=1}^{N(t)} \mathbf{Z}_i\,, \quad t \geq 0\,, \tag{2.2.2}$$

where $(N(t))_{t\geq 0}$ is a *homogeneous Poisson process* with intensity $\lambda > 0$ and $(\mathbf{Z}_i)_{i\in\mathbb{N}}$ is a sequence of i.i.d. random variables with values in $\mathbb{R}^d$. Moreover $(N(t))_{t\geq 0}$ and $(\mathbf{Z}_i)_{i\in\mathbb{N}}$ are independent and the $\mathbf{Z}_i$'s have no atom in $\mathbf{0}$, i.e. $P(\mathbf{Z}_1 = \mathbf{0}) = 0$.

To prepare the ground for our statistical analysis based on a Lévy copula to come, we present a bivariate CPP in more detail. In particular, we give three approaches to understand the dependence structure of such a process in more detail.

Assume that for $i \in \mathbb{N}$ the bivariate vector $\mathbf{Z}_i$ has df G with components Z_{1i} and Z_{2i} with dfs G_1 and G_2, respectively. It is, of course, possible that single jumps in one of the marginal processes occur, in which case the probability measure of the marks Z_{1i} and Z_{2i} have atoms in 0; i.e. they are not continuous.

In our first approach we write

$$\mathbf{S}(t) \;=\; \sum_{i=1}^{N(t)} (Z_{1i}, Z_{2i}) = \Big(\sum_{i=1}^{N(t)} Z_{1i}, \sum_{i=1}^{N(t)} Z_{2i}\Big), \quad t \geq 0, \tag{2.2.3}$$

where we set $p_1 := P(Z_{1i} = 0)$ and $p_2 := P(Z_{2i} = 0)$ and recall that possibly $p_1, p_2 > 0$. Then for almost all $\omega \in \Omega$,

$$S_1(t) = \sum_{i=1}^{N_1(t)} X_i, \quad t \geq 0 \quad \text{ and } \quad S_2(t) = \sum_{i=1}^{N_2(t)} Y_i, \quad t \geq 0, \tag{2.2.4}$$

where X_i and Y_i take only the non-zero values of Z_1 and Z_2, respectively, and inherit the independence of $N_1(\cdot) = (1-p_1)N(\cdot)$ and $N_2(\cdot) = (1-p_2)N(\cdot)$. To make this precise, for $i = 1, 2$ and a Borel set $A \subset \mathbb{R}\setminus\{0\}$ we can write

$$P\Big(\sum_{j=1}^{N(t)} Z_{ij} \in A\Big) \;=\; P\Big(\sum_{j=1}^{N^{i1}(t)} Z_{ij}\ 1_{\{Z_{ij}\neq 0\}} + \sum_{j=1}^{N^{i2}(t)} Z_{ij}\ 1_{\{Z_{ij}=0\}} \in A\Big), \; t \geq 0\,, \tag{2.2.5}$$

where $N^{i1}(\cdot)$ and $N^{i2}(\cdot)$ count the non-zero and zero jumps, respectively. By the thinning property of the Poisson process, they are again Poisson processes. Since the last summation in (2.2.5) is zero, we conclude that for almost all $\omega \in \Omega$

$$\sum_{j=1}^{N^{11}(t)} Z_{1j} = \sum_{j=1}^{N_1(t)} X_j, \quad t \geq 0, \quad \text{and} \quad \sum_{j=1}^{N^{21}(t)} Z_{2j} = \sum_{j=1}^{N_2(t)} Y_j\,, \quad t \geq 0,$$

are compound Poisson processes. Here $(N_1(t))_{t\geq 0}$, $(N_2(t))_{t\geq 0}$ are Poisson processes with intensities $\lambda_1 = (1-p_1)\lambda$ and $\lambda_2 = (1-p_2)\lambda$, respectively, and $(X_i)_{i\in\mathbb{N}}$ and $(Y_i)_{i\in\mathbb{N}}$ are sequences of i.i.d. random variables with dfs given for all $x \in \mathbb{R}$ by

$$F_1(x) = P(Z_1 \leq x \mid Z_1 \neq 0) \quad \text{and} \quad F_2(y) = P(Z_2 \leq y \mid Z_2 \neq 0)$$

Unlike G_1 and G_2, the dfs F_1 and F_2 have no mass in 0.

Our second approach is based on the representation of a compound Poisson process as an integral with respect to a Poisson random measure M; cf. Sato [32], Theorems 19.2 and 19.3. For almost all $\omega \in \Omega$ we have the representation

$$\begin{aligned}
\mathbf{S}(t) &= \sum_{i=1}^{N(t)} (Z_{1i}, Z_{2i}) \\
&= \int_0^t \int_{\mathbb{R}^2\setminus\{\mathbf{0}\}} \mathbf{z} M(ds \times d\mathbf{z}) \qquad (2.2.6) \\
&= \int_0^t \int_{(\mathbb{R}\setminus\{0\})\times\{0\}} \mathbf{z} M(ds \times d\mathbf{z}) + \int_0^t \int_{\{0\}\times(\mathbb{R}\setminus\{0\})} \mathbf{z} M(ds \times d\mathbf{z}) \\
&\quad + \int_0^t \int_{(\mathbb{R}\setminus\{0\})^2} \mathbf{z} M(ds \times d\mathbf{z})
\end{aligned}$$

where M is a Poisson random measure on $[0,\infty) \times (\mathbb{R}^2 \setminus \{\mathbf{0}\})$ with intensity measure $ds\,\Pi(d\mathbf{z})$.

Corresponding to the first two integrals we can introduce two compound Poisson processes $S_1^{\perp}$ and $S_2^{\perp}$, which are called the *independent parts of* (S_1, S_2). They are independent of each other and never jump together. On the other hand, the third integral corresponds to a compound Poisson process which is supported on sets in $(\mathbb{R} \setminus \{\mathbf{0}\})^2$, and this part of (S_1, S_2) measures the simultaneous jumps of S_1 and S_2. We denote it by $(S_1^{\parallel}, S_2^{\parallel})$, and it is the *(jump) dependent part of* (S_1, S_2). Since its components $S_1^{\|}$ and $S_2^{\|}$ always jump together, they must have the same jump intensity parameter, which we denote by $\lambda^{\|}$. Now we can decompose $(S_1(t), S_2(t))_{t\geq 0}$ for almost all $\omega \in \Omega$ into

$$\begin{aligned} S_1(t) &= S_1^{\perp}(t) + S_1^{\|}(t)\,, \quad t \geq 0\,, \\ S_2(t) &= S_2^{\perp}(t) + S_2^{\|}(t)\,, \quad t \geq 0\,. \end{aligned} \tag{2.2.7}$$

Here we see clearly the decomposition of the bivariate compound Poisson process in single jumps in each marginal process and the process of common jumps in both components. It is clear from the properties of the Poisson random measure that the three processes $S_1^{\perp}$, $S_2^{\perp}$ and $(S_1^{\|}, S_2^{\|})$ are compound Poisson and independent.

Our last approach is similar to the previous one, but based on a decomposition of the Lévy measure. It also prepares the ground for the following Section 2.3. Recall that for any Borel set $A \subseteq \mathbb{R}^2 \setminus \{0\}$ its Lévy measure $\Pi(A)$ denotes the expected number of jumps per unit time with size in A. This can be formulated as

$$\Pi(A) = \mathbb{E}\Big[\#\{(t, (\Delta S_1(t), \Delta S_2(t))) \in (0, 1] \times A\}\Big].$$

This set, and hence Π can be decomposed into the following components:

$$\begin{aligned} \Pi_1(A) &= \mathbb{E}\Big[\#\{(t, (\Delta S_1(t), \Delta S_2(t))) \in (0, 1] \times A \mid \Delta S_1(t) \neq 0 \text{ and } \Delta S_2(t) = 0\}\Big], \\ \Pi_2(A) &= \mathbb{E}\Big[\#\{(t, (\Delta S_1(t), \Delta S_2(t))) \in (0, 1] \times A \mid \Delta S_1(t) = 0 \text{ and } \Delta S_2(t) \neq 0\}\Big], \\ \Pi_3(A) &= \mathbb{E}\Big[\#\{(t, (\Delta S_1(t), \Delta S_2(t))) \in (0, 1] \times A \mid \Delta S_1(t) \neq 0 \text{ and } \Delta S_2(t) \neq 0\}\Big]. \end{aligned}$$

Since $\Pi(A) = \Pi_1(A) + \Pi_2(A) + \Pi_3(A)$, the integral of the characteristic function in (2.2.1) can be decomposed into three integrals with different Lévy measures $\Pi_1(A)$, $\Pi_2(A)$ and $\Pi_3(A)$, respectively. Clearly Π_1 is supported by the set $\{(x, 0) \in \mathbb{R}^2 \mid x \in \mathbb{R}\}$. We set $\Pi_1(A) = \Pi_1^{\perp}(A_1)$, where $A_1 = \{x \in \mathbb{R} \mid (x, 0) \in A\}$. Then the first integral reduces to a one-dimensional integral related only to the component S_1. Similarly, for the second part $\Pi_2(A) = \Pi_2^{\perp}(A_2)$, where $A_2 = \{y \in \mathbb{R} \mid (0, y) \in A\}$; hence the second integral also reduces to a one-dimensional integral. By introducing the notation $\Pi^{\|}$ for Π_3, the characteristic function in (2.2.1) can be decomposed into

$$
\begin{aligned}
&\mathbb{E}\left[e^{iz_1S_1(t)+iz_2S_2(t)}\right] \\
&= \exp\left\{t\int_{\mathbb{R}}(e^{iz_1x}-1)\Pi_1^{\perp}(dx)+t\int_{\mathbb{R}}(e^{iz_2y}-1)\Pi_2^{\perp}(dy)+t\int_{\mathbb{R}^2}(e^{iz_1x+iz_2y}-1)\Pi^{\parallel}(dx\times dy)\right\} \\
&= E\left[e^{iz_1S_1^{\perp}(t)}\right]E\left[e^{iz_2S_2^{\perp}(t)}\right]E\left[e^{iz_1S_1^{\parallel}(t)+iz_2S_2^{\parallel}(t)}\right]. \qquad (2.2.8)
\end{aligned}
$$

Note that the Lévy measure $\Pi_1^{\perp}$ gives the mean number of jumps of S_1 such that S_2 does not have a jump at the same time. Similarly, the mean number of jumps for S_2, when S_1 has no jump, is measured by $\Pi_2^{\perp}$. Corresponding to $\Pi_1^{\perp}$ and $\Pi_2^{\perp}$ we find again the two processes $S_1^{\perp}$ and $S_2^{\perp}$ which we called the *independent parts of* (S_1, S_2). On the other hand, $\Pi^{\parallel}$ is supported by sets in $(R\setminus\{0\})^2$, and we denoted this part by $(S_1^{\parallel}, S_2^{\parallel})$, which is the *(jump) dependent part of* (S_1, S_2). This results in the same representation (2.2.7) as above.

This decomposition has also been presented in Cont and Tankov [13], Section 5.5 and Böcker and Klüppelberg [9], Section 3. Note that for completely dependent components we have $S_1^{\perp} = S_2^{\perp} = 0$ a.s. On the other hand, for independent components, the third part of the integral (2.2.6) or (2.2.8) is zero and this means that the components a.s. never jump together.

2.3 The Lévy copula

We shall present an estimation procedure for a bivariate compound Poisson process based on Lévy copulas. The reason for this is two-fold. Working with real data it may not be so easy to estimate statistically the components on the right-hand side of (2.2.7) so that the resulting statistical model is a bivariate compound Poisson process. Moreover, the ingredients require quite a number of parameters, which makes it desirable to find a parsimonious model. We are convinced that the notion of a Lévy copula plays here the same important role as a copula does for multivariate dfs.

Mainly for ease of notation we shall present our Lévy copula concept for spectrally non-negative CPPs only; i.e. for CPPs with non-negative jumps only. Since the Lévy copula for a general CPP is defined for each quadrant separately, this is no restriction of the theory developed. Furthermore, the insurance claims data considered later also justify this restriction.

Lévy copulas are defined via the tail integral of a Lévy process.

Definition 2.3.1. *Let* Π *be a Lévy measure on* $\mathbb{R}_+^d$*. The* tail integral *is a function* $\overline{\Pi} : [0,\infty]^d \to [0,\infty]$ *defined by*

$$\overline{\Pi}(x_1,\ldots,x_d) = \begin{cases} \Pi([x_1,\infty)\times\cdots\times[x_d,\infty))\,, & (x_1,\ldots,x_d)\in[0,\infty)^d \\ 0\,, & \textit{if } x_i=\infty \textit{ for at least one } i. \end{cases} \tag{2.3.1}$$

The marginal tail integrals are defined analogously for $i = 1,\ldots,d$ *as* $\overline{\Pi}_i(x) = \Pi_i([x,\infty))$ *for* $x \geq 0$.

Next we define the Lévy copula for a spectrally positive Lévy process; for details see Nelson [27], Tankov [34] or Cont and Tankov [13].

Definition 2.3.2. *The* Lévy copula *of a spectrally positive Lévy process is a d-increasing grounded function* $\mathfrak{C} : [0,\infty]^d \to [0,\infty]$ *with margins* $\mathfrak{C}_k(u) = u$ *for all* $u \in [0,\infty]$ *and* $k = 1,\ldots,d$.

The notion of groundedness guarantees that $\mathfrak{C}$ defines a measure on $[0,\infty]^d$; indeed a Lévy copula is a d-dimensional measure with Lebesgue margins.

The following theorem is a version of Sklar's theorem for spectrally positive Lévy process; for a proof we refer to Tankov [34], Theorem 3.1.

Theorem 2.3.3 (Sklar's Theorem for Lévy copulas)**.**
Let $\overline{\Pi}$ *denote the tail integral of a spectrally positive d-dimensional Lévy process, whose components have Lévy measures* $\Pi_1,\ldots,\Pi_d$*. Then there exists a Lévy copula* $\mathfrak{C} : [0,\infty]^d \to [0,\infty]$ *such that for all* $(x_1,x_2,\ldots,x_d) \in [0,\infty]^d$

$$\overline{\Pi}(x_1,\ldots,x_d) = \mathfrak{C}(\overline{\Pi}_1(x_1),\ldots,\overline{\Pi}_d(x_d)). \tag{2.3.2}$$

If the marginal tail integrals are continuous, then this Lévy copula is unique. Otherwise, it is unique on $Ran\overline{\Pi}_1 \times \ldots \times Ran\overline{\Pi}_d$*.*

Conversely, if $\mathfrak{C}$ is a Lévy copula and $\overline{\Pi}_1, \ldots, \overline{\Pi}_d$ are marginal tail integrals of a spectrally positive Lévy process, then the relation (2.3.2) *defines the tail integral of a d-dimensional spectrally positive Lévy process and $\overline{\Pi}_1, \ldots, \overline{\Pi}_d$ are tail integrals of its components.*

This result opens up now a way of estimating multivariate compound Poisson processes by separating the marginal compound Poisson processes and coupling them with the dependence structure given by the Lévy copula. We shall show the procedure in details in the next section.

2.4 Maximum likelihood estimation of the parameters of a Lévy measure

Now the stage is set to tackle our main problem, namely the maximum likelihood estimation of the parameters of a bivariate spectrally positive CPP based on the observation of a sample path of the bivariate model in $[0, T]$ for fixed $T > 0$.

Obviously, representation (2.2.3) suggests estimating the rate of the compound Poisson process based on the i.i.d. exponential arrival times and, independently, the bivariate distribution function of (Z_1, Z_2). Since both marginal random variables may have an atom in 0, and in the examples we are concerned about, they indeed have, we are faced with the estimation of a mixture model. This is one reason, why we base our estimation on representation (2.2.8). The other motivation comes from possible extensions of our estimation method to general Lévy processes; cf. Esmaeili and Klüppelberg [18].

Consequently, we assume throughout that the decomposition (2.2.7) holds for the observed path. We write for $t \in [0, T]$,

$$\begin{aligned}\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} &= \begin{pmatrix} \sum_{i=1}^{N_1(t)} X_i \\ \sum_{j=1}^{N_2(t)} Y_j \end{pmatrix} = \begin{pmatrix} S_1^{\perp}(t) + S_1^{\parallel}(t) \\ S_2^{\perp}(t) + S_2^{\parallel}(t) \end{pmatrix} \\ &= \begin{pmatrix} \sum_{i=1}^{N_1^{\perp}(t)} X_i^{\perp} + \sum_{j=1}^{N^{\parallel}(t)} X_j^{\parallel} \\ \sum_{i=1}^{N_2^{\perp}(t)} Y_i^{\perp} + \sum_{j=1}^{N^{\parallel}(t)} Y_j^{\parallel} \end{pmatrix} \end{aligned} \quad (2.4.1)$$

with the familiar independence structure of the Poisson counting processes and the jump variables. Although, as described above, every bivariate CPP has three independent parts, the parts are linked by a common set of parameters in the frequency part as well as in the jump size distributions.

Our approach is an extension of the maximum likelihood method for the one-dimensional compound Poisson model; see e.g. Basawa and Prakasa Rao [6], Chapter 6.

Assume that we observe the bivariate CPP (S_1, S_2) continuously over a fixed time interval $[0, T]$. The process S_1 has frequency parameter $\lambda_1 > 0$ and jump size distribution F_1 and the process S_2 has frequency parameter $\lambda_2 > 0$ and jump size distribution F_2. Observing a CPP continuously over a time period is equivalent to observing all jump times and jump sizes in this time interval.

Let $N(T) = n$ denote the total number of jumps occurring in $[0, T]$, which decompose in the number $N_1^{\perp}(T) = n_1^{\perp}$ of jumps occurring only in the first component, the number $N_2^{\perp}(T) = n_2^{\perp}$ of jumps occurring only in the second component, and the number $N^{\parallel}(T) = n^{\parallel}$ of jumps occurring in both components. We denote by $\widetilde{x}_1, \ldots, \widetilde{x}_{n_1^{\perp}}$ the observed jumps occurring only in the first component, by $\widetilde{y}_1, \ldots, \widetilde{y}_{n_2^{\perp}}$ the observed jumps occurring only in the second component, and by $(x_1, y_1), \ldots, (x_{n^{\parallel}}, y_{n^{\parallel}})$ the observed jumps occurring in both components.

Theorem 2.4.1. *Assume an observation scheme as above. Assume that $\boldsymbol{\theta_1}$ is a parameter of the marginal density f_1 of the first jump component only, and $\boldsymbol{\theta_2}$ a parameter of the marginal density f_2 of the second jump component only, and that $\boldsymbol{\delta}$ is a parameter of the Lévy copula. Assume further that $\frac{\partial^2}{\partial u \partial v}\mathfrak{C}(u, v; \boldsymbol{\delta})$ exists for all $(u, v) \in (0, \lambda_1) \times (0, \lambda_2)$,* *which is the domain of $\mathfrak{C}$. Then the full likelihood of the bivariate CPP is given by*

$$\begin{aligned}
&L(\lambda_1, \lambda_2, \boldsymbol{\theta_1}, \boldsymbol{\theta_2}, \boldsymbol{\delta}) \\
= \quad & (\lambda_1)^{n_1^{\perp}} e^{-(\lambda_1^{\perp})T} \prod_{i=1}^{n_1^{\perp}} \left[f_1(\widetilde{x}_i; \boldsymbol{\theta_1}) \left(1 - \frac{\partial}{\partial u} \mathfrak{C}(u, \lambda_2; \boldsymbol{\delta}) \Big|_{u=\lambda_1 \overline{F}_1(\widetilde{x}_i; \boldsymbol{\theta_1})} \right) \right] \\
& \times (\lambda_2)^{n_2^{\perp}} e^{-(\lambda_2^{\perp})T} \prod_{i=1}^{n_2^{\perp}} \left[f_2(\widetilde{y}_i; \boldsymbol{\theta_2}) \left(1 - \frac{\partial}{\partial v} \mathfrak{C}(\lambda_1, v; \boldsymbol{\delta}) \Big|_{v=\lambda_2 \overline{F}_2(\widetilde{y}_i; \boldsymbol{\theta_2})} \right) \right] \qquad (2.4.2) \\
& \times (\lambda_1 \lambda_2)^{n^{\parallel}} e^{-\lambda^{\parallel} T} \prod_{i=1}^{n^{\parallel}} \left[f_1(x_i; \boldsymbol{\theta_1}) f_2(y_i; \boldsymbol{\theta_2}) \frac{\partial^2}{\partial u \partial v} \mathfrak{C}(u, v; \boldsymbol{\delta}) \Big|_{u=\lambda_1 \overline{F}_1(x_i; \boldsymbol{\theta_1}), v=\lambda_2 \overline{F}_2(y_i; \boldsymbol{\theta_2})} \right]
\end{aligned}$$

with $\lambda^{\parallel} = \lambda^{\parallel}(\boldsymbol{\delta}) = \mathfrak{C}(\lambda_1, \lambda_2, \boldsymbol{\delta})$ and $\lambda_i^{\perp}(\boldsymbol{\delta}) = \lambda_i - \lambda^{\parallel}(\boldsymbol{\delta})$ for $i = 1, 2$.

Proof. To calculate the likelihood function, we use representation (2.2.7) in combination with the independence as it is manifested in (2.2.8). This corresponds to the representation of the tail integrals for $i = 1, 2$ as

$$\overline{\Pi}_i =: \overline{\Pi}_i^{\perp} + \overline{\Pi}_i^{\parallel},$$

where $\overline{\Pi}_i$ denotes the marginal tail integral and $\overline{\Pi}_i^{\perp}$ and $\overline{\Pi}_i^{\parallel}$ are the tail integrals of the independent and jump dependent parts, respectively. Then, setting

$$\lambda^{\parallel} = \lim_{x,y \to 0^+} \overline{\Pi}(x, y) = \mathfrak{C}(\lambda_1, \lambda_2; \boldsymbol{\delta}) \quad \text{and} \quad \lambda_i^{\perp} = \lambda_i - \lambda^{\parallel} \quad \text{for} \quad i = 1, 2,$$

we obtain the independent parts and the jump dependent part of (S_1, S_2) as

$$\begin{aligned}
\lambda_1^{\perp} \overline{F}_1^{\perp}(x) &= \lambda_1 \overline{F}_1(x) - \lambda^{\parallel} \overline{F}_1^{\parallel}(x) = \lambda_1 \overline{F}_1(x) - \mathfrak{C}(\lambda_1 \overline{F}_1(x), \lambda_2; \boldsymbol{\delta}), \\
\lambda_2^{\perp} \overline{F}_2^{\perp}(y) &= \lambda_2 \overline{F}_2(y) - \lambda^{\parallel} \overline{F}_2^{\parallel}(y) = \lambda_2 \overline{F}_2(y) - \mathfrak{C}(\lambda_1, \lambda_2 \overline{F}_2(y); \boldsymbol{\delta}), \qquad (2.4.3) \\
\lambda^{\parallel} \overline{F}^{\parallel}(x, y) &= \mathfrak{C}(\lambda_1 \overline{F}_1(x), \lambda_2 \overline{F}_2(y); \boldsymbol{\delta}), \quad x, y > 0
\end{aligned}$$

Let now $L_1(\lambda_1^{\perp}, \boldsymbol{\theta_2})$ be the marginal likelihood function based on the observations of the jump times and jump sizes of the first component $S_1^{\perp}$. To derive L_1 let $\widetilde{t}_1, \ldots, \widetilde{t}_{n_1^{\perp}}$ denote the jump times of $S_1^{\perp}$, and define the sequence of inter-arrival times $\widetilde{T}_k = \widetilde{t}_k - \widetilde{t}_{k-1}$ for $k = 1, \ldots, n_1^{\perp}$ with $\widetilde{t}_0 = 0$. Then the $\widetilde{T}_k$ are i.i.d. exponential random variables with mean

$1/\lambda_1^\perp$ and they are independent of the observed jump sizes $\widetilde{x}_1, \ldots, \widetilde{x}_{n_1^\perp}$. The likelihood function of the observations concerning $S_1^\perp$ is given by

$$\begin{aligned} L_1(\lambda_1^\perp, \boldsymbol{\theta_1}) &= \prod_{i=1}^{n_1^\perp} \left(\lambda_1^\perp e^{-\lambda_1^\perp \widetilde{T}_i}\right) \times \; e^{-\lambda_1^\perp (T - \widetilde{t}_{n_1^\perp})} \times \prod_{i=1}^{n_1^\perp} f_1^\perp(\widetilde{x}_i; \boldsymbol{\theta_1}) \\ &= (\lambda_1^\perp)^{n_1^\perp} e^{-\lambda_1^\perp T} \prod_{i=1}^{n_1^\perp} f_1^\perp(\widetilde{x}_i; \boldsymbol{\theta_1}) \,, \end{aligned} \tag{2.4.4}$$

where the density $f_1^\perp$ is found by taking the derivative in the first equation of (2.4.3). The second part $S_2^\perp$ is treated analogously and we obtain $L_2(\lambda_2^\perp, \boldsymbol{\theta_2})$ as (2.4.4) with $\lambda_1^\perp$ replaced by $\lambda_2^\perp$ and $f_1^\perp(\widetilde{x}_i, \boldsymbol{\theta_1})$ replaced by $f_2^\perp(\widetilde{y}_i, \boldsymbol{\theta_2})$. For the joint jump part of the process, that is $(S_1^\parallel, S_2^\parallel)$, we observe the number $n^\parallel = n_1 - n_1^\perp = n_2 - n_2^\perp$ of joint jumps with frequency $\lambda^\parallel$ at times $t_1, \ldots, t_{n^\parallel}$ with the observed bivariate jump sizes $(x_1, y_1), \ldots, (x_{n^\parallel}, y_{n^\parallel})$. Denote $T_k = t_k - t_{k-1}$ and $F^\parallel(x, y)$ the joint distribution of the jump sizes with joint density $f^\parallel(x, y)$. These are observations of a jump dependent CPP with frequency parameter $\lambda^\parallel$ and Lévy measure concentrated in $(0, \infty)^2$. Recall the formula for $(x, y) \in (0, \infty)^2$, which is a consequence of the formula after Theorem 5.4 on p. 148 in Cont and Tankov [13],

$$\Pi(dx, dy) = \frac{\partial^2}{\partial u \partial v} \mathfrak{C}(u, v, \boldsymbol{\delta}) \Big|_{u = \lambda_1 \overline{F}_1(x, \boldsymbol{\theta_1}), v = \lambda_2 \overline{F}_2(y, \boldsymbol{\theta_2})} \Pi_1(dx) \Pi_2(dy) \,.$$

In our case the joint density of the Lévy measure on the left hand side is given by $\lambda^\parallel f^\parallel(x, y)$. The derivative $\frac{\partial^2}{\partial u \partial v} \mathfrak{C}(u, v, \boldsymbol{\delta})$ exists by assumption. Then the likelihood of the joint jump process is given by the product in the third line of (2.4.2). This concludes the proof. □

Remark 2.4.2. Note that this estimation procedure ensures that the estimated model is again a bivariate CPP.

This applies for instance to the following parametric Lévy copula family.

Example 2.4.3. [Clayton Lévy copula]
The *Clayton Lévy copula* is defined as

$$\mathfrak{C}(u, v) = (u^{-\delta} + v^{-\delta})^{-1/\delta} \,, \quad u, v > 0 \,,$$

where $\delta > 0$ is the Lévy copula parameter. We calculate

$$\begin{aligned}\frac{\partial}{\partial u}\mathfrak{C}(u,v) &= \Big(1+(\frac{u}{v})^{\delta}\Big)^{-1/\delta-1},\\ \frac{\partial^2}{\partial u \partial v}\mathfrak{C}(u,v) &= (\delta+1)(uv)^{-\delta-1}(u^{-\delta}+v^{-\delta})^{-1/\delta-2},\\ &= (\delta+1)(uv)^{\delta}(u^{\delta}+v^{\delta})^{-1/\delta-2}, \quad u,v>0\,.\end{aligned}$$

We observe that the joint jump intensity is given by

$$\lambda^{\|} = (\lambda_1^{-\delta}+\lambda_2^{-\delta})^{-\frac{1}{\delta}}\,.$$

Two specific examples, which will be used later, are the following:

(i) For the *exponential Clayton model* the marginal jump distributions are for $i=1,2$ exponentially distributed with parameters $\theta_i > 0$ and densities $f_i(z;\theta_i) = \theta_i e^{-\theta_i z}$ for $z \geq 0$. The likelihood function for the continuously observed bivariate process $(S_1(t), S_2(t))_{0\leq t\leq T}$ with the notation as in Theorem 2.4.1 is given by

$$\begin{aligned}L(\lambda_1,\lambda_2,\theta_1,\theta_2,\delta) &= (\theta_1\lambda_1)^{n_1^{\perp}} e^{-\lambda_1^{\perp}T-\theta_1\sum_{i=1}^{n_1^{\perp}}\widetilde{x}_i} \prod_{i=1}^{n_1^{\perp}}\left[1-\Big(1+(\frac{\lambda_1}{\lambda_2})^{\delta}e^{-\delta\theta_1\widetilde{x}_i}\Big)^{-\frac{1}{\delta}-1}\right],\\ &\times\ (\theta_2\lambda_2)^{n_2^{\perp}} e^{-\lambda_2^{\perp}T-\theta_2\sum_{i=1}^{n_2^{\perp}}\widetilde{y}_i} \prod_{i=1}^{n_2^{\perp}}\left[1-\Big(1+(\frac{\lambda_2}{\lambda_1})^{\delta}e^{-\delta\theta_2\widetilde{y}_i}\Big)^{-\frac{1}{\delta}-1}\right],\\ &\times\ \Big((1+\delta)\theta_1\theta_2(\lambda_1\lambda_2)^{\delta+1}\Big)^{n^{\|}} e^{-\lambda^{\|}T-(1+\delta)(\theta_1\sum_{i=1}^{n^{\|}}x_i+\theta_2\sum_{i=1}^{n^{\|}}y_i)}\\ &\times\ \prod_{i=1}^{n^{\|}}(\lambda_1^{\delta}e^{-\theta_1\delta x_i}+\lambda_2^{\delta}e^{-\theta_2\delta y_i})^{-\frac{1}{\delta}-2}\,.\end{aligned}$$

(ii) For the *Weibull Clayton model* the marginal jump distributions are for $i=1,2$ Weibull distributed with parameters $a_i, b_i > 0$ and densities $w_i(z;a_i,b_i) = \frac{b_i}{a_i^{b_i}} z^{b_i-1}e^{-(z/a_i)^{b_i}}$ for $z \geq 0$. The likelihood function for the continuously observed bivariate process $(S_1(t), S_2(t))_{0\leq t\leq T}$

is given by

$$
\begin{aligned}
& L(\lambda_1,\lambda_2,a_1,b_1,a_2,b_2,\delta) \qquad (2.4.5)\\
= \quad & \left(\lambda_1 b_1 a_1^{-b_1}\right)^{n_1^\perp} e^{-\lambda_1^\perp T - \sum_{i=1}^{n_1^\perp} (\widetilde{x}_i/a_1)^{b_1}} \prod_{i=1}^{n_1^\perp} \left[\widetilde{x}_i^{b_1-1} \left(1 - \left(1 + \left(\frac{\lambda_1 e^{-(\widetilde{x}_i/a_1)^{b_1}}}{\lambda_2} \right)^{\delta} \right)^{-1/\delta-1} \right) \right] \\
\times \quad & \left(\lambda_2 b_2 a_2^{-b_2}\right)^{n_2^\perp} e^{-\lambda_2^\perp T - \sum_{i=1}^{n_2^\perp} (\widetilde{y}_i/a_2)^{b_2}} \prod_{i=1}^{n_2^\perp} \left[\widetilde{y}_i^{b_2-1} \left(1 - \left(1 + \left(\frac{\lambda_2 e^{-(\widetilde{y}_i/a_2)^{b_2}}}{\lambda_1} \right)^{\delta} \right)^{-1/\delta-1} \right) \right] \\
\times \quad & \left((1+\delta)(\lambda_1\lambda_2)^{1+\delta} b_1 b_2 a_1^{-b_1} a_2^{-b_2}\right)^{n^\parallel} e^{-\lambda^\parallel T - (1+\delta)\sum_{i=1}^{n^\parallel}\left((x_i/a_1)^{b_1} + (y_i/a_2)^{b_2}\right)} \\
\times \quad & \prod_{i=1}^{n^\parallel} \left[{x_i}^{b_1-1} {y_i}^{b_2-1} \left(\left(\lambda_1 e^{-(x_i/a_1)^{b_1}}\right)^{\delta} + \left(\lambda_2 e^{-(y_i/a_2)^{b_2}}\right)^{\delta} \right)^{-1/\delta-2} \right]
\end{aligned}
$$

□

For a bivariate CPP, where dependence is modelled by a Lévy copula, the bivariate distribution of the joint jumps of the process exhibits a specific dependence structure, which can also be described by a distributional copula, or better by the corresponding survival copula. We explain this for the Clayton Lévy copula.

Example 2.4.4. [Continuation of Example 2.4.3]
Denote by $\overline{C}$ the survival copula of the joint jumps of $(S_1^\parallel(t), S_2^\parallel(t))_{t\geq 0}$ given by

$$\overline{F}^\parallel(x,y) = \overline{C}\left(\overline{F}_1^\parallel(x), \overline{F}_2^\parallel(y)\right). \qquad (2.4.6)$$

Assume further that the jump distributions F_1 and F_2 have no atom at 0. From the last equation of (2.4.3) we see that

$$\overline{F}_1^\parallel(x) = \lim_{y\to 0} \frac{1}{\lambda^\parallel} \mathfrak{C}(\lambda_1 \overline{F}_1(x), \lambda_2 \overline{F}_2(y))$$

and analogously for $\overline{F}_2^\parallel$. Here equation (2.4.6) can be rewritten as

$$\frac{1}{\lambda^\parallel}\mathfrak{C}(\lambda_1\overline{F}_1(x),\lambda_2\overline{F}_2(y)) = \overline{C}\left(\frac{1}{\lambda^\parallel}\mathfrak{C}(\lambda_1\overline{F}_1(x),\lambda_2), \frac{1}{\lambda^\parallel}\mathfrak{C}(\lambda_1,\lambda_2\overline{F}_2(y))\right).$$

For the Clayton Lévy copula $\mathfrak{C}$ the right hand side is equal to

$$\overline{C}\left(\left(\frac{(\lambda_1\overline{F}_1(x))^{-\delta}+\lambda_2^{-\delta}}{\lambda_1^{-\delta}+\lambda_2^{-\delta}}\right)^{-\frac{1}{\delta}}, \left(\frac{\lambda_1^{-\delta}+(\lambda_2\overline{F}_2(y))^{-\delta}}{\lambda_1^{-\delta}+\lambda_2^{-\delta}}\right)^{-\frac{1}{\delta}}\right) = \left(\frac{(\lambda_1\overline{F}_1(x))^{-\delta}+(\lambda_2\overline{F}_2(y))^{-\delta}}{\lambda_1^{-\delta}+\lambda_2^{-\delta}}\right)^{-\frac{1}{\delta}}$$

Abbreviating the arguments of $\overline{C}$ by u and v (note that $u, v \in (0,1)$) gives

$$(\lambda_1 \overline{F}_1(x))^{-\delta} = u^{-\delta}(\lambda_1^{-\delta} + \lambda_2^{-\delta}) - \lambda_2^{-\delta} \quad \text{and} \quad (\lambda_2 \overline{F}_2(y))^{-\delta} = v^{-\delta}(\lambda_1^{-\delta} + \lambda_2^{-\delta}) - \lambda_1^{-\delta},$$

such that

$$\begin{aligned} \overline{C}(u,v) &= \left(\frac{u^{-\delta}(\lambda_1^{-\delta} + \lambda_2^{-\delta}) - \lambda_2^{-\delta} + v^{-\delta}(\lambda_1^{-\delta} + \lambda_2^{-\delta}) - \lambda_1^{-\delta}}{\lambda_1^{-\delta} + \lambda_2^{-\delta}} \right)^{-\frac{1}{\delta}} \\ &= (u^{-\delta} + v^{-\delta} - 1)^{-\frac{1}{\delta}}, \end{aligned}$$

which is the well-known distributional Clayton copula; cf. Cont and Tankov [13], eq. (5.3) or Joe [23], Family B4 on p. 141. □

2.5 A simulation study

In this section we study the quality of our estimates in a small simulation study. This means that we first have to simulate sample paths of a bivariate CPP on $[0, T]$ for prespecified $T > 0$, equivalently, we simulate the jump times and jump sizes (independently) in this time interval.

In Section 6 of Cont and Tankov [13] various simulation algorithms for Lévy processes have been suggested. We extend here their Algorithm 6.2 to a bivariate setting by invoking decomposition (2.4.1) for given λ_1, λ_2, marginal jump distribution functions F_1, F_2 and a Lévy copula $\mathfrak{C}$.

As we work with a fully parametric bivariate model, we assume that we are given frequency parameters $\lambda_1, \lambda_2 > 0$, the parameters of the marginal jump size distributions $\boldsymbol{\theta_1} \in \mathbb{R}^{k_1}$, $\boldsymbol{\theta_2} \in \mathbb{R}^{k_2}$ for some $k_1, k_2 \in \mathbb{N}$ and, finally, the dependence parameter $\boldsymbol{\delta} \in \mathbb{R}^m$ of the Lévy copula. Moreover, we choose a time interval $[0, T]$.

Then the number of points in the first component is Poisson distributed with frequency $\lambda_1 T$, so generate a Poisson random number $N_1(T)$ with mean $\lambda_1 T$. The number of points in the second component is Poisson distributed with frequency $\lambda_2 T$, so generate a Poisson random number $N_2(T)$ with mean $\lambda_2 T$. Then $\lambda^{\|} T = \mathfrak{C}(\lambda_1, \lambda_2) T$ is the frequency parameter

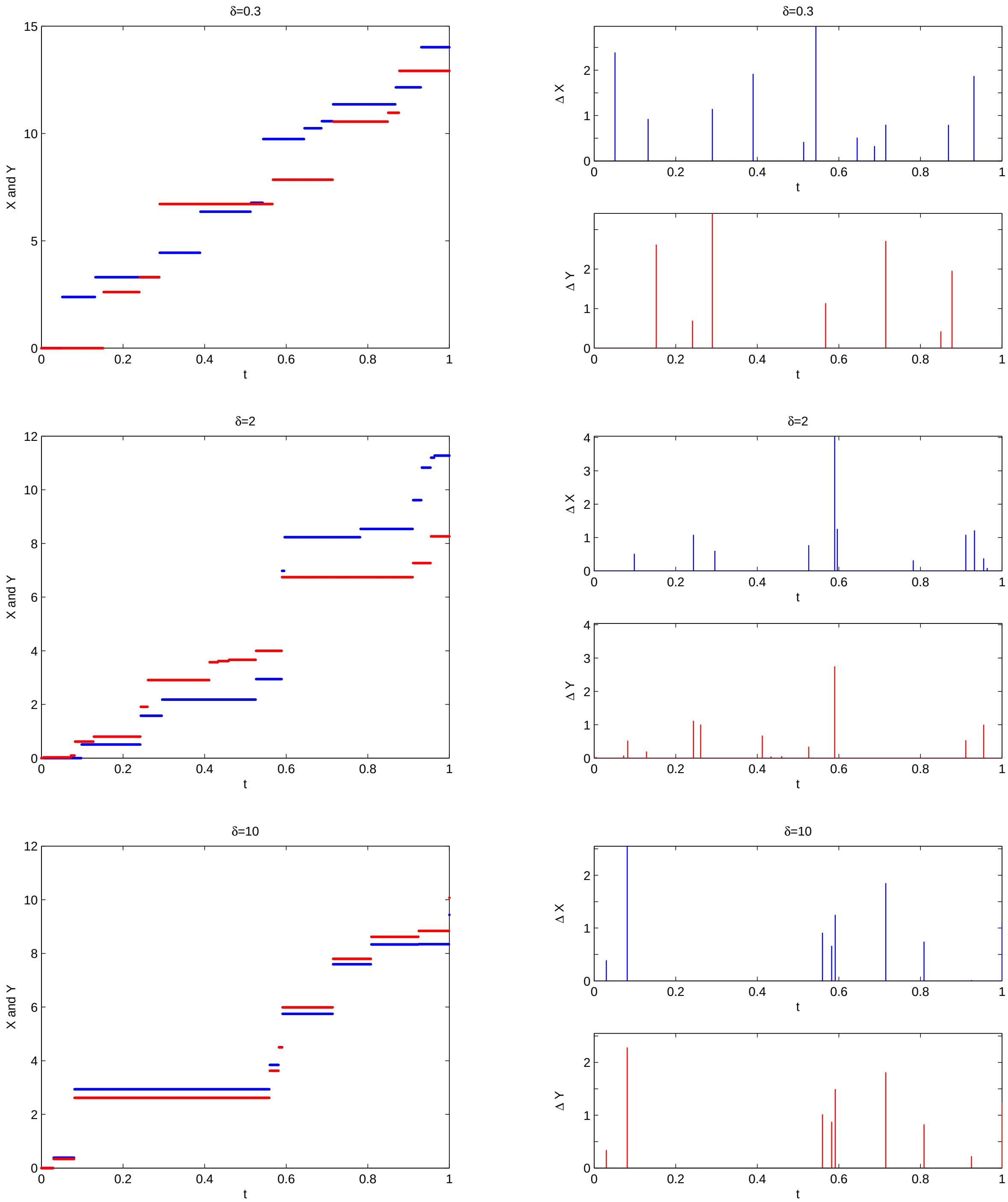

Figure 2.1: Simulation of three bivariate CPPs with exponentially distributed jumps and a Clayton Lévy copula with dependence parameter $\delta = 0.3$ (top), $\delta = 2$ (middle) and $\delta = 10$ (below). The left hand figures show the sample paths of the CPPs, whereas the right hand figures present the same paths as marked Poisson process.

of the joint jumps, so simulate another Poisson random number $N^{\|}(T)$ with frequency $\lambda^{\|}T$. This implies then that $N_1^{\perp}(T) = N_1(T) - N^{\|}(T)$ and $N_2^{\perp}(T) = N_2(T) - N^{\|}(T)$.

Now conditional on these numbers, the Poisson points are uniformly distributed in the interval $[0,T]$, so simulate the correct number of $[0,T]$-uniformly distributed random variables, independently for the three components: $U_{1,i}^{\perp}$ for $i = 1,\dots,N_1^{\perp}(T)$, $U_{2,i}^{\perp}$ for $i = 1,\dots,N_2^{\perp}(T)$, and $U_i^{\|}$ for $i = 1,\dots,N^{\|}(\mathrm{T})$.

Next we simulate the jump sizes. Denote by $(U_{1,i}^{\perp}, X_i^{\perp})$ for $i = 1,\dots,N_1^{\perp}(T)$, $(U_{2,i}^{\perp}, Y_i^{\perp})$ for $i = 1,\dots,N_2^{\perp}(T)$ and $(U_i^{\|}, X_i^{\|}, Y_i^{\|})$ for $i = 1,\dots,N^{\|}(T)$ the marked points of the single jumps and the joint jumps, respectively, then the bivariate trajectory is given by

$$\begin{pmatrix} S_1(t) \\ S_2(t) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N_1^{\perp}(T)} 1_{\{U_{1,i}^{\perp}<t\}} X_i^{\perp} + \sum_{i=1}^{N^{\|}(T)} 1_{\{U_i^{\|}<t\}} X_i^{\|} \\ \sum_{i=1}^{N_2^{\perp}(T)} 1_{\{U_{2,i}^{\perp}<t\}} Y_i^{\perp} + \sum_{i=1}^{N^{\|}(T)} 1_{\{U_i^{\|}<t\}} Y_i^{\|} \end{pmatrix}, \quad 0 < t < T\,.$$

For the marks on these points given by the corresponding jump sizes we need then $N_1^{\perp}(T)$ i.i.d. jump sizes with df $F_1^{\perp}$, $N_2^{\perp}(T)$ i.i.d. jump sizes with df $F_2^{\perp}$, and $N^{\|}(T)$ bivariate jump sizes with df $F^{\|}$, all of them independent. Single jump sizes are generated by $X_i^{\perp} \stackrel{d}{=} F_1^{\perp\leftarrow}(U_i)$, $i = 1,\cdots,N_1^{\perp}(T)$ and $Y_i^{\perp} \stackrel{d}{=} F_2^{\perp\leftarrow}(U_i)$, $i = 1,\cdots,N_2^{\perp}(T)$, where for any increasing function h its generalized inverse is defined as

$$h^{\leftarrow}(u) := \inf\{s \in \mathbb{R} : h(s) \geq u\}\,,$$

(which coincides with the analytical inverse, provided h is strictly monotone).

It remains to simulate the joint jumps $(X_j^{\|}, Y_j^{\|})$ for $j = 1,\dots,N^{\|}(T)$. We use the joint survival copula $\overline{C}$ as in (2.4.6). We simulate standard uniform independent random variables $U_1,\dots,U_{N^{\|}(T)}, V_1,\dots,V_{N^{\|}(T)}$ and recall that $X_j^{\|} \stackrel{d}{=} F_1^{\|\leftarrow}(U_j)$. Then the following standard calculation for a generic pair $(X^{\|}, Y^{\|})$ is well-known:

$$\begin{aligned} \lim_{\Delta x\to 0} P(Y^{\|} > y \mid x < X^{\|} \leq x + \Delta x) &= \lim_{\Delta x\to 0} \frac{\overline{F}^{\|}(x,y) - \overline{F}^{\|}(x+\Delta x, y)}{P(x < \Delta X^{\|} \leq x + \Delta x)} \\ = -\frac{\partial \overline{F}^{\|}(x,y)}{\partial x}\frac{1}{f_1^{\|}(x)} &= -\frac{\partial \overline{C}(\overline{F}_1^{\|}(x), \overline{F}_2^{\|}(y))}{\partial x}\frac{1}{f_1^{\|}(x)} \\ = \frac{\partial}{\partial u}\overline{C}(u, \overline{F}_2^{\|}(y))\Big|_{u=\overline{F}_1^{\|}(x)} &=: \overline{H}_x(y)\,. \end{aligned} \qquad (2.5.1)$$

Value	$\widehat{\text{mean}}$	$\widehat{\text{MSE}}$	$\widehat{\text{MAE}}$	$\widehat{\text{MRB}}$
$\delta = 0.5$	0.4995 (0.0597)	0.0036 (0.0054)	0.0492 (0.0339)	-0.0012
$\delta = 1$	0.9896 (0.0964)	0.0094 (0.0150)	0.0748 (0.0617)	0.0147
$\delta = 3$	3.0583 (0.2828)	0.0834 (0.1121)	0.2314 (0.1727)	-0.0321
$\delta = 5$	5.0279 (0.4494)	0.2027 (0.2764)	0.3511 (0.2819)	0.0147

Table 2.1: Mean, mean squared errors (MSE), mean absolute error (MAE) and mean relative bias (MRB) are presented for 100 MLEs of the Lévy copula parameter of a bivariate exponential Clayton model (Example 2.5.1). Each estimate is calculated from an observed sample path of a bivariate CPP with parameters $\lambda_1 = 100, \lambda_2 = 80, \theta_1 = 1, \theta_2 = 2$ (which are assumed to be known) and unknown dependence parameter δ. The values in brackets show the standard deviation of estimates.

Now we take the generalized inverse $H_x^{\leftarrow}$ and define $Y_j^{\parallel} \stackrel{d}{=} H_x^{\leftarrow}(V_j)$. Then the following calculation convinces us that this algorithm works:

$$
\begin{aligned}
P(F_1^{\parallel\leftarrow}(U) > x, H_{X^{\parallel}}^{\leftarrow}(V) > y) &= P(X^{\parallel} > x) P(H_{X^{\parallel}}^{\leftarrow}(V) > y \mid X^{\parallel} > x) \\
&= P(X^{\parallel} > x) \int_x^{\infty} P(Y^{\parallel} > y \mid X^{\parallel} = t) dF_1^{\parallel}(t) \\
&= P(X^{\parallel} > x) P(Y^{\parallel} > y \mid X^{\parallel} > x) \\
&= P(X^{\parallel} > x, Y^{\parallel} > y)\,.
\end{aligned}
$$

Example 2.5.1. [Simulation of a bivariate exponential Clayton model, continuation of Examples 2.4.3 and 2.4.4]

Let (S_1, S_2) be a bivariate CPP with exponentially distributed jump sizes, i.e. $\overline{F}_i(z) = e^{-\theta_i z}$, $z > 0$, for $i = 1, 2$, and the dependence structure of a Clayton Lévy copula $\mathfrak{C}$ with parameter $\delta > 0$. Assume further $\lambda_1, \lambda_2 > 0$ are the intensities of the marginal Poisson processes. We simulate a bivariate exponential Clayton model over the time interval $[0, 1]$.

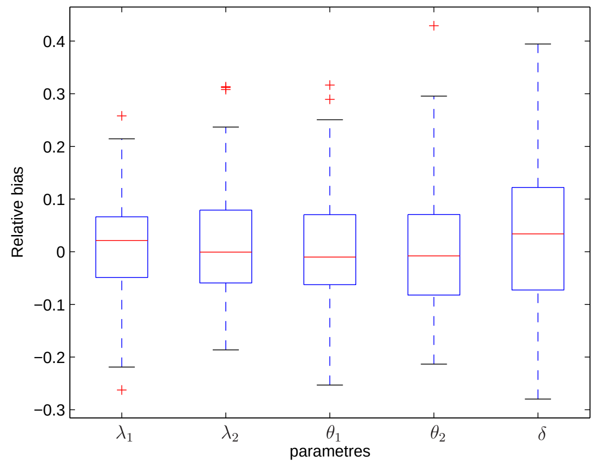

Figure 2.2: Box-plots of the relative bias for the estimates of the exponential Clayton model with parameter values as in Table 2.2.

We apply the above simulation algorithm. The distribution functions of the single jump sizes of the process are for $i = 1, 2$ given by

$$\overline{F}_i^{\perp}(z) = \frac{1}{\lambda_i^{\perp}} \left\{ \lambda_i e^{-\theta_i z} - \left(\lambda_1^{-\delta} e^{\theta_1 \delta z(2-i)} + \lambda_2^{-\delta} e^{\theta_2 \delta z(i-1)} \right)^{-\frac{1}{\delta}} \right\}, \quad z > 0\,,$$

and the bivariate distribution function for the joint jumps has the form

$$\overline{F}^{\parallel}(x, y) = \frac{1}{\lambda^{\parallel}} \left(\lambda_1^{-\delta} e^{\theta_1 \delta x} + \lambda_2^{-\delta} e^{\theta_2 \delta y} \right)^{-\frac{1}{\delta}}, \quad x, y > 0$$

with margins $\overline{F}_1^{\parallel}(x) = \frac{1}{\lambda^{\parallel}} \left(\lambda_1^{-\delta} e^{\theta_1 \delta x} + \lambda_2^{-\delta} \right)^{-\frac{1}{\delta}}$, $x > 0$, and $\overline{F}_2^{\parallel}(y) = \frac{1}{\lambda^{\parallel}} \left(\lambda_1^{-\delta} + \lambda_2^{-\delta} e^{\theta_2 \delta y} \right)^{-\frac{1}{\delta}}$, $y > 0$.

The simulation algorithm:

(i) Generate two random numbers N_1 and N_2 from Poisson distributions with parameters λ_1 and λ_2, respectively. Generate $N^{\parallel}$ from a Poisson distribution with parameter $\lambda^{\parallel} = \mathfrak{C}(\lambda_1, \lambda_2) = (\lambda_1^{-\delta} + \lambda_2^{-\delta})^{-1/\delta}$.

(ii) Generate $N^{\parallel}$, $N_1^{\perp} = N_1 - N^{\parallel}$ and $N_2^{\perp} = N_2 - N^{\parallel}$ independent $[0, 1]$-uniformly distributed random variables. These are the Poisson points of joint and single jumps.

	$\widehat{\lambda}_1$	$\widehat{\lambda}_2$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\delta}$
Values	100	80	1.00	2.00	1.00
$\widehat{\text{mean}}$	100.8377 (9.8302)	80.4022 (8.7985)	1.0105 (0.0979)	2.0326 (0.2158)	1.0097 (0.1197)
$\widehat{\text{MSE}}$	97.3344 (141.0848)	78.9570 (113.1887)	0.0097 (0.0168)	0.0476 (0.0714)	0.0144 (0.0202)
$\widehat{\text{MRB}}$	0.0116	0.0203	-0.0087	0.0022	0.0423

Table 2.2: Estimated mean, mean squared error (MSE) and mean relative bias (MRB) of 100 MLEs of an exponential Clayton model with estimated standard deviations for mean and MSE in brackets.

(iii) Generate independent $U_1, \dots, U_{N_1^\perp}$ and $V_1, \dots, V_{N_2^\perp}$ standard uniform random variables. Then the single jump sizes of both components are found by taking the inverse of $F_1^\perp$ and $F_2^\perp$, that is, $X_i^\perp \stackrel{d}{=} F_1^{\perp\leftarrow}(U_i)$, $i = 1, \dots, N_1^\perp$ and $Y_j^\perp \stackrel{d}{=} F_2^{\perp\leftarrow}(V_j)$, $j = 1, \dots, N_2^\perp$.

(iv) For the bivariate jump sizes, generate new independent $[0,1]$-uniform $U_1, \dots, U_{N^\parallel}$ and $V_1, \dots, V_{N^\parallel}$ random variables. Then $X_i^\parallel \stackrel{d}{=} F_1^{\parallel\leftarrow}(U_i)$ and, given $X_i^\parallel = x$, $Y_i^\parallel \stackrel{d}{=} H_x^\leftarrow(V_i)$, $i = 1, \dots, N^\parallel$, where for fixed $x > 0$, as shown in (2.5.1),

$$\begin{aligned}
\overline{H}_x(y) = \frac{\partial}{\partial u}\overline{C}(u, \overline{F}_2^\parallel(y))\Big|_{u=\overline{F}_1^\parallel(x)} &= \Big(1 + (\frac{u}{v})^\delta - u^\delta\Big)^{-1/\delta-1}\Big|_{u=\overline{F}_1^\parallel(x), v=\overline{F}_2^\parallel(y)} \\
&= \left(1 + \frac{\lambda_1^{-\delta} + \lambda_2^{-\delta}e^{\theta_2\delta y}}{\lambda_1^{-\delta}e^{\theta_1\delta x} + \lambda_2^{-\delta}} - \frac{\lambda^{\parallel^{-\delta}}}{\lambda_1^{-\delta}e^{\theta_1\delta x} + \lambda_2^{-\delta}}\right)^{-1/\delta-1} \\
&= \left(\frac{\lambda_1^{-\delta}e^{\theta_1\delta x} + \lambda_2^{-\delta}e^{\theta_2\delta y}}{\lambda_1^{-\delta}e^{\theta_1\delta x} + \lambda_2^{-\delta}}\right)^{-\frac{1}{\delta}-1}, \quad y > 0.
\end{aligned}$$

Various scenarios are depicted in Figure 2.1. □

Next we show the performance of the MLE estimation from Section 2.4 based on simulated sample paths.

Example 2.5.2. [Estimation of a bivariate exponential Clayton model, continuation of Examples 2.4.3, 2.4.4, and 2.5.1]

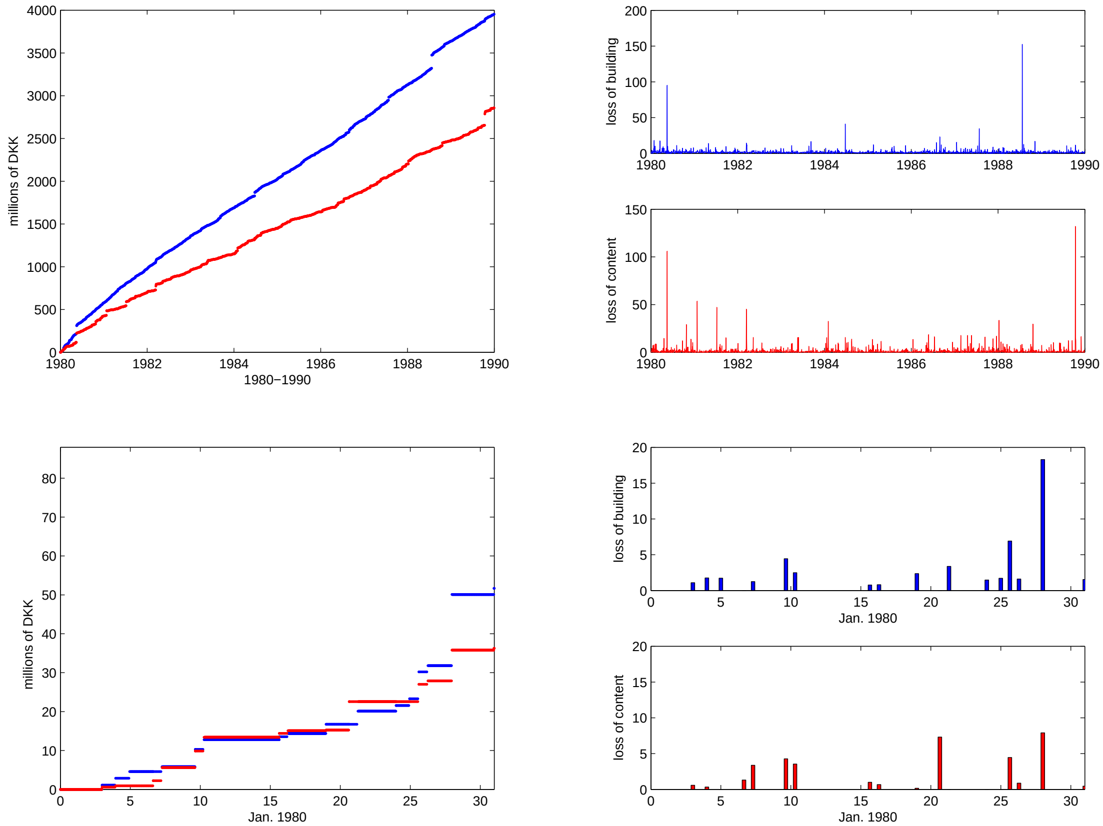

Figure 2.3: The Danish fire insurance data: The top figures show the total losses (left) and the individual losses (right) over the period 1980-1990. The figures below depict the data only for the one-month period of January 1980.

Let (S_1, S_2) be a bivariate CPP with exponentially distributed jump sizes, i.e. $\overline{F}_i(z) = e^{-\theta_i z}$, $z > 0$, for $i = 1, 2$, and the dependence structure of a Clayton Lévy copula $\mathfrak{C}$ with parameter $\delta > 0$. Assume further $\lambda_1, \lambda_2 > 0$ are the intensities of the marginal Poisson processes. We simulate 100 sample paths of a bivariate exponential Clayton model with parameters $\lambda_1 = 100, \lambda_2 = 80, \theta_1 = 1, \theta_2 = 2$ and different δ over the time interval $[0, 1]$ and estimate for each sample path the parameters. The results are summarized in Tables 2.1 and 2.2 and Figure 2.2. Note that the critical parameter is the dependence parameter δ; cf. Figure 2.2. From Table 2.1 we note from the estimated MSE and MAE that its estimation is more precise for small δ than for large. The mean relative bias, on the other hand, remains for all δ near 0. Similar interpretations can be read off from Table 2.2. □

2.6 A real data analysis

In this section we fit a CPP to a bivariate data set. The data we fit is called the Danish fire insurance data and appears in aggregated form in Embrechts et al. [16], Figure 6.2.11. The data are available at *www.ma.hw.ac.uk/∼mcneil/*. As described there, the data were collected at Copenhagen Reinsurance and comprise 2167 fire losses over the period 1980 to 1990. They have been adjusted for inflation to reflect 1985 values and are expressed in millions of Danish Kroner. Every total claim has been divided into loss of building, loss of content and loss of profit. Since the last variable rarely has non-zero value, we restrict ourselves to the first two variables. Figure 2.3 shows the time series and the aggregated process of the data in the whole and in a one-month period of time.

We shall estimate a full bivariate parametric model based on the likelihood function of Theorem 2.4.1. This means that we have to specify the marginal distributions for the losses of buildings and the losses of contents, and we do this for the logarithmic data. As explained above the bivariate data come from originally aggregated data, where the claims (sum of losses of buildings, contents and profits) are larger than one million Danish Kroner. Due to the splitting of the data in losses of buildings and losses of contents, certain losses have become smaller than the threshold for the aggregated data, such effects also appear due to the inflation adjustment. To guarantee that the bivariate data we want to fit come from the same distribution, we have based our analysis on those data, which are larger than one million Danish Kroner after inflation adjustment in both coordinates. This amounts to 940 data points.

An explorative data analysis shows that the family of two-parameter Weibull distributions are appropriate for the log-data. We present the histograms of the log-transformed data in Figures 2.4 with fitted marginal Weibull densities as presented in Example 2.4.3(ii).

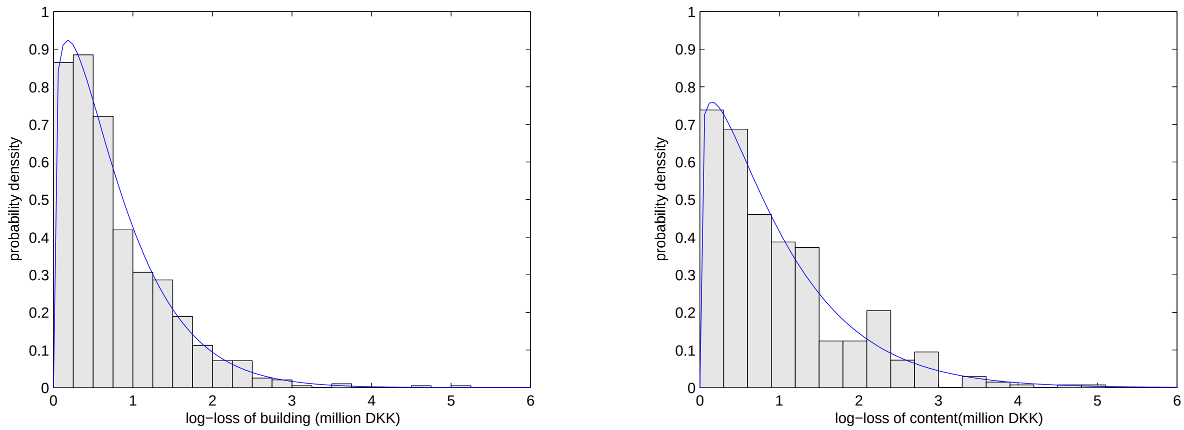

Figure 2.4: Histogram (and estimated Weibull density) of the logarithmic losses of buildings (left) and logarithmic losses of content based on the Danish fire insurance data larger than 1 million Danish Kroner in both variables.

The marginal parameters have been fitted by maximum likelihood estimation giving

$$f_1(x) = 1.5225(\log x)^{0.1954} \exp\left(-1.2737(\log x)^{1.1954}\right), \quad x > 1$$

$$f_2(y) = 1.0863(\log y)^{0.1289} \exp\left(-0.9622(\log y)^{1.1289}\right), \quad y > 1.$$

The corresponding QQ-plots are depicted in Figure 2.5.

It is worth mentioning that modelling with Lévy copulas is useful, when the dependence structure of the Poisson processes matches the dependence of the jump sizes. The reason for this is that the parameter of the Lévy copula models the dependence structure of the Lévy measure, which comprises the intensity of the jumps and the distribution of the jump sizes. By Sklar's theorem for Lévy copulas (cf. Theorem 2.3.3), if the data follow a bivariate compound Poisson process, this kind of dependence structure is exactly, what we expect.

To check the suitability of the model for these data, we first estimate the parameter of the Clayton Lévy copula based on the point processes only. This results in solving the equation

$$(\widehat{\lambda}_1^{-\delta} + \widehat{\lambda}_2^{-\delta})^{-\frac{1}{\delta}} = \widehat{\lambda}^{\|},$$

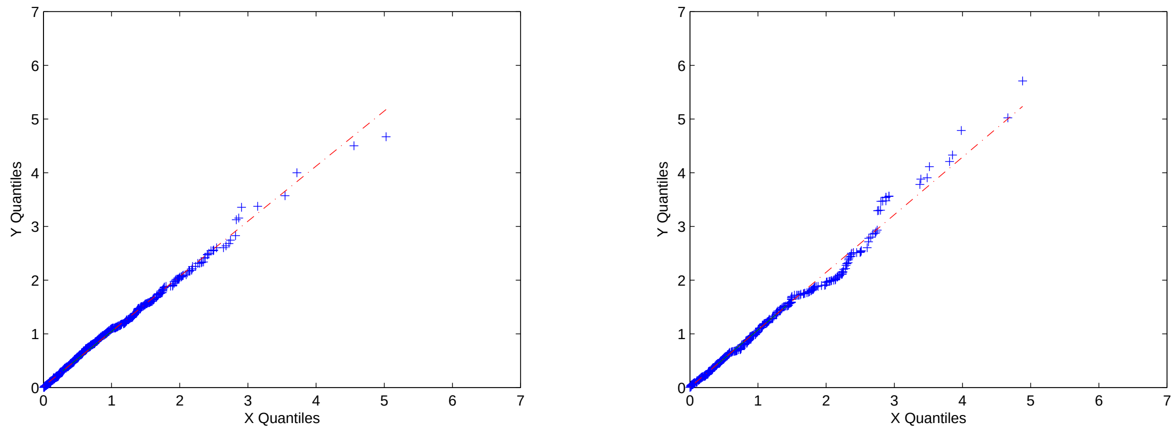

Figure 2.5: QQ-plot of the logarithmic Danish fire insurance data versus their estimated Weibull distributions, with parameters estimated from the data set. Left (loss of building), right (loss of content)

where $\widehat{\lambda}_1$, $\widehat{\lambda}_2$ and $\widehat{\lambda}^{\|}$ are the estimated intensities for each of the marginal univariate Poisson processes and the jump dependent part of the process. We obtain $\widehat{\delta} = 1.0546$.

Second, we should compare this estimator with the corresponding estimator based on the jump sizes. For this we invoke Example 2.4.4, which shows that the Clayton Lévy copula for a bivariate CPP implies a distributional Clayton copula for the joint jump sizes of the process. The maximum likelihood estimator of the parameter δ based on the joint jumps is obtained as $\widehat{\delta} = 0.8675$. This is close enough to convince us that a bivariate compound Poisson process is a good model for the Danish fire insurance data, and that the Clayton Lévy copula is an appropriate model.

Now we consider the full likelihood as given in equation (2.4.2), with two-parametric marginal Weibull distributions for the log-sizes of the claims and a Clayton Lévy copula $\mathfrak{C}$. Furthermore, we denote by λ_1 and λ_2, the intensities of losses in each component. Then the full likelihood including seven parameters is given by equation (2.4.5).

The resulting maximum likelihood estimates of the parameters are as follows.

Parameters	λ_1	λ_2	a_1	b_1	a_2	b_2	δ
Estimates	76.5643	44.7933	0.8302	1.1308	1.0898	1.0805	0.9531

From this table it can be seen that the estimator of the Lévy copula parameter $\widehat{\delta} = 0.9531$ for a Weibull-Clayton model is between the estimator only based on point processes and the estimator only based on joint jumps of the process This is as expected.

2.7 Conclusion

We have suggested a maximum likelihood estimation procedure for a multivariate compound Poisson process, which guarantees that the estimated model is again a compound Poisson process. This is achieved on the basis of Sklar's theorem for Lévy copulas by a detailed analysis of the dependence structure. We have also suggested a new simulation algorithm for a multivariate compound Poisson process. A small simulation study has shown that the estimation procedure works well also for small sample sizes. For the Danish fire insurance data, after some explorative data analysis to find a convincing model, we have fitted a seven parameter compound Poisson process model. The use of a Lévy copula approach for the dependence modelling has proved extremely useful in this context.

Chapter 3

Parametric estimation of a bivariate stable Lévy process

SUMARRY

We propose a parametric model for a bivariate stable Lévy process based on a Lévy copula as a dependence model. We estimate the parameters of the full bivariate model by maximum likelihood estimation. As an observation scheme we assume that we observe all jumps larger than some $\varepsilon > 0$ and base our statistical analysis on the resulting compound Poisson process. We derive the Fisher information matrix and prove asymptotic normality of all estimates when the truncation point ε tends to zero. A simulation study investigates the loss of efficiency because of the truncation.

3.1 Introduction

The problem of parameter estimation of one-dimensional stable Lévy processes has been investigated already in the seventies of the last century by Basawa and Brockwell [4, 5]. Starting with a subordinator model, they assumed that it is possible to observe $n^{(\varepsilon)}$ jumps in a time interval $[0, t]$, all larger than a certain small $\varepsilon > 0$. Based on this observation scheme, they estimated the parameters by a maximum likelihood procedure, and investigated the distributional limits of the MLEs for $n^{(\varepsilon)} \to \infty$, which in this model happens by $t \to \infty$ and/or $\varepsilon \to 0$.

The task of estimating multivariate stable processes is usually solved by estimating the parameters of the marginal processes and the spectral measure separately; cf. Nolan, Panorska and McCulloch [28] and Höpfner [21] and references therein.

The rather recent modelling of multivariate Lévy processes by their marginal processes and a Lévy copula for the dependence structure (cf. Cont and Tankov [13], Kallsen and Tankov [25], and Eder and Klüppelberg [15]) allows for the construction of new parametric models. This approach is similar to the representation of a multivariate distribution function by its marginal distributions and a copula and is valid for all multivariate Lévy processes.

Moreover, various estimation methods of the parameters of the marginal processes and the dependence structure either together or separately can be applied. Obviously, it is more efficient to estimate all parameters of a model in one go, but often the attempt fails. Problems may occur because of the complexity of the numerical optimization involved to obtain the MLEs of the parameters or, given the estimates, their asymptotic properties are not clear concerning their asymptotic covariance structure.

This is an important point in the context of Lévy processes, since these properties may depend on the observation scheme. In reality it is usually not possible to observe the continuous-time sample path, but it may be possible to observe all jumps larger than ε as in the one-dimensional problem studied by Basawa and Brockwell [5]. For a stable subordinator we obtain asymptotic normality for such an observation scheme, provided that $n^{(\varepsilon)} \to \infty$ or equivalently $\varepsilon \to 0$.

With this paper we want to start an investigation concerning statistical estimation of multivariate Lévy processes in a parametric framework. In a certain sense the present paper is a follow-up of Esmaeili and Klüppelberg [17], where we concentrated on parametric estimation of multivariate compound Poisson processes.

Since our observation scheme involves only jumps larger than ε, the observed process is a multivariate compound Poisson process. But in contrast to [17], we now assume that the Lévy process has infinite Lévy measure and we investigate asymptotic normality also for $\varepsilon \to 0$.

Our paper is organised as follows. In Section 3.2 we present some basic facts about Lévy copulas and recall the estimation procedure as presented in Basawa and Brockwell [4, 5] for one-dimensional α-stable subordinators in Section 3.3. Section 3.4 contains the theoretical body of our new results. In Section 3.4.1 we present the small jumps truncation and its consequences for the Lévy copula; Section 3.4.2 presents the maximum likelihood estimation for the α-stable Clayton subordinator, including an explicit calculation of the Fisher information matrix, which ensures joint asymptotic normality of all estimates. Section 3.5 presents a simulation study and Section 3.6 concludes and gives an outlook to further work.

3.2 Lévy processes and Lévy copulas

Let $\mathbf{S} = (\mathbf{S}(t))_{t\geq 0}$ be a Lévy process with values in $\mathbb{R}^d$ defined on a filtered probability space $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathcal{F}, \mathcal{P})$; i.e $\mathbf{S}$ has independent and stationary increments, and we assume that it has cádlág sample paths. For each $t > 0$, the random variable $\mathbf{S}(t)$ has an infinitely divisible distribution, whose characteristic function has a Lévy-Khintchine representation:

$$\mathbb{E}[e^{i(z,\mathbf{S}(t))}] = \exp\left\{t\left(i(\gamma, z) - \frac{1}{2}z^\top Az + \int_{\mathbb{R}^d}\left(e^{i(z,x)} - 1 - i(z,x)1_{|x|\leq 1}\right)\Pi(dx)\right)\right\},\ z \in \mathbb{R}^d,$$

where $(\cdot,\cdot)$ denotes the inner product in $\mathbb{R}^d$, $\gamma \in \mathbb{R}^d$ and A is a symmetric nonnegative definite $d \times d$ matrix. The *Lévy measure* Π is a measure on $\mathbb{R}^d$ satisfying $\Pi(\{\mathbf{0}\}) = 0$ and $\int_{\mathbb{R}^d\setminus\{\mathbf{0}\}} \min\{1, |x|^2\}\Pi(dx) < \infty$. For every Lévy process its distribution is defined by (γ, A, Π), which is called the *characteristic triplet*. It is worth mentioning that the Lévy measure $\Pi(B)$ for $B \in \mathcal{B}(\mathbb{R}^d)$ is the expected number of jumps per unit time with size in B.

Brownian motion is characterised by $(0, A, 0)$ and Brownian motion with drift by $(\gamma, A, 0)$. Poisson processes and compound Poisson processes have characteristic triplet $(\gamma_1, 0, \Pi)$. The class of Lévy processes is very rich including prominent examples like stable processes, gamma processes, variance gamma processes, inverse Gaussian and normal inverse Gaussian processes. Their applications reach from finance and insurance applications to the natural sciences and engineering. A particular role is played by *subordinators*, which are Lévy processes with increasing sample paths. Other important classes are *spectrally one-sided Lévy processes*, which have only positive or only negative jumps.

We are concerned with dependence in the jump behaviour $\mathbf{S}$, which we model by an appropriate functional of the marginals of the Lévy measure Π. Since, with the exception of a compound Poisson model, all Lévy measures have a singularity in 0, we follow Cont and Tankov [13] and introduce a (survival) copula on the *tail integral*, which is called *Lévy copula* and, because of the singularity in 0, is defined for each quadrant separately; for details we refer to Kallsen and Tankov [25] and to Eder and Klüppelberg [15] for a different approach.

Throughout this paper we restrict the presentation to the positive cone $\mathbb{R}_+^d$, where only common positive jumps in all component processes happen. To extend this theory to general Lévy processes is not difficult, but notationally involved.

We present the definition of the tail integral on the positive cone $\mathbb{R}_+^d$. For a spectrally positive Lévy process this characterises the jump behaviour completely.

Definition 3.2.1. *Let* Π *be a Lévy measure on* $\mathbb{R}_+^d$. *The* tail integral *is a function* $\overline{\Pi} : [0,\infty]^d \to [0,\infty]$ *defined by*

$$\overline{\Pi}(x_1,\dots,x_d) = \begin{cases} \Pi([x_1,\infty)\times\cdots\times[x_d,\infty))\,, & (x_1,\dots,x_d)\in[0,\infty)^d\setminus\{\mathbf{0}\} \\ 0\,, & x_i=\infty \text{ for at least one } i \\ \infty\,, & (x_1,\dots,x_d)=\mathbf{0}. \end{cases}$$

The marginal tail integrals are defined analogously for $i=1,\dots,d$ *as* $\overline{\Pi}_i(x)=\Pi_i([x,\infty))$ *for* $x\geq 0$.

Also the Lévy copula is defined quadrantwise and the following characterises the dependence structure of a spectrally positive Lévy process completely.

Definition 3.2.2. *A* positive Lévy copula *is a* d*-increasing grounded function* $\mathfrak{C} : [0,\infty]^d \to [0,\infty]$ *with margins* $\mathfrak{C}_k(u)=u$ *for all* $u\in[0,\infty]$ *and* $k=1,\dots,d$.

The following theorem is a version of Sklar's theorem for Lévy processes with positive jumps, proved in Tankov [34], Theorem 3.1; for the corresponding result for general Lévy processs we refer again to Kallsen and Tankov [25].

Theorem 3.2.3 (Sklar's Theorem for Lévy copulas)**.** *Let $\overline{\Pi}$ denote the tail integral of a spectrally positive d-dimensional Lévy process, whose components have Lévy measures $\Pi_1, \ldots, \Pi_d$. Then there exists a Lévy copula $\mathfrak{C} : [0,\infty]^d \to [0,\infty]$ such that for all $x_1, x_2, \ldots, x_d \in [0,\infty]$*

$$\overline{\Pi}(x_1, \ldots, x_d) = \mathfrak{C}(\overline{\Pi}_1(x_1), \ldots, \overline{\Pi}_d(x_d)). \tag{3.2.1}$$

If the marginal tail integrals are continuous, then this Lévy copula is unique. Otherwise, it is unique on $Ran\overline{\Pi}_1 \times \cdots \times Ran\overline{\Pi}_d$.
Conversely, if $\mathfrak{C}$ is a Lévy copula and $\overline{\Pi}_1, \ldots, \overline{\Pi}_d$ are marginal tail integrals of a spectrally positive Lévy process, then the relation (3.2.1) *defines the tail integral of a d-dimensional spectrally positive Lévy process and $\overline{\Pi}_1, \ldots, \overline{\Pi}_d$ are tail integrals of its components.*

Remark 3.2.4. *In the case of multivariate stable Lévy processes the Lévy copula carries the same information as the spectral measure. By choosing a slightly different approach this was shown in Eder and Klüppelberg [15]. Note, however, that the spectral measure restricts to stable processes, whereas the Lévy copula models the dependence for all Lévy processes.*

We are concerned with the estimation of the parameters of a multivariate Lévy process and assume for simplicity that we observe all jumps larger than $\varepsilon > 0$ of a subordinator. This results in a compound Poisson process and we recall the following well-known results; see e.g. Sato [32], Theorem 21.2 and Corollary 8.8.

Proposition 3.2.5. *A pure jump Lévy process $\mathbf{S}$ in $\mathbb{R}^d$ is compound Poisson if and only if it has a finite Lévy measure Π with $\lim_{\boldsymbol{x}\to\mathbf{0}} \overline{\Pi}(\boldsymbol{x}) = \lambda$, the intensity of the d-dimensional Poisson process, and jump distribution $F(d\boldsymbol{x}) = \lambda^{-1}\Pi(d\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^d$.*

3.3 Maximum likelihood estimation of the parameters of a one-dimensional stable Lévy process

3.3.1 Small jumps truncation

With the understanding that the Lévy measure can be decomposed in positive and negative jumps we restrict ourselves to subordinators.

Let S be a one-dimensional subordinator with unbounded Lévy measure Π, without drift or Gaussian part. For all $t \geq 0$ its characteristic function has the representation $Ee^{iuS(t)} = e^{t\psi(u)}$ for $u \in \mathbb{R}$ with

$$\psi(u) = \int_{0<x<\varepsilon} \big(e^{iux} - 1\big)\Pi(dx) + \int_{x\geq\varepsilon} \big(e^{iux} - 1\big)\Pi(dx), \quad u \in \mathbb{R}\,, \tag{3.3.1}$$

for arbitrary $\varepsilon > 0$. The last integral in (3.3.1) is the characteristic exponent of a compound Poisson process with Poisson intensity $\lambda^{(\varepsilon)} \in (0, \infty)$ and jump distribution function $F^{(\varepsilon)}$

$$\lambda^{(\varepsilon)} = \int_{\varepsilon}^{\infty} \Pi(dx) \quad \text{and} \quad F^{(\varepsilon)}(dx) = \Pi(dx)/\lambda^{(\varepsilon)} \text{ on } [\varepsilon, \infty).$$

As an observation scheme we assume that we observe the whole sample path of S over a time interval $[0, t]$, but that we only observe jumps of size larger than ε. Then our observation scheme is equivalent to observing a compound Poisson process, say $S^{(\varepsilon)}$, given in its marked point process representation as $\{(T_k^{(\varepsilon)}, X_k^{(\varepsilon)}), k = 1, \ldots, n^{(\varepsilon)}\}$, where $n^{(\varepsilon)} = n^{(\varepsilon)}(t) = \text{card}\{T_k^{(\varepsilon)} \in [0, t] : k \in \mathbb{N}\}$. We also assume that $\Pi(dx) = \nu(x; \theta)dx$ where θ is a vector of parameters of the Lévy measure so that the density of $X_k^{(\varepsilon)}$ is given by $f^{(\varepsilon)}(x; \theta) = \nu(x, \theta)/\lambda^{(\varepsilon)}$ for $x \geq \varepsilon$. The likelihood function of this compound Poisson process is well-known, see e.g. Basawa and Prakasa Rao [6], and is given by

$$L^{(\varepsilon)}(\theta) = (\lambda^{(\varepsilon)})^{n^{(\varepsilon)}} e^{-\lambda^{(\varepsilon)} t} \times \prod_{i=1}^{n^{(\varepsilon)}} f^{(\varepsilon)}(x_i, \theta) = e^{-\lambda^{(\varepsilon)} t} \times \prod_{i=1}^{n^{(\varepsilon)}} \nu(x_i; \theta) \mathbf{1}_{\{x_i \geq \varepsilon\}}. \tag{3.3.2}$$

3.3.2 Asymptotic behaviour of the MLEs

MLE is a well established estimation procedure and the asymptotic properties of the estimators is well-known for i.i.d. data, but also for continuous-time stochastic processes, see e.g. Küchler and Sorensen [26] and references therein. However, this theory is usually concerned about letting the observation time, i.e. t tends to infinity. We are more interested in the case of fixed t and $\varepsilon \downarrow 0$, and here there exist to our knowledge only some specific results in the literature; see e.g. Basawa and Brockwell [4, 5] and Höpfner and Jacod [22].

We start with a general Lévy process S and base the maximum likelihood estimation on the jumps $\Delta S_v > \varepsilon$ for $v \in [0, t]$. The MLEs are, in fact, those obtained from the CPP $S^{(\varepsilon)}$ as described in Section 3.3.1 above. Therefore, under some regularity conditions (see e.g. Prakasa Rao [29], Section 3.11) the MLEs are consistent and asymptotically normal. In the context of a compound Poisson process the asymptotic behavior of estimators is considered for $t \to \infty$. In our set-up, however, it is also relevant to consider the performance of estimators as $\varepsilon \to 0$ with t fixed.

We investigate the asymptotic behavior of estimators for a stable Lévy process as $n^{(\varepsilon)} \to \infty$ and shall show that this covers the cases of $t \to \infty$ as well as $\varepsilon \to 0$. Asymptotic normality of the estimators has been derived in Basawa and Brockwell [4, 5]. For comparison and later reference we summarize these results in some detail.

Example 3.3.1. *[α-stable subordinator]*
Let $(S(t))_{t\geq 0}$ be a one dimensional α-stable subordinator with parameters $c > 0$ and $0 < \alpha < 1$, such that the tail integral is given by $\overline{\Pi}(x) = cx^{-\alpha}$ for $x > 0$. Observing all jumps larger than some $\varepsilon > 0$, the resulting CPP has intensity and jump size density

$$\lambda^{(\varepsilon)} = \int_\varepsilon^\infty \Pi(dx) = c\varepsilon^{-\alpha} \qquad , \qquad f^{(\varepsilon)}(x) = \frac{\Pi(dx)/dx}{\lambda^{(\varepsilon)}} = \alpha\varepsilon^\alpha x^{-1-\alpha}, \quad x > \varepsilon.$$

If we observe $n^{(\varepsilon)}$ jumps larger than ε in $[0, t]$, we estimate the intensity by $\widehat{\lambda}^{(\varepsilon)} = \frac{n^{(\varepsilon)}}{t}$. Moreover, by (3.3.2) the loglikelihood function for $\theta = (\alpha, \log c)$ is given by

$$\ell(\alpha, c) = n^{(\varepsilon)}(\log\alpha + \log c) - e^{\log c}\varepsilon^{-\alpha}t - (1+\alpha)\sum_{i=1}^{n^{(\varepsilon)}} \log x_i\,.$$

We calculate the score functions as

$$\begin{aligned}\frac{\partial \ell}{\partial \alpha} &= \frac{n^{(\varepsilon)}}{\alpha} + te^{\log c}\varepsilon^{-\alpha}\log\varepsilon - \sum_{i=1}^{n^{(\varepsilon)}} \log x_i\,,\\ \frac{\partial \ell}{\partial \log c} &= n^{(\varepsilon)} - e^{\log c}\varepsilon^{-\alpha}t\end{aligned}$$

To obtain candidates for maxima we calculate

$$
\begin{aligned}
e^{\log c} &= \frac{n^{(\varepsilon)}}{t\varepsilon^{-\alpha}} = \frac{\widehat{\lambda}^{(\varepsilon)}}{\varepsilon^{-\alpha}}\,, \\
\frac{1}{\alpha} &= -\frac{t}{n^{(\varepsilon)}} c\varepsilon^{-\alpha}\log\varepsilon + \frac{1}{n^{(\varepsilon)}}\sum_{i=1}^{n^{(\varepsilon)}}\log x_i \\
&= \frac{1}{n^{(\varepsilon)}}\sum_{i=1}^{n^{(\varepsilon)}}(\log x_i - \log\varepsilon) + \log\varepsilon\Big(1 - \frac{\lambda^{(\varepsilon)}}{\widehat{\lambda}^{(\varepsilon)}}\Big)\,.
\end{aligned}
$$

Consequently, we have the maximum likelihood estimators

$$
\begin{aligned}
\widehat{\alpha} &= \left(\frac{1}{n^{(\varepsilon)}}\sum_{i=1}^{n^{(\varepsilon)}}\Big(\log X_i^{(\varepsilon)} - \log\varepsilon\Big) + \log\varepsilon\Big(1 - \frac{\lambda^{(\varepsilon)}}{\widehat{\lambda}^{(\varepsilon)}}\Big)\right)^{-1}\,, \\
\widehat{\log c} &= \log\widehat{\lambda}^{(\varepsilon)} + \widehat{\alpha}\log\varepsilon\,.
\end{aligned}
$$

Next we calculate the second derivatives as

$$
\begin{aligned}
\frac{\partial^2\ell}{\partial\alpha^2} &= -n^{(\varepsilon)}\frac{1}{\alpha^2} - ct\varepsilon^{-\alpha}(\log\varepsilon)^2 = -tc\varepsilon^{-\alpha}\left(\frac{\widehat{\lambda}^{(\varepsilon)}}{\lambda^{(\varepsilon)}}\frac{1}{\alpha^2} + (\log\varepsilon)^2\right) \\
\frac{\partial^2\ell}{\partial\alpha\,\partial\log c} &= ct\varepsilon^{-\alpha}\log\varepsilon = \frac{\partial^2\ell}{\partial\log c\partial\alpha} \\
\frac{\partial^2\ell}{\partial(\log c)^2} &= -tc\varepsilon^{-\alpha}\,.
\end{aligned}
$$

Consequently, the Fisher information matrix is given by

$$
I^{(\varepsilon)}_{\alpha,\log c} = tc\varepsilon^{-\alpha}\begin{pmatrix} \frac{1}{\alpha^2} + (\log\varepsilon)^2 & -\log\varepsilon \\ -\log\varepsilon & 1 \end{pmatrix}.
$$

We calculate the determinant as $\det(I^{(\varepsilon)}_{\alpha,\log c}) = c^2t^2\alpha^{-2}\varepsilon^{-2\alpha}$*. Using Cramer's rule of inversion easily gives*

$$
(I^{(\varepsilon)}_{\alpha,\log c})^{-1} = (ct)^{-1}\varepsilon^{\alpha}\alpha^2\begin{pmatrix} 1 & \log\varepsilon \\ \log\varepsilon & \frac{1}{\alpha^2} + (\log\varepsilon)^2 \end{pmatrix}.
$$

We are interested in the asymptotic behaviour of the MLEs $\widehat{\alpha}$ *and* $\widehat{\log c}$ *based on a fixed time interval* $[0,t]$ *and letting* $\varepsilon \to 0$*. Note that we have to get the variance-covariance* *matrix asymptotically independent of* ε. *Division of* $\widehat{\log c}$ *by* $\log\varepsilon$ *changes the matrix* $I^{(\varepsilon)^{-1}}_{\alpha,\log c}$ *into*

$$(\widetilde{I}^{(\varepsilon)}_{\alpha,\frac{\log c}{\log\varepsilon}})^{-1} = t^{-1}c^{-1}\varepsilon^{\alpha}\alpha^2 \begin{pmatrix} 1 & 1 \\ 1 & \frac{1}{\alpha^2(\log\varepsilon)^2}+1 \end{pmatrix}.$$

Since

$$\sqrt{n^{(\varepsilon)}}\left(\frac{\widehat{\lambda}^{(\varepsilon)}}{\lambda^{(\varepsilon)}}-1\right) = \sqrt{n^{(\varepsilon)}}\left(\frac{n^{(\varepsilon)}}{tc\varepsilon^{-\alpha}}-1\right) \xrightarrow{d} N(0,1)\,,\quad n^{(\varepsilon)}\to\infty\,, \tag{3.3.3}$$

and the regularity conditions of Section 3.11 of Prakasa Rao [29] are satisfied, classical likelihood theory ensures that

$$\sqrt{n^{(\varepsilon)}}\begin{pmatrix} \widehat{\alpha}-\alpha \\ \dfrac{\log\widehat{c}-\log c}{\log\varepsilon} \end{pmatrix} \sim \mathrm{AN}\left(\mathbf{0},\alpha^2\begin{pmatrix} 1 & 1 \\ 1 & \frac{1}{\alpha^2}\frac{1}{(\log\varepsilon)^2}+1 \end{pmatrix}\right)\,,\quad n^{(\varepsilon)}\to\infty\,.$$

Consistency of $\widehat{\lambda}^{(\varepsilon)}$, *obtained from* (3.3.3), *and a Taylor expansion of* $\log x$ *around* c *ensures with Slutzky's theorem that for* $\varepsilon\to 0$,

$$\sqrt{ct}\varepsilon^{-\alpha/2}\begin{pmatrix} \dfrac{\widehat{\alpha}}{\alpha}-1 \\ \frac{1}{\alpha\log\varepsilon}\left(\dfrac{\widehat{c}}{c}-1\right) \end{pmatrix} \xrightarrow{d} N\left(\mathbf{0},\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right) = \begin{pmatrix} N_1 \\ N_2 \end{pmatrix},$$

where N_1, N_2 *are standard normal random variables with* $\mathrm{cov}(N_1,N_2)=1$, *which implies that* $N_1=N_2=N$. *So the limit law is degenerate.*

It has been shown in Jacod and Höpfner [22] that the natural parameterization is not (c,α), *but* $(\lambda^{(\varepsilon)},\alpha)$, *which leads to asymptotically independent normal limits. Indeed, we have*

$$\sqrt{n^{(\varepsilon)}}\begin{pmatrix} \dfrac{\widehat{\alpha}}{\alpha}-1 \\ \dfrac{\widehat{\lambda}^{(\varepsilon)}}{\lambda^{(\varepsilon)}}-1 \end{pmatrix} \xrightarrow{d} N\left(\mathbf{0},\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right),\quad n^{(\varepsilon)}\to\infty\,,$$

where $n^{(\varepsilon)}$ *can again be replaced by* $tc\varepsilon^{-\alpha}$ *and the same result holds for* $t\to\infty$, *equivalently,* $\varepsilon\to 0$. □

3.4 Maximum likelihood estimation of the parameters of a bivariate Lévy process

3.4.1 Small jumps truncation

Let $\mathbf{S}$ be a bivariate Lévy process with unbounded Lévy measure Π in both components and marginal Lévy measures Π_1 and Π_2 corresponding to the components S_1 and S_2, respectively. It has an infinite number of jumps in the observation interval $[0,t]$. Several observation schemes are possible here concerning the truncation of the small jumps.

We consider only jumps (x,y), where both $x \geq \varepsilon$ and $y \geq \varepsilon$ at the same time. This leads to a bivariate compound Poisson model with joint jumps larger than ε.

Consider the truncated process $\mathbf{S}^{(\varepsilon)}$ with total Lévy measure

$$\Pi^{(\varepsilon)}(\mathbb{R}_+^2) = \Pi\{(x,y) \in \mathbb{R}_+^2 : x \geq \varepsilon, y \geq \varepsilon\} =: \lambda^{(\varepsilon)} < \infty \,.$$

Then there exists a representation

$$\mathbf{S}^{(\varepsilon)}(t) = \int_0^t \int_{\boldsymbol{x} \geq \varepsilon} \boldsymbol{x} M(ds \times d\boldsymbol{x}) = \sum_{i=1}^{N(t)} \mathbf{X}_i \,, \quad t \geq 0 \,,$$

where $\geq$ is taken componentwise and M is a Poisson random measure, which has support $[0,\infty) \times [\varepsilon,\infty)^2$ with intensity measure $ds\Pi^{(\varepsilon)}(d\boldsymbol{x})$ on its support; cf. Sato [32], Theorem 19.2. This means that $\mathbf{S}^{(\varepsilon)}$ is a compound Poisson process with intensity $\lambda^{(\varepsilon)}$ and leads to the observation scheme as described in Section 4 of Esmaeili and Klüppelberg [17] in detail, where now all jumps are larger than ε in both components. We now investigate the influence of the truncation on the Lévy copula (see Figure 3.1).

Lemma 3.4.1. *Let $\mathbf{S}$ be a bivariate Lévy process with continuous marginal tail integrals. Assume that the unbounded Lévy measure Π concentrates on $\mathbb{R}_+^2$ with Lévy copula $\mathfrak{C}$, which is different from the independent Lévy copula. Consider only those jumps, which are larger than ε in both component processes.Then the Lévy copula of the resulting CPP is given by*

$$\widetilde{\mathfrak{C}}^{(\varepsilon)}(u,v) = \mathfrak{C}(\mathfrak{C}_1^{\leftarrow}(u,\lambda_2^{(\varepsilon)}), \mathfrak{C}_2^{\leftarrow}(\lambda_1^{(\varepsilon)},v)), \quad 0 < u,v \leq \lambda^{(\varepsilon)}. \qquad (3.4.1)$$

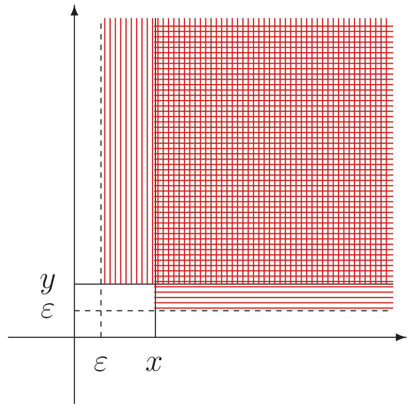

Figure 3.1: Illustration of the support of the bivariate tail integral $\overline{\Pi}^{(\varepsilon)}(x,y)$ and the marginal tail integrals $\overline{\Pi}_1^{(\varepsilon)}(x)$ and $\overline{\Pi}_2^{(\varepsilon)}(x)$.

where $\mathfrak{C}_k^{\leftarrow}$, $k=1,2$ *is the inverse of* $\mathfrak{C}$ *with respect to the* k*-th argument,* $\lambda_k^{(\varepsilon)} = \overline{\Pi}_k(\varepsilon)$, $k = 1,2$, *and* $\lambda^{(\varepsilon)} = \overline{\Pi}(\varepsilon,\varepsilon)$.

Proof. If the Lévy copula is the independent Lévy copula, then there is only mass on the axes and there are a.s. no jumps in both components at the same time. So, assume that the Lévy copula is different from the independent Lévy copula. The marginal tail integrals of the CPP are given by

$$\begin{aligned} \overline{\Pi}_1^{(\varepsilon)}(x) &= \overline{\Pi}(x,\varepsilon) = \mathfrak{C}(\overline{\Pi}_1(x),\overline{\Pi}_2(\varepsilon)) = \mathfrak{C}(\overline{\Pi}_1(x),\lambda_2^{(\varepsilon)}), \quad x>\varepsilon, \\ \overline{\Pi}_2^{(\varepsilon)}(y) &= \overline{\Pi}(\varepsilon,y) = \mathfrak{C}(\overline{\Pi}_1(\varepsilon),\overline{\Pi}_2(y)) = \mathfrak{C}(\lambda_1^{(\varepsilon)},\overline{\Pi}_2(y)), \quad y>\varepsilon, \end{aligned} \tag{3.4.2}$$

whereas the bivariate tail integral is

$$\overline{\Pi}^{(\varepsilon)}(x,y) = \overline{\Pi}(x,y) = \mathfrak{C}(\overline{\Pi}_1(x),\overline{\Pi}_2(y)), \quad x,y>\varepsilon. \tag{3.4.3}$$

Denote by $\widetilde{\mathfrak{C}}^{(\varepsilon)}$ the Lévy copula of the CPP, and from (3.4.3) we have

$$\widetilde{\mathfrak{C}}^{(\varepsilon)}(\overline{\Pi}_1^{(\varepsilon)}(x),\overline{\Pi}_2^{(\varepsilon)}(y)) = \mathfrak{C}(\overline{\Pi}_1(x),\overline{\Pi}_2(y)), \quad x,y>\varepsilon.$$

Together with (3.4.2) this implies that

$$\widetilde{\mathfrak{C}}^{(\varepsilon)}(\mathfrak{C}(\overline{\Pi}_1(x),\lambda_2^{(\varepsilon)}),\mathfrak{C}(\lambda_1^{(\varepsilon)},\overline{\Pi}_2(y))) = \mathfrak{C}(\overline{\Pi}_1(x),\overline{\Pi}_2(y)), \quad x,y>\varepsilon.$$

Setting $u := \mathfrak{C}(\overline{\Pi}_1(x), \lambda_2^{(\varepsilon)})$ and $v := \mathfrak{C}(\lambda_1^{(\varepsilon)}, \overline{\Pi}_2(y))$, we see that for $x, y > \varepsilon$

$$\overline{\Pi}_1(x) = \mathfrak{C}_1^{\leftarrow}(u, \lambda_2^{(\varepsilon)}) \quad \text{and} \quad \overline{\Pi}_2(y) = \mathfrak{C}_2^{\leftarrow}(\lambda_1^{(\varepsilon)}, v),$$

and, hence, for $0 < u, v \leq \lambda^{(\varepsilon)}$

$$\widetilde{\mathfrak{C}}^{(\varepsilon)}(u, v) = \mathfrak{C}\left(\mathfrak{C}_1^{\leftarrow}(u, \lambda_2^{(\varepsilon)}), \mathfrak{C}_2^{\leftarrow}(\lambda_1^{(\varepsilon)}, v)\right).$$

□

Proposition 3.4.2. *Assume that the conditions of Lemma 3.4.1 hold and that Lévy copula $\mathfrak{C}$ is continuous on $[0,\infty]^2$. Then*

$$\lim_{\varepsilon \to 0} \widetilde{\mathfrak{C}}^{(\varepsilon)}(u, v) = \mathfrak{C}(u, v), \quad u, v > 0.$$

Proof. Take arbitrary $u, v > 0$. Then there exists some $\varepsilon > 0$ such that $0 < u, v \leq \lambda^{(\varepsilon)}$. Invoking the Lipschitz condition for Lévy copula (Theorem 2.1, Barndorff-Nielsen and Lindner [3])and (3.4.1), we have

$$\begin{aligned} |\widetilde{\mathfrak{C}}^{(\varepsilon)}(u, v) - \mathfrak{C}(u, v)| &= \left|\mathfrak{C}\left(\mathfrak{C}_1^{\leftarrow}(u, \lambda_2^{(\varepsilon)}), \mathfrak{C}_2^{\leftarrow}(\lambda_1^{(\varepsilon)}, v)\right) - \mathfrak{C}(u, v)\right| \\ &\leq \left|\mathfrak{C}_1^{\leftarrow}(u, \lambda_2^{(\varepsilon)}) - u\right| + \left|\mathfrak{C}_2^{\leftarrow}(\lambda_1^{(\varepsilon)}, v) - v\right|. \end{aligned}$$

Since the Lévy copula $\mathfrak{C}$ has Lebesgue margins, i.e. $\mathfrak{C}(u, \infty) = u$ and $\mathfrak{C}(\infty, v) = v$, we have $\mathfrak{C}_1^{\leftarrow}(u, \infty) = u$ and $\mathfrak{C}_2^{\leftarrow}(\infty, v) = v$. This implies that

$$|\widetilde{\mathfrak{C}}^{(\varepsilon)}(u, v) - \mathfrak{C}(u, v)| \leq \left|\mathfrak{C}_1^{\leftarrow}(u, \lambda_2^{(\varepsilon)}) - \mathfrak{C}_1^{\leftarrow}(u, \infty)\right| + \left|\mathfrak{C}_2^{\leftarrow}(\lambda_1^{(\varepsilon)}, v) - \mathfrak{C}_2^{\leftarrow}(\infty, v)\right|.$$

The terms on the rhs tend to zero because the Lévy measure is unbounded and $\lim_{\varepsilon \to 0} \lambda_1^{(\varepsilon)} = \lim_{\varepsilon \to 0} \lambda_2^{(\varepsilon)} = \infty$. □

Now we proceed as in Esmaeili and Klüppelberg [17] and use the same notation. Denote by $(x_1, y_1), \ldots, (x_{n^{(\varepsilon)}}, y_{n^{(\varepsilon)}})$ the observed jumps larger than ε in both components, i.e. occurring at the same time during the observation interval $[0, t]$. Assume further that the dependence structure of the process $\mathbf{S} = (S_1, S_2)$ is defined by a Lévy copula $\mathfrak{C}$ with a parameter vector δ. We also assume that γ_1 and γ_2 are the parameter vectors of the marginal Lévy measures Π_1 and Π_2.

Using the notation $\nu_k(\cdot) = \lambda_k^{(\varepsilon)} f_k^{(\varepsilon)}(\cdot)$ for the marginal Lévy densities on (ε, ∞) for $k = 1, 2$ we can reformulate Theorem 4.1 of Esmaeili and Klüppelberg [17] as follows.

Theorem 3.4.3. *Assume an observation scheme as above for a bivariate Lévy process with only non-negative jumps. Assume that γ_1 and γ_2 are the parameters of the marginal Lévy measures Π_1 and Π_2 with Lévy densities ν_1 and ν_2, respectively, and a Lévy copula $\mathfrak{C}$ with parameter vector δ. Assume further that $\frac{\partial^2}{\partial u \partial v}\mathfrak{C}(u, v; \delta)$ exists for all $(u, v) \in (0, \infty)^2$, which is the domain of $\mathfrak{C}$. Then the full likelihood of the bivariate CPP is given by*

$$L^{(\varepsilon)}(\gamma_1, \gamma_2, \delta) = e^{-\lambda^{(\varepsilon)} t} \prod_{i=1}^{n^{(\varepsilon)}} \left[\nu_1(x_i; \gamma_1)\nu_2(y_i; \gamma_2) \frac{\partial^2}{\partial u \partial v} \mathfrak{C}(u, v; \delta) \Big|_{\substack{u=\overline{\Pi}_1(x_i;\gamma_1), \\ v=\overline{\Pi}_2(y_i;\gamma_2)}} \right] \tag{3.4.4}$$

where

$$\lambda^{(\varepsilon)} = \int_\varepsilon^\infty \int_\varepsilon^\infty \Pi(dx, dy) = \mathfrak{C}(\overline{\Pi}_1(\varepsilon; \gamma_1), \overline{\Pi}_2(\varepsilon; \gamma_2); \delta).$$

3.4.2 Asymptotic behaviour of the MLEs of a bivariate stable Clayton model

A spectrally positive Lévy process is an α-stable subordinator if and only if $0 < \alpha < 1$ and there exists a finite measure $\tilde{\rho}$ on the unit sphere $\mathcal{S}^{d-1} := \{x \in \mathbb{R}_+^d \mid \|x\| = 1\}$ in $\mathbb{R}_+^d$ (for an arbitrary norm $\|\cdot\|$) such that the Lévy measure

$$\Pi(B) = \int_{\mathcal{S}^{d-1}} \tilde{\rho}(d\xi) \int_0^\infty 1_B(r\xi) \frac{dr}{r^{1+\alpha}}, \quad B \in \mathcal{B}(\mathbb{R}_+^d),$$

(cf. Theorem 14.3(ii) and Example 21.7 in Sato [32]).

From Kallsen and Tankov [25], Theorem 4.6, it is known that a bivariate process is α-stable if and only if it has α-stable marginal processes and a homogeneous Lévy copula of order 1; i.e. $\mathfrak{C}(tu, tv) = t\mathfrak{C}(u, v)$. The *Clayton Lévy copula*

$$\mathfrak{C}(u, v) = \left(u^{-\delta} + v^{-\delta}\right)^{-1/\delta}, \quad u, v > 0,$$

is homogeneous of order 1. Hence it is a valid model to define a bivariate α-stable process.

Suppose S_1 and S_2 are two α-stable subordinators with same tail integrals

$$\overline{\Pi}_k(x) = cx^{-\alpha}\,, \quad x > 0\,, \text{ for } k = 1, 2.$$

Assume further that $\mathbf{S} = (S_1, S_2)$ is a bivariate α-stable process with dependence structure modelled by a Clayton Lévy copula. The joint tail integral is then given by

$$\overline{\Pi}(x, y) \;=\; \mathfrak{C}(\overline{\Pi}_1(x), \overline{\Pi}_2(y)) = c\big(x^{\alpha\delta} + y^{\alpha\delta}\big)^{-\frac{1}{\delta}}\,, \quad x, y > 0\,. \tag{3.4.5}$$

The bivariate Lévy density is given by

$$\nu(x, y) \;=\; c(1+\delta)\alpha^2(xy)^{\alpha\delta-1}\big(x^{\alpha\delta} + y^{\alpha\delta}\big)^{-\frac{1}{\delta}-2}\,, \quad x, y > 0\,, \tag{3.4.6}$$

We assume the observation scheme as in Section 3.4.1. The Lévy measure Π will be considered on the set $[\varepsilon, \infty) \times [\varepsilon, \infty)$ with jump intensity

$$\lambda^{(\varepsilon)} = \overline{\Pi}(\varepsilon, \varepsilon) = c\left(\varepsilon^{\alpha\delta} + \varepsilon^{\alpha\delta}\right)^{-\frac{1}{\delta}} = c2^{-1/\delta}\varepsilon^{-\alpha}. \tag{3.4.7}$$

and marginal tail integrals

$$\overline{\Pi}_k^{(\varepsilon)}(x) = c(x^{\alpha\delta} + \varepsilon^{\alpha\delta})^{-1/\delta}\,, \quad k = 1, 2. \tag{3.4.8}$$

Moreover, for $k = 1, 2$,

$$\overline{\Pi}_k^{(\varepsilon)}(\varepsilon) = c2^{-1/\delta}\varepsilon^{-\alpha} = \lambda^{(\varepsilon)}\,,$$

and

$$\begin{aligned}\overline{G}_k^{(\varepsilon)}(x) &= \mathbb{P}(X > x) = \mathbb{P}(Y > x) = \frac{\overline{\Pi}_k^{(\varepsilon)}(x)}{\lambda^{(\varepsilon)}} \\ &= \left[\frac{1}{2}\Big(1 + \big(\frac{x}{\varepsilon}\big)^{\alpha\delta}\Big)\right]^{-1/\delta}\,, \quad x > \varepsilon\,.\end{aligned} \tag{3.4.9}$$

The Lévy copula of the CPP is by Lemma 3.4.1 given by

$$\begin{aligned}\widetilde{\mathfrak{C}}^{(\varepsilon)}(u, v) &= \mathfrak{C}\left(\mathfrak{C}_1^{\leftarrow}(u, \lambda_2^{(\varepsilon)}), \mathfrak{C}_2^{\leftarrow}(\lambda_1^{(\varepsilon)}, v)\right) \\ &= \mathfrak{C}\left(\left(u^{-\delta} - \lambda_2^{(\varepsilon)-\delta}\right)^{-1/\delta}, \left(v^{-\delta} - \lambda_1^{(\varepsilon)-\delta}\right)^{-1/\delta}\right) \\ &= \left(u^{-\delta} + v^{-\delta} - 2c^{-\delta}\varepsilon^{\alpha\delta}\right)^{-1/\delta}\end{aligned}$$

From the Lévy density in (3.4.6) the intensity in (3.4.7) the joint probability density of the bivariate jumps is given by

$$g^{(\varepsilon)}(x,y) = \alpha^2(1+\delta)\varepsilon^{\alpha}2^{\frac{1}{\delta}}(xy)^{\alpha\delta-1}(x^{\alpha\delta}+y^{\alpha\delta})^{-\frac{1}{\delta}-2}\,, \quad x,y>\varepsilon\,, \tag{3.4.10}$$

we note that our model is a bivariate generalized Pareto distribution (GPD); cf. Model I of Section 5.4 in Arnold et al. [2]. They present some properties of the model, and in our case X, Y are positively correlated.

We now turn to the MLE procedure. Noting that the parameterisation (c, α, δ) creates various problems taking derivatives, we propose a different choice of parameters. First we set $\alpha\delta = \theta$. Furthermore, recalling from the one-dimensional case that $\lambda^{(\varepsilon)}$ is a more natural choice than c, we decided to use the parameters $(\lambda^{(\varepsilon)}, \alpha, \theta)$. Recall from (3.3.2) for the bivariate CPP based on observations $(x_i, y_i) > \varepsilon$ for $i = 1, \dots, n^{(\varepsilon)}$,

$$\begin{aligned} L^{(\varepsilon)}(\lambda^{(\varepsilon)},\alpha,\theta) &= e^{-\lambda^{(\varepsilon)}t}\prod_{i=1}^{n^{(\varepsilon)}}\nu(x_i,y_i) \\ &= e^{-\lambda^{(\varepsilon)}t}(\lambda^{(\varepsilon)})^{n^{(\varepsilon)}}(\alpha(\alpha+\theta))^{n^{(\varepsilon)}}\varepsilon^{\alpha n^{(\varepsilon)}}2^{\frac{n^{(\varepsilon)}\alpha}{\theta}}\prod_{i=1}^{n^{(\varepsilon)}}\left[(x_iy_i)^{\theta-1}\left(x_i^{\theta}+y_i^{\theta}\right)^{-\frac{\alpha}{\theta}-2}\right]. \end{aligned}$$

Then the log-likelihood is given by

$$\begin{aligned} \ell^{(\varepsilon)}(\lambda^{(\varepsilon)},\alpha,\theta) &= -\lambda^{(\varepsilon)}t + n^{(\varepsilon)}\log\lambda^{(\varepsilon)} + n^{(\varepsilon)}(\log\alpha + \log(\alpha+\theta)) + \alpha n^{(\varepsilon)}\log\varepsilon + n^{(\varepsilon)}\frac{\alpha}{\theta}\log 2 \\ &\quad +(\theta-1)\sum_{i=1}^{n^{(\varepsilon)}}(\log x_i + \log y_i) - (2+\frac{\alpha}{\theta})\sum_{i=1}^{n^{(\varepsilon)}}\log(x_i^{\theta}+y_i^{\theta})\,. \end{aligned}$$

Note that the last term prevents the model to belong to an exponential family, so we have to be very careful concerning exchanging differentiation and integration. For the score functions we obtain

$$
\begin{aligned}
\frac{\partial \ell^{(\varepsilon)}}{\partial \lambda^{(\varepsilon)}} &= -t + \frac{n^{(\varepsilon)}}{\lambda^{(\varepsilon)}} \\
\frac{\partial \ell^{(\varepsilon)}}{\partial \alpha} &= \frac{n^{(\varepsilon)}}{\alpha} + \frac{n^{(\varepsilon)}}{\alpha+\theta} + n^{(\varepsilon)} \log \varepsilon + \frac{n^{(\varepsilon)} \log 2}{\theta} - \frac{1}{\theta} \sum_{i=1}^{n^{(\varepsilon)}} \log(x_i^\theta + y_i^\theta) \\
\frac{\partial \ell^{(\varepsilon)}}{\partial \theta} &= \frac{n^{(\varepsilon)}}{\alpha+\theta} - \frac{n^{(\varepsilon)}\alpha}{\theta^2} \log 2 + \sum_{i=1}^{n^{(\varepsilon)}} (\log x_i + \log y_i) + \frac{\alpha}{\theta^2} \sum_{i=1}^{n^{(\varepsilon)}} \log(x_i^\theta + y_i^\theta) \\
&\quad -(2 + \frac{\alpha}{\theta}) \sum_{i=1}^{n^{(\varepsilon)}} \frac{\partial}{\partial \theta} \log(x_i^\theta + y_i^\theta) \,.
\end{aligned}
$$

From this we obtain the MLE $\widehat{\lambda}^{(\varepsilon)} = \frac{n^{(\varepsilon)}}{t}$, whose asymptotic properties are well-known, and note that $\widehat{\lambda}^{(\varepsilon)}$ is independent of $\widehat{\alpha}$ and $\widehat{\theta}$. So we concentrate on $\widehat{\alpha}$ and $\widehat{\theta}$.

Note first that, as a consequence of (3.4.9), the d.f. of $X^* = \frac{X}{\varepsilon}$ is given by

$$
P(X^* > x) \;=\; P(X > \varepsilon x) = 2^{\alpha/\theta}(x^\theta + 1)^{-\alpha/\theta} \quad \text{for} \quad x > 1.
$$

Since also the distributions of $(X^*, Y^*) = (\frac{X}{\varepsilon}, \frac{Y}{\varepsilon})$ is independent of ε, the following quantities are independent of ε.

Lemma 3.4.4. *The following moments are finite.*

$$
\begin{aligned}
\mathbb{E}\left[\log\left(\frac{X}{\varepsilon}\right)\right] &= 2^{\frac{\alpha}{\theta}} \int_1^\infty \frac{(1+y^\theta)^{-\frac{\alpha}{\theta}}}{y} dy \\
\mathbb{E}\left[\log\left(\frac{1}{2}\left(\left(\frac{X}{\varepsilon}\right)^\theta + \left(\frac{Y}{\varepsilon}\right)^\theta\right)\right)\right] &= \frac{\theta}{\alpha} + \frac{\theta}{\alpha+\theta} \\
\mathbb{E}\left[\frac{\partial}{\partial\theta} \log\left(\left(\frac{X}{\varepsilon}\right)^\theta + \left(\frac{Y}{\varepsilon}\right)^\theta\right)\right] &= (2 + \frac{\alpha}{\theta}) \log \varepsilon + \frac{2}{\theta} + \mathbb{E}\left[\log\left(\frac{X}{\varepsilon}\right) + \log\left(\frac{Y}{\varepsilon}\right)\right] \\
&= (\frac{2\theta}{2\theta+\alpha}) \left(\frac{1}{\theta} + 2^{\frac{\alpha}{\theta}} \int_1^\infty \frac{(y^\theta+1)^{-\frac{\alpha}{\theta}}}{y} dy\right) .
\end{aligned}
$$

Proof. The first equality is a consequence of the joint density (3.4.10) and marginal tail distribution (3.4.9) with some standard analysis.

The second equality is calculated from the score function for α and (3.4.11).

For the last identity we calculate

$$
\begin{aligned}
& (2+\frac{\alpha}{\theta})\mathbb{E}\left[\frac{\partial}{\partial\theta}\log(X^\theta+Y^\theta)\right] \\
= & \frac{1}{\alpha+\theta}-\frac{\alpha}{\theta^2}\log 2+\mathbb{E}\left[\log X+\log Y\right]+\frac{\alpha}{\theta^2}\mathbb{E}\left[\log(X^\theta+Y^\theta)\right] \\
= & \frac{1}{\alpha+\theta}-\frac{\alpha}{\theta^2}\log 2+\mathbb{E}\left[\log X+\log Y\right]+\frac{\alpha}{\theta^2}(\theta\log\varepsilon+\frac{\theta}{\alpha}+\frac{\theta}{\alpha+\theta}+\log 2) \\
= & \frac{1}{\alpha+\theta}(1+\frac{\alpha}{\theta})+\mathbb{E}\left[\log X+\log Y\right]+\frac{1}{\theta}+\frac{\alpha}{\theta}\log\varepsilon \\
= & \mathbb{E}\left[\log X+\log Y\right]+\frac{2}{\theta}+\frac{\alpha}{\theta}\log\varepsilon \\
= & (2+\frac{\alpha}{\theta})\log\varepsilon+\frac{2}{\theta}+2^{\frac{\alpha}{\theta}+1}\int_1^\infty\frac{(y^\theta+1)^{-\frac{\alpha}{\theta}}}{y}dy\,.
\end{aligned}
$$

□

The following is a first step for calculating the Fisher information matrix.

Lemma 3.4.5. *For all $\varepsilon>0$,*

$$
\mathbb{E}\Big[\frac{\partial\ell^{(\varepsilon)}}{\partial\alpha}\Big]=\mathbb{E}\Big[\frac{\partial\ell^{(\varepsilon)}}{\partial\theta}\Big]=0. \tag{3.4.11}
$$

Proof. We show the result for the partial derivative with respect to α, where we use a dominated convergence argument. Since derivatives are local objects, it suffices to show that for each $\alpha_0\in(0,1)$ there exist a $\xi>0$ such that for all α in a neighbourhood of α_0, given by $N_\xi(\alpha_0):=\{\alpha\in(0,1):0<\alpha_0-\xi\le\alpha\le\alpha_0+\xi<1\}$ there exists a dominating integrable function, independent of α. We obtain

$$
\left|\frac{\partial\ell^{(\varepsilon)}}{\partial\alpha}\right| \;\le\; \frac{n^{(\varepsilon)}}{\alpha_0-\xi}+\frac{n^{(\varepsilon)}}{\alpha_0-\xi+\theta}+n^{(\varepsilon)}\log\varepsilon+\frac{n^{(\varepsilon)}\log 2}{\theta}+\frac{1}{\theta}\sum_{i=1}^{n^{(\varepsilon)}}\Big|\log(x_i^\theta+y_i^\theta\Big|.
$$

The right-hand side is integrable by Lemma 3.4.4, which can be seen by multiplying and dividing the x_i and y_i by ε and using the second identity of Lemma 3.4.4.
The proof for the partial derivative with respect to θ is similar, invoking Lemma 3.4.4. □

Next we calculate the second derivatives

$$\begin{aligned}
\frac{\partial^2 \ell^{(\varepsilon)}}{\partial \alpha^2} &= n^{(\varepsilon)} \left(-\frac{1}{\alpha^2} - \frac{1}{(\alpha+\theta)^2} \right) \\
\frac{\partial^2 \ell^{(\varepsilon)}}{\partial \alpha \partial \theta} &= n^{(\varepsilon)} \left(-\frac{1}{(\alpha+\theta)^2} - \frac{1}{\theta^2} \log 2 \right) - \frac{1}{\theta} \sum_{i=1}^{n^{(\varepsilon)}} \frac{\partial}{\partial \theta} \log(x_i^\theta + y_i^\theta) + \frac{1}{\theta^2} \sum_{i=1}^{n^{(\varepsilon)}} \log(x_i^\theta + y_i^\theta) \\
\frac{\partial^2 \ell^{(\varepsilon)}}{\partial \theta \partial \alpha} &= \frac{\partial^2 \ell^{(\varepsilon)}}{\partial \alpha \partial \theta} \\
\frac{\partial^2 \ell^{(\varepsilon)}}{\partial \theta^2} &= -n^{(\varepsilon)} \left(\frac{1}{(\alpha+\theta)^2} - \frac{2\alpha}{\theta^3} \log 2 \right) - \frac{2\alpha}{\theta^3} \sum_{i=1}^{n^{(\varepsilon)}} \log(x_i^\theta + y_i^\theta) + \frac{2\alpha}{\theta^2} \sum_{i=1}^{n^{(\varepsilon)}} \frac{\partial}{\partial \theta} \log(x_i^\theta + y_i^\theta) \\
&\quad -(2 + \frac{\alpha}{\theta}) \sum_{i=1}^{n^{(\varepsilon)}} \frac{\partial^2}{\partial \theta^2} \log(x_i^\theta + y_i^\theta).
\end{aligned}$$

In order to calculate the Fisher information matrix we invoke Lemma 3.4.4.

The components of the Fisher information matrix are then given by

$$\begin{aligned}
\widetilde{i}_{11} &= \mathbb{E}\left[-\frac{\partial^2}{\partial \alpha^2} \ell^{(\varepsilon)} \right] = \lambda^{(\varepsilon)} t \left[\frac{1}{\alpha^2} + \frac{1}{(\alpha+\theta)^2} \right] \\
&=: \lambda^{(\varepsilon)} i_{11}\, t \\
\widetilde{i}_{12} &= \widetilde{i}_{21} = \mathbb{E}\left[-\frac{\partial^2}{\partial \alpha \partial \theta} \ell^{(\varepsilon)} \right] \\
&= \lambda^{(\varepsilon)} t \left[\frac{1}{(\alpha+\theta)^2} - \frac{1}{\alpha\theta} - \frac{1}{\theta(\alpha+\theta)} + \frac{1}{\theta} \mathbb{E}\left[\frac{\partial}{\partial \theta} \log\left(\left(\frac{X}{\varepsilon}\right)^\theta + \left(\frac{Y}{\varepsilon}\right)^\theta \right) \right] \right] \\
&= \lambda^{(\varepsilon)} t \left[\frac{1}{(\alpha+\theta)^2} + \frac{2}{\theta(2\theta+\alpha)} - \frac{1}{\alpha\theta} - \frac{1}{\theta(\alpha+\theta)} + \frac{2^{\alpha/\theta+1}}{2\theta+\alpha} \int_1^\infty \frac{(1+u^\theta)^{-\frac{\alpha}{\theta}}}{u} du \right] \\
&=: \lambda^{(\varepsilon)} i_{12}\, t = \lambda_{21}^{(\varepsilon)} i_{12}\, t
\end{aligned}$$

$$\begin{aligned}
\widetilde{i}_{22} &= \mathbb{E}\left[-\frac{\partial^2}{\partial \theta^2} \ell^{(\varepsilon)} \right] = \lambda^{(\varepsilon)} t \left[\frac{1}{(\alpha+\theta)^2} - \frac{2\alpha \log 2}{\theta^3} + \frac{2\alpha}{\theta^3} \mathbb{E}\left(\log(X^\theta + Y^\theta) \right) \right. \\
&\quad \left. -\frac{2\alpha}{\theta^2} \mathbb{E}\left(\frac{\partial}{\partial \theta} \log(X^\theta + Y^\theta) \right) + (\frac{\alpha}{\theta} + 2) \mathbb{E}\left(\frac{\partial^2}{\partial \theta^2} \log(X^\theta + Y^\theta) \right) \right] \\
&= \lambda^{(\varepsilon)} t \left[\frac{1}{(\alpha+\theta)^2} + \frac{2}{\theta^2} + \frac{2\alpha}{\theta^2(\alpha+\theta)} - \frac{4\alpha}{\theta^2(\alpha+2\theta)} - \frac{\alpha 2^{\alpha/\theta+2}}{\theta(2\theta+\alpha)} \int_1^\infty \frac{(u^\theta+1)^{-\alpha/\theta}}{u} du \right. \\
&\quad \left. +(\frac{\alpha}{\theta} + 2) g(\alpha, \theta) \right] \\
&=: \lambda^{(\varepsilon)} i_{22}\, t\,,
\end{aligned}$$

where

$$g(\alpha,\theta) := \mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log(X^\theta + Y^\theta)\right] = \mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log\left(\left(\frac{X}{\varepsilon}\right)^\theta + \left(\frac{Y}{\varepsilon}\right)^\theta\right)\right]$$

does not depend on ε. This implies in particular that all i_{kl} are independent of ε. Consequently, the Fisher information matrix is given by

$$I^{(\varepsilon)}_{\alpha,\theta} = \lambda^{(\varepsilon)} t \begin{pmatrix} i_{11} & i_{12} \\ i_{12} & i_{22} \end{pmatrix}.$$

Recall the asymptotic normality of the estimated parameters in the one-dimensional case of Example 3.3.1. In our bivariate model we have additionally to those parameters the dependence parameter θ. This means that we have to check the regularity conditions (A1)-(A4) in Section 3.11 of Prakasa Rao [29] for this model. (A1) and (A2) are differentiability conditions, which are satisfied. As a prerequisite for (A3) and (A4) we need to show invertibility of the Fisher information matrix $I^{(\varepsilon)}_{\alpha,\theta}$, which we are not able to do analytically. A numerical study for a large number of values for α and θ, however, always gave a positive determinant, indicating that the inverse indeed exist. Since the Fisher information matrix depends on t only by the common factor, it is not difficult to convince ourselves that (A3) and (A4) are satisfied. Hence, classical likelihood theory applies and ensures that

$$\sqrt{n^{(\varepsilon)}} \begin{pmatrix} \widehat{\alpha} - \alpha \\ \widehat{\theta} - \theta \end{pmatrix} \sim \mathrm{AN}\left(\mathbf{0}, \begin{pmatrix} i_{11} & i_{12} \\ i_{12} & i_{22} \end{pmatrix}^{-1}\right), \quad n^{(\varepsilon)} \to \infty.$$

As in the one-dimensional case, we use the consistency result in (3.3.3) and Slutzky's theorem, which gives for $n^{(\varepsilon)} \to \infty$, equivalently, $\varepsilon \to 0$,

$$\sqrt{c2^{-\alpha/\theta}\varepsilon^{-\alpha}t} \begin{pmatrix} \widehat{\alpha} - \alpha \\ \widehat{\theta} - \theta \end{pmatrix} \xrightarrow{d} N\left(\mathbf{0}, \begin{pmatrix} i_{11} & i_{12} \\ i_{12} & i_{22} \end{pmatrix}^{-1}\right).$$

In reality the parameters are estimated from the data and plugged into the rate and the i_{kl}. Moreover, the unknown expectations in the Fisher information matrix have to be either numerically calculated by the corresponding integrals or estimated by Monte Carlo simulation. In Section 3.5 we shall perform a simulation study and also present an example of the covariance matrix for some specific choice of parameters.

Before this we want to come back to our change of parameters and, in particular, want to discuss estimation of the parameter c of the stable margins. From (3.3.3) and the fact that $\widehat{\lambda}^{(\varepsilon)}$, $\widehat{\alpha}$ and $\widehat{\theta}$ are consistent, we know that for $n^{(\varepsilon)} \to \infty$,

$$\widehat{c} := \widehat{\lambda}^{(\varepsilon)} 2^{\widehat{\alpha}/\widehat{\theta}} \varepsilon^{\widehat{\alpha}}$$

is a consistent estimator of c.

We calculate as follows

$$\begin{aligned}
\widehat{\log c} &= \log \widehat{\lambda}^{(\varepsilon)} + \frac{\widehat{\alpha}}{\widehat{\theta}} \log 2 + \widehat{\alpha} \log \varepsilon \\
&= \log \frac{\widehat{\lambda}^{(\varepsilon)}}{\lambda^{(\varepsilon)}} + \log c - \frac{\alpha}{\theta} \log 2 - \alpha \log \varepsilon + \frac{\widehat{\alpha}}{\widehat{\theta}} \log 2 + \widehat{\alpha} \log \varepsilon \,.
\end{aligned}$$

Consistency of $\widehat{\lambda}^{(\varepsilon)}$ implies that

$$\begin{aligned}
\widehat{\log c} - \log c &= o_P(1) + (\widehat{\alpha} - \alpha) \log \varepsilon + \Big(\frac{\widehat{\alpha}}{\widehat{\theta}} - \frac{\alpha}{\theta} \Big) \log 2 \\
&= o_P(1) + (\widehat{\alpha} - \alpha) \log \varepsilon + \Big(\frac{1}{\widehat{\theta}} - \frac{1}{\theta} \Big) \alpha \log 2 + (\widehat{\alpha} - \alpha) \frac{\log 2}{\widehat{\theta}} \\
&= o_P(1) + (\widehat{\alpha} - \alpha) \log \varepsilon + \Big(\frac{\theta}{\widehat{\theta}} - 1 \Big) \frac{\alpha}{\theta} \log 2 \\
&\quad + (\widehat{\alpha} - \alpha) \frac{\log 2}{\theta} (1 + o_P(1)) \,,
\end{aligned}$$

where we have used the consistency of $\widehat{\alpha}$ and $\widehat{\theta}$. This implies for $\varepsilon \to 0$

$$\frac{\widehat{\log c} - \log c}{\log \varepsilon} = (\widehat{\alpha} - \alpha)(1 + o_P(1)) \,.$$

Consequently, analogously to the one-dimensional case, we obtain the following result.

Theorem 3.4.6. *Let $(\widehat{c}, \widehat{\alpha}, \widehat{\theta})$ denote the MLEs of the bivariate α-stable Clayton subordinator. Then, as $\varepsilon \to 0$,*

$$\sqrt{c 2^{-\alpha/\theta} \varepsilon^{-\alpha} t} \begin{pmatrix} \dfrac{\widehat{\log c} - \log c}{\log \varepsilon} \\ \widehat{\alpha} - \alpha \\ \widehat{\theta} - \theta \end{pmatrix} \xrightarrow{d} \begin{pmatrix} N_1 \\ N_1 \\ N_2 \end{pmatrix} ,$$

where $\mathrm{Cov}(N_1, N_2) = \begin{pmatrix} i_{11} & i_{12} \\ i_{12} & i_{22} \end{pmatrix}^{-1}$ *is independent of* ε.

Obviously, we can do all again a Taylor expansion to obtain the limit law for $\widehat{c}$ instead of $\widehat{\log c}$ as in the one-dimensional case.

3.5 Simulation study for a bivariate α-stable Clayton subordinator

We start generating data from a bivariate α-stable Clayton subordinator over a time span $[0, t]$, where we choose $t = 1$ for simplicity. Recall that our observation scheme introduced in Section 3.4.1. assumes that from the α-stable Clayton subordinator we only observe bivariate jumps larger than ε. Obviously, we cannot simulate a trajectory of a stable process, since we are restricted to the simulation of a finite number of jumps. For simulation purpose we choose a threshold ξ (which should be much smaller than ε) and simulate jumps larger than ξ in one component, and arbitrary in the second component. To this end we invoke Algorithm 6.15 in Cont and Tankov [13].

The simulation of a bivariate stable Clayton subordinator is explained in detail in Example 6.18 of [13]. The algorithm starts by fixing a number τ determined by the required precision. This number coincides with $\lambda_1^{(\xi)}$ and fixes the average number of terms in (3.5.1) below.

We generate an i.i.d. sequence of standard exponential random numbers $E_1, E_2, \ldots$. Then we set $\Gamma_0^{(1)} = 0$ and $\Gamma_i^{(1)} = \Gamma_{i-1}^{(1)} + E_i$ until $\Gamma_{n^{(\xi)}}^{(1)} \leq \tau$ and $\Gamma_{n^{(\xi)}+1}^{(1)} > \tau$ resulting in the jump times of a standard Poisson process $\Gamma_0^{(1)}, \Gamma_1^{(1)}, \ldots, \Gamma_{n^{(\xi)}}^{(1)}$. Besides the marginal tail integrals we also need to know for every i the conditional distribution function given for $\Gamma_i^{(1)} = u > 0$ by

$$F_{2|1}(v \mid u) = \left(1 + (u/v)^{\delta}\right)^{-1/\delta - 1}, \quad v > 0\,.$$

We simulate $\Gamma_i^{(2)}$ from the d.f. $F_{2|1}(v \mid u = \Gamma_i^{(1)})$. Finally, we simulate a sequence $U_1, U_2, \ldots$ of i.i.d. uniform random numbers on $(0, 1)$. The trajectory of the bivariate Clayton sub-

ε		$\delta = 2$	$\alpha = 0.5$	$c = 1$
0.001	Mean	2.0861 (0.8245)	0.5323 (0.1233)	1.0642 (0.6848)
	$\sqrt{MSE}$	0.8290 (1.3074)	0.1275 (0.0340)	0.6878(0.9855)
	MRB	0.0476	0.0658	0.0232
0.0001	Mean	2.0180 (0.4333)	0.5110 (0.0637)	1.0531 (0.5174)
	$\sqrt{MSE}$	0.4337 (0.2831)	0.0647 (0.0078)	0.5201(0.5170)
	MRB	0.0108	0.0216	0.0423
0.00001	Mean	2.0029 (0.2364)	0.5041 (0.0348)	1.0270 (0.3713)
	$\sqrt{MSE}$	0.2364 (0.0781)	0.0350 (0.0021)	0.3722 (0.2730)
	MRB	0.0015	0.0081	0.0240

Table 3.1: Estimation of the bivariate $\frac{1}{2}$-stable Clayton process with jumps truncated at different ε: the mean of MLEs of the copula and the margins parameter δ, α and c with $\sqrt{MSE}$ and standard deviations (in brackets). This is based on a simulation of the process in a unit of time, $0 \le t < 1$, for $\tau = 1000$, equivalent to truncation of small jumps at the cut-off point $\xi = \overline{\Pi}^{\leftarrow}(\tau) = 10^{-6}$.

ordinator has the following representation

$$\begin{pmatrix} S_1^{(\xi)}(t) \\ S_2^{(\xi)}(t) \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n^{(\xi)}} 1_{\{U_i \le t\}} \overline{\Pi}_1^{\leftarrow}(\Gamma_i^{(1)}) \\ \sum_{i=1}^{n^{(\xi)}} 1_{\{U_i \le t\}} \overline{\Pi}_2^{\leftarrow}(\Gamma_i^{(2)}) \end{pmatrix}, \quad 0 < t < 1\,, \tag{3.5.1}$$

where $(\Gamma_i^{(1)}, \Gamma_i^{(2)})$ carry the dependence structure of the Lévy copula. Note that the jump times in both components always coincide.

Table 4.1 summarizes the results of a simulation study based on 100 trajectories of the bivariate α-stable Clayton subordinator with parameters $\alpha = 0.5$, $c = 1$ and Clayton dependence parameter $\delta = 2$.

Finally, we also want to give an idea about the theoretical properties of our MLE procedure. To this end we calculate the theoretical asymptotic covariance matrix for the same set of parameters $(c, \alpha, \theta) = (1, 0.5, 1)$. Note that in this case we can calculate the integral in the Fisher information matrix explicitly. The expectation of the second derivative we obtain from a Monte Carlo simulation based on simulated (X, Y).

We conclude this section with an example of the covariance matrix $\mathrm{Cov}(N_1, N_2)$ of the normal limit vector of the parameter estimates as given in Theorem 3.4.6. We do this for the model with parameters $c = 1$, $\alpha = 0.5$ and $\theta = 1$ as used for the simulation with results summarized in Table 4.1. We present the matrix resulting from two different methods. The left hand matrix has been calculated by numerical integration, whereas the right hand matrix is the result of a Monte Carlo simulation based on 1000 observations from the bivariate Pareto distribution (3.4.10).

$$\begin{array}{cc} \text{Numerical integration} & \text{Monte Carlo simulation} \\ \begin{bmatrix} 0.2492 & -0.1885 \\ -0.1885 & 1.4686 \end{bmatrix} & \begin{bmatrix} 0.2487 & -0.1867 \\ -0.1867 & 1.4700 \end{bmatrix} \end{array}$$

3.6 Conclusion and outlook

For the specific bivariate α-stable Clayton subordinator with equal marginal Lévy processs we have estimated all parameters in one go and proved asymptotic normality for $n^{(\varepsilon)} \to \infty$. Observation scheme were joint jumps larger than ε in both components and a fixed observation interval $[0, t]$. This limit result holds for $t \to \infty$ or, equivalently, for $\varepsilon \to 0$.

Since this estimation procedure requires even for a bivariate model with the same marginal processes a non-trivial numerical procedure to estimate the parameters, it seems to be advisable to investigate also two-step procedures like IFM (inference for the marginals). In such a procedure the parameters of the marginals may well be different, and the model of arbitrary dimension, since marginals are estimated first and then the marginals are transformed before estimating the dependence structure in form of the Lévy copula. This well-known estimation procedure in the copula framework will be investigated in a follow-up paper.

Alternatively, one can apply non-parametric estimation procedures for Lévy measures as e.g. in [35].

Chapter 4

Two-step estimation of a multivariate Lévy process

SUMARRY

Based on the concept of a Lévy copula to describe the dependence structure of a multivariate Lévy process we present a new estimation procedure. We consider a parametric model for the marginal Lévy processes as well as for the Lévy copula and estimate the parameters by a two-step procedure. We first estimate the parameters of the marginal processes, and then estimate in a second step only the dependence structure parameter. For infinite Lévy measures we truncate the small jumps and base our statistical analysis on the large jumps of the model. Prominent example will be a bivariate stable Lévy process, which allows for analytic calculations and, hence, for a comparison of different methods. We prove asymptotic normality of the parameter estimates from the two-step procedure and, in particular, we derive the Godambe information matrix, whose inverse is the covariance matrix of the normal limit law. A simulation study investigates the loss of efficiency because of the two-step procedure and the truncation.

4.1 Introduction

In Esmaeili and Klüppelberg [18] we presented the maximum likelihood estimation (MLE) for a bivariate stable subordinator. We assumed for the marginal subordinators to be both stable with the same parameters and modelled the dependence structure by a Clayton Lévy copula. Estimation was based on observed jumps larger than some predefined $\varepsilon > 0$ in both components within a fixed interval $[0, t]$. For this model we computed the MLEs numerically and proved asymptotic normality for $\varepsilon \to 0$ and/or for $t \to \infty$, respectively. It is certainly useful to know that such a procedure works; but for more general models as, for instance, for higher dimensional models with different marginal Lévy processes, this estimation method becomes computationally very expensive.

Consequently, we present in this paper an alternative, which is a Lévy equivalent of the so-called IFM (inference functions for margins) method, a standard method in multivariate statistics; cf. Godambe [20], Joe [23], Ch. 10, and Xu [36], Ch. 2. The observation scheme as chosen in Esmaeili and Klüppelberg [18] was simple in the sense that we only considered observations with jumps in both components larger than some $\varepsilon > 0$. For this observation scheme, however, the marginally truncated processes are not independent of the Lévy copula parameter.

The appropriate observation scheme to separate marginal and dependence parameters of the small jumps truncated processes requires to consider each component process separately and observe jumps larger than ε in each single component. This results again in a compound Poisson process (CPP), where jumps larger than ε in both components are seen as joint jumps, and those jumps with sizes larger than ε only in one component (and smaller in the other) are treated as positive jumps in one component and jump size 0 in the other.

Separation of the marginals and the Lévy copula is based on Sklar's theorem for Lévy measures. Due to the fact that all Lévy processes with the exception of a CPP have a singularity in 0, the Lévy measure is considered on quadrants in $\mathbb{R}^d$ avoiding the origin. The simplest object to consider is, hence, a d-dimensional subordinator, which allows for only positive jumps in all components. We restrict ourselves in this paper to such processes, extensions to general Lévy processes are not difficult, but notationally involved; see Kallsen and Tankov [25] or Eder and Klüppelberg [15].

Our paper is organised as follows. In Section 4.2 we introduce the notion of a Lévy copula needed later to model the dependence structure between the components of a multivariate Lévy process. Here we also explain the truncation scheme of the observed jumps and present our prominent example, the bivariate α-stable Clayton subordinator. Section 4.3 is dedicated to the two-step estimation procedure. We prove asymptotic normality of the estimates in Section 4.4 including the calculation of the covariance matrix as the inverse of the Godambe information matrix. For a comparison with MLE based on the full model we calculate the log-likelihood function in Section 4.5. Finally, in Section 4.6, we perform a small simulation study, where we compare the quality of all three estimation methods: the full MLE and the MLE based on joint jumps only, and the two-step procedure.

4.2 Preliminaries

4.2.1 The Lévy copula

Throughout this paper we denote by $\mathbf{S} = (\mathbf{S}(t))_{t\geq 0}$ an increasing Lévy process with values in $\mathbb{R}_+^d$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. This means that $\mathbf{S}$ is a subordinator without Gaussian component, drift γ and a Lévy measure Π on $\mathbb{R}_+^d$ satisfying $\Pi(\{\mathbf{0}\}) = 0$ and $\int_{\mathbb{R}_+^d} \min\{x, 1\}\Pi(dx) < \infty$; cf. Sato [32], Th. 21.5, or Cont and Tankov [13], Prop. 3.10.

The tail integral of the Lévy measure Π is the function $\overline{\Pi} : [0,\infty]^d \to [0,\infty]$ defined by

$$\overline{\Pi}(x_1,\dots,x_d) = \begin{cases} \Pi([x_1,\infty) \times \cdots \times [x_d,\infty)) , & (x_1,\dots,x_d) \in [0,\infty)^d \setminus \{\mathbf{0}\} \\ 0 , & x_i = \infty \text{ for at least one } i, \\ \infty , & (x_1,\dots,x_d) = \mathbf{0}. \end{cases} \qquad (4.2.1)$$

The marginal tail integrals are defined analogously for $i = 1, \ldots, d$ as $\overline{\Pi}_i(x) = \Pi_i([x, \infty))$ for $x \geq 0$; cf. Cont and Tankov [13], Def. 5.7, and Kallsen and Tankov [25], Def. 3.3 and 3.4.

The jump dependence of the process $\mathbf{S}$ is part of the multivariate tail integral and can be described by a so-called Lévy copula. We recall the notion of a Lévy copula from [13, 25] to be a measure defining function $\mathfrak{C} : [0, \infty]^d \to [0, \infty]$ with Lebesgue margins $\mathfrak{C}_k(u) = u$ for all $u \in [0, \infty]$ and $k = 1, \ldots, d$.

The following result, called Sklar's Theorem for Lévy copulas, is central for our set-up; it has been proved in Cont and Tankov [13], Th. 5.4, for a bivariate Lévy process and in Kallsen and Tankov [25], Th. 3.6, for a d-dimensional Lévy process.

Theorem 4.2.1. *Let $\overline{\Pi}$ denote the tail integral of a spectrally positive d-dimensional Lévy process, whose components have Lévy measures $\Pi_1, \ldots, \Pi_d$. Then there exists a Lévy copula $\mathfrak{C} : [0, \infty]^d \to [0, \infty]$ such that for all $x_1, x_2, \ldots, x_d \in [0, \infty]$*

$$\overline{\Pi}(x_1, \ldots, x_d) = \mathfrak{C}\left(\overline{\Pi}_1(x_1), \ldots, \overline{\Pi}_d(x_d)\right). \tag{4.2.2}$$

If the marginal tail integrals are continuous, then this Lévy copula is unique. Otherwise, it is unique on $Ran(\overline{\Pi}_1) \times \cdots \times Ran(\overline{\Pi}_d)$.
Conversely, if $\mathfrak{C}$ is a Lévy copula and $\overline{\Pi}_1, \ldots, \overline{\Pi}_d$ are one-dimensional tail integrals of spectrally positive Lévy processes, then the relation (4.2.2) *defines the tail integral of a d-dimensional spectrally positive Lévy process and $\overline{\Pi}_1, \ldots, \overline{\Pi}_d$ are tail integrals of its components.*

4.2.2 Truncation of the small jumps

For notational convenience we proceed with a bivariate subordinator. As truncation point we choose $\varepsilon > 0$. Figure 4.1 shows how the Lévy measure Π on $\mathbb{R}_+^2 \setminus (0, \varepsilon)^2$ is divided into two parts, the part concentrated on $[\varepsilon, \infty)^2$, and the part concentrated on the axes, which is in fact the projected measure of Π on $[\varepsilon, \infty) \times (0, \varepsilon)$ and $(0, \varepsilon) \times [\varepsilon, \infty)$ to the horizontal and vertical axes, respectively.

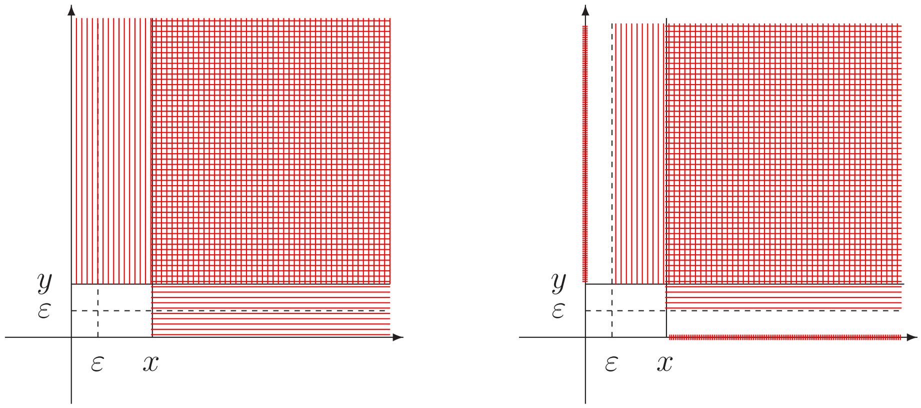

Figure 4.1: Illustration of the tail integral $\overline{\Pi}$ of a truncated bivariate Lévy process at (x, y) for a process with jump sizes of $\max\{x, y\} \geq \varepsilon$ (left) and a process with jump sizes of $x \geq \varepsilon$ and $y \geq \varepsilon$ (right). Note that in the right plot the mass, where only one component is larger than $\varepsilon > 0$ and the other smaller, is projected to the axes.

4.2.3 The observation scheme

It is based on all jumps of the process larger than some ε componentwise within the observation interval $[0, t]$. That is, we may observe a single jump x or y either in the first or in the second component. The other observed jumps are (x, y), where $x \geq \varepsilon$ and $y \geq \varepsilon$ at the same time. Let $n = n_1 + n_2$ denote the total number of jumps occurring in $[0, t]$ in either component, where we denote by n_1 and n_2 the number of jumps in each marginal process, respectively. This means that we count every joint jump in both components as two jumps. Then n decomposes in the number $n_1^{\perp}$ of jumps occurring only in the first component, the number $n_2^{\perp}$ of jumps occurring only in the second component, and the number $2n^{\parallel}$ of jumps occurring in both components.

We denote by $(x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_2})$ the observed jumps. By the independence property of the jumps of a Lévy process the order does not matter as long as concurrent jumps remain in the same coordinate. Consequently, throughout the paper we place w.l.o.g. all joint jumps at the beginning of the x- and y-observations so that $(\boldsymbol{x}^{\parallel}, \mathbf{y}^{\parallel}) = ((x_1, y_1), \ldots, (x_{n^{\parallel}}, y_{n^{\parallel}}))$.

The resulting $n^{\parallel} + n_1^{\perp} + n_2^{\perp}$ observations can be attributed to a bivariate CPP similar to the model considered in Esmaeili and Klüppelberg [17]. We shall need the marginal truncated Lévy measures $\Pi_k^{(\varepsilon)}$ for $k = 1, 2$. They will be calculated by first determining the Lévy measures of those processes representing joint jumps larger than ε, denoted by $\Pi^{(\varepsilon)\|}$, single jumps larger than ε in the first or second component, denoted by $\Pi_1^{(\varepsilon)\perp}$ and $\Pi_2^{(\varepsilon)\perp}$, respectively.

The tail integrals of the observed CPP are given for $x, y > \varepsilon$ by

$$\begin{aligned} \overline{\Pi}^{(\varepsilon)\|}(x,y) &= \overline{\Pi}(x,y)\,, \\ \overline{\Pi}_1^{(\varepsilon)\perp}(x) &= \overline{\Pi}(x,0) - \overline{\Pi}(x,\varepsilon)\,, \\ \overline{\Pi}_2^{(\varepsilon)\perp}(y) &= \overline{\Pi}(0,y) - \overline{\Pi}(\varepsilon,y)\,. \end{aligned} \tag{4.2.3}$$

The jump intensities of these CPPs are

$$\begin{aligned} \lambda^{(\varepsilon)\|} &= \overline{\Pi}(\varepsilon,\varepsilon)\,, \\ \lambda_1^{(\varepsilon)\perp} &= \overline{\Pi}(\varepsilon,0) - \overline{\Pi}(\varepsilon,\varepsilon)\,, \\ \lambda_2^{(\varepsilon)\perp} &= \overline{\Pi}(0,\varepsilon) - \overline{\Pi}(\varepsilon,\varepsilon). \end{aligned} \tag{4.2.4}$$

The corresponding jump size distributions are given by the Lévy measures divided by the intensities, respectively. The marginal tail integrals of the truncated processes are now calculated as

$$\begin{aligned} \overline{\Pi}_1^{(\varepsilon)}(x) &= \overline{\Pi}^{(\varepsilon)\|}(x,\varepsilon) + \overline{\Pi}_1^{(\varepsilon)\perp}(x) = \overline{\Pi}(x,0)\,, \quad x \geq \varepsilon \\ \overline{\Pi}_2^{(\varepsilon)}(y) &= \overline{\Pi}^{(\varepsilon)\|}(\varepsilon,y) + \overline{\Pi}_2^{(\varepsilon)\perp}(y) = \overline{\Pi}(0,y)\,, \quad y \geq \varepsilon\,, \end{aligned}$$

which implies intensities $\lambda_k^{(\varepsilon)} = \lambda^{(\varepsilon)\|} + \lambda_k^{(\varepsilon)\perp} = \overline{\Pi}_k(\varepsilon)$.

Lemma 4.1 in Esmaeili and Klüppelberg [18] explains the consequence of the small jumps truncation to the Lévy copula. We shall need the notion of a generalized inverse function: for $g : \mathbb{R} \to \mathbb{R}$ increasing define the *generalized inverse* of g as $g^{\leftarrow}(x) = \inf\{u \in \mathbb{R} : g(u) \geq x\}$. The definition extends naturally to other supports. For more details and properties of the generalized inverse we refer to Resnick [30], Section 0.2.

From Lemma 4.1 in Esmaeili and Klüppelberg [18] the Lévy copula of the CPP is given by

$$\mathfrak{C}^{(\varepsilon)}(u,v) = \mathfrak{C}\left(\mathfrak{C}_1^{\leftarrow}(u,\lambda_2^{(\varepsilon)}), \mathfrak{C}_2^{\leftarrow}(\lambda_1^{(\varepsilon)},v)\right), \quad 0 < u,v < \lambda^{(\varepsilon)\|}, \tag{4.2.5}$$

where for $k = 1,2$ the symbol $\mathfrak{C}_k^{\leftarrow}$ denotes the generalized inverse of $\mathfrak{C}$ with respect to the k-th argument.

The following will be our prominent example.

Example 4.2.2. *[Bivariate α-stable Clayton subordinator]*

Let $c_1, c_2 > 0$ and $0 < \alpha_1, \alpha_2 < 1$. Assume that $\overline{\Pi}_1(x) = c_1 x^{-\alpha_1}$ for $x > 0$ and $\overline{\Pi}_2(y) = c_2 y^{-\alpha_2}$ for $y > 0$ and that dependence is modelled by a Clayton Lévy copula, which is given by

$$\mathfrak{C}(u,v) = \left(u^{-\delta} + v^{-\delta}\right)^{-1/\delta}, \quad u,v > 0\,,$$

with dependence parameter $\delta > 0$.

By (4.2.3) the tail integrals of the observed CPP are given by

$$\begin{aligned}
\overline{\Pi}^{(\varepsilon)\|}(x,y) &= \left((c_1 x^{-\alpha_1})^{-\delta} + (c_2 y^{-\alpha_2})^{-\delta}\right)^{-\frac{1}{\delta}}, \quad x,y \geq \varepsilon\,, \\
\overline{\Pi}_1^{(\varepsilon)\perp}(x) &= c_1 x^{-\alpha_1}\left[1 - \left(1 + \left(\frac{c_2 \varepsilon^{-\alpha_2}}{c_1 x^{-\alpha_1}}\right)^{-\delta}\right)^{-1/\delta}\right], \quad x \geq \varepsilon\,, \\
\overline{\Pi}_2^{(\varepsilon)\perp}(y) &= c_2 y^{-\alpha_2}\left[1 - \left(1 + \left(\frac{c_1 \varepsilon^{-\alpha_1}}{c_2 y^{-\alpha_2}}\right)^{-\delta}\right)^{-1/\delta}\right], \quad y \geq \varepsilon\,.
\end{aligned} \tag{4.2.6}$$

From (4.2.4) we calculate the jump intensities

$$\begin{aligned}
\lambda^{(\varepsilon)\|} &= \left((c_1 \varepsilon^{-\alpha_1})^{-\delta} + (c_2 \varepsilon^{-\alpha_2})^{-\delta}\right)^{-\frac{1}{\delta}}, \\
\lambda_1^{(\varepsilon)\perp} &= c_1 \varepsilon^{-\alpha_1}\left[1 - \left(1 + \left(\frac{c_2 \varepsilon^{-\alpha_2}}{c_1 \varepsilon^{-\alpha_1}}\right)^{-\delta}\right)^{-1/\delta}\right], \\
\lambda_2^{(\varepsilon)\perp} &= c_2 \varepsilon^{-\alpha_2}\left[1 - \left(1 + \left(\frac{c_1 \varepsilon^{-\alpha_1}}{c_2 \varepsilon^{-\alpha_2}}\right)^{-\delta}\right)^{-1/\delta}\right].
\end{aligned} \tag{4.2.7}$$

The marginal tail integrals and intensities of the truncated process are now calculated for $k = 1,2$ as

$$\overline{\Pi}_k^{(\varepsilon)}(x) = c_k x^{-\alpha_k}, \quad x \geq \varepsilon\,, \quad \textit{and} \quad \lambda_k^{(\varepsilon)} = c_k \varepsilon^{-\alpha_k}\,.$$

This implies for the marginal jump size distributions

$$\begin{aligned} P(X > x) &= \overline{\Pi}_1^{(\varepsilon)}(x)/\lambda_1^{(\varepsilon)} = \varepsilon^{\alpha_1} x^{-\alpha_1}\,, \quad x \ge \varepsilon\,, \\ P(Y > y) &= \overline{\Pi}_2^{(\varepsilon)}(y)/\lambda_2^{(\varepsilon)} = \varepsilon^{\alpha_2} y^{-\alpha_2}\,, \quad y \ge \varepsilon\,. \end{aligned}$$

By (4.2.5) *the Lévy copula of the observed CPP is for* $0 < u, v < \lambda^{(\varepsilon)\|}$ *given by*

$$\begin{aligned} \mathfrak{C}^{(\varepsilon)}(u,v) &= \mathfrak{C}\left(\left(u^{-\delta} - (\lambda_2^{(\varepsilon)})^{-\delta}\right)^{-1/\delta}, \left(v^{-\delta} - (\lambda_1^{(\varepsilon)})^{-\delta}\right)^{-1/\delta}\right) \\ &= \left(u^{-\delta} + v^{-\delta} - (c_1^{-\delta}\varepsilon^{\alpha_1\delta} + c_2^{-\delta}\varepsilon^{\alpha_2\delta})\right)^{-1/\delta}\,. \end{aligned}$$

4.3 Two-step parameter estimation of a Lévy process

The idea of a two-step procedure for subordinators is similar to the IFM method for multivariate distributions. The term IFM is the acronym for “inference functions for margins” and has been applied in various areas of multivariate statistics; cf. Godambe [20] and Joe [23], Ch. 10. Obviously, the maximization of a likelihood with many parameters can be numerically sophisticated and computationally time-consuming; in a two-step method the parameters of the marginal components are estimated first and the Lévy copula parameters in a second step, thus reducing the dimensionality of the problem. For multivariate distribution functions, the algorithm is explained, for instance, in Joe [23], Ch. 10.

For a multivariate Lévy process in $\mathbb{R}^d$ for arbitrary dimension $d \in \mathbb{N}$, the two-step algorithm can be formalized as follows.

Step 1 : We do not distinguish between single and common jumps, but make use of all data available; i.e., we take all observations $x_{ik} > \varepsilon$ for $i = 1, \dots, n_k$ and all $k = 1, \dots, d$. We denote by $\gamma = (\theta_1, \dots, \theta_d)$ the vector of all marginal parameters (the θ_i are usually vectors) and let $l_1^{(\varepsilon)}, \dots, l_d^{(\varepsilon)}$ be the marginal log-likelihood functions with respect to the parameters. Determine

$$\widetilde{\gamma} := \operatorname{argmax}_\gamma \sum_{k=1}^{d} l_k^{(\varepsilon)}(\theta_k \mid \boldsymbol{x}_k)\,, \tag{4.3.1}$$

where $\boldsymbol{x}_k = (x_{1k}, x_{2k}, \ldots, x_{n_k k})$ are all observations in component k larger than ε.

Step 2 : Write the log-likelihood $l^{(\varepsilon)}$ of a CPP, whose jumps are only the common jumps of $x_{ik} > \varepsilon$ for $i = 1, \ldots, n^{\parallel}$ and $k = 1, \ldots, d$, plug in the marginal parameter estimates from Step 1, resulting in the log-likelihood of a CPP with only dependence structure parameter δ. Maximize the log-likelihood function over δ; i.e., estimate the Lévy copula parameter vector $\delta \in \mathbb{R}^m$ for some $m \in \mathbb{N}$, based on the common jumps:

$$\widetilde{\delta} := \operatorname{argmax}_{\delta} l^{(\varepsilon)}(\delta \mid \widetilde{\gamma}, \boldsymbol{x}_1^{\parallel}, \ldots, \boldsymbol{x}_d^{\parallel}), \tag{4.3.2}$$

where $\widetilde{\gamma} = (\widetilde{\theta}_1, \ldots, \widetilde{\theta}_d)$ and $\boldsymbol{x}_k^{\parallel} = (x_{1k}, \ldots, x_{n^{\parallel}k})$ for $k = 1, \ldots, d$.

Remark 4.3.1. *The MLE $\hat{\eta}$ of the parameter vector $\eta = (\theta_1, \ldots, \theta_d, \delta)$ is derived by maximization of the log-likelihood of the multivariate CPP $l^{(\varepsilon)}$ over the parameter vector η (as done in [18]). The estimate $\hat{\eta}$ is the solution of*

$$\left(\frac{\partial l^{(\varepsilon)}}{\partial \theta_1}, \ldots, \frac{\partial l^{(\varepsilon)}}{\partial \theta_d}, \frac{\partial l^{(\varepsilon)}}{\partial \delta} \right) = 0.$$

This is in contrast with the two-step method, where the estimate $\widetilde{\eta}$ is the solution of

$$\left(\frac{\partial l_1^{(\varepsilon)}}{\partial \theta_1}, \ldots, \frac{\partial l_k^{(\varepsilon)}}{\partial \theta_k}, \frac{\partial l^{(\varepsilon)}}{\partial \delta} \right) = 0.$$

Remark 4.3.2. *The aim of the two-step method is in fact to reduce the dimension of the parameter vector to have a simpler structure for the optimization of the likelihood function. Note that the observation scheme in [18], which takes only the $n^{\parallel}$ observations of the joint jumps in both steps into account, fails this goal, since the observation scheme used there introduces the dependence parameter into the marginal likelihoods.*

4.3.1 Two-step estimation method of an α-stable Clayton subordinator with different marginal parameters

The following algorithm works in principle in every dimension. For notational simplicity we formulate it only for dimension $d = 2$. Let $\mathbf{S} = (S_1, S_2)$ be a bivariate α-stable Clayton subordinator as introduced in Example 4.2.2 with different marginal parameters

$\theta_1 = (\alpha_1, c_1)$ and $\theta_2 = (\alpha_2, c_2)$ with $\alpha_k \in (0,1)$ and $c_k \in (0,\infty)$ for $k = 1,2$ and a Lévy copula parameter $\delta \in (0,\infty)$. We assume the observation scheme as described in Section 2.4. We denote by $(X_1, \dots, X_{n^\|}, \dots, X_{n_1}, Y_1, \dots, Y_{n^\|}, \dots, Y_{n_2})$ the vector of jumps larger than ε for the component processes $S_1^{(\varepsilon)}$ and $S_2^{(\varepsilon)}$, respectively. As before, all double jumps are numbered as (X_i, Y_i) for $i = 1, \dots, n^\|$.

Step 1 : Since the marginal log-likelihoods have the same structure with no common parameters, (4.3.1) decomposes in its components for S_1 and S_2, and maximization is done separately. We proceed as in Basawa and Brockwell [4, 5]; cf. Esmaeili and Klüppelberg [18], Example 3.1, and exemplify it for the first component:

$$l_1^{(\varepsilon)}(\log c_1, \alpha_1; \boldsymbol{x}) = -c_1 t \varepsilon^{-\alpha_1} + n_1(\log c_1 + \log \alpha_1) - (\alpha_1 + 1) \sum_{i=1}^{n_1} \log x_i\,.$$

From Basawa and Brockwell [4, 5] we know that the marginal MLEs of c_1 and α_1 and the intensity parameter $\lambda_1^{(\varepsilon)}$ are given by

$$\begin{aligned}
\widetilde{\lambda}_1^{(\varepsilon)} &= \frac{n_1}{t}\,, \\
\widetilde{\alpha}_1 &= \left(\frac{1}{n_1}\sum_{i=1}^{n_1}\Big(\log X_i - \log\varepsilon\Big) + \log\varepsilon\Big(1 - \frac{\lambda_1^{(\varepsilon)}}{\widetilde{\lambda}_1^{(\varepsilon)}}\Big)\right)^{-1}, \qquad (4.3.3) \\
\log \widetilde{c}_1 &= \log \widetilde{\lambda}_1^{(\varepsilon)} + \widetilde{\alpha}_1 \log \varepsilon\,.
\end{aligned}$$

Furthermore, asymptotic normality holds with degenerate limit for $(\widetilde{c}_1, \widetilde{\alpha}_1)$ and with asymptotic independence for $(\widetilde{\lambda}_1, \widetilde{\alpha}_1)$ as $n_1 \to \infty$. Limit laws hold for both situations, $t \to \infty$ or $\varepsilon \to 0$. The first limit was derived in Basawa and Brockwell [5]. Asymptotic independence for the second vector was shown in Höpfner and Jacod [22]. Both results are reported with precise rates and the asymptotic covariance matrix in Esmaeili and Klüppelberg [18], Example 3.1.

Step 2 : We first determine the log-likelihood function in (4.3.2) for the bivariate CPP of common jumps larger than ε. By (4.2.7) the intensity is $\lambda^{(\varepsilon)\|} = (c_1^{-\delta}\varepsilon^{\alpha_1\delta} + c_2^{-\delta}\varepsilon^{\alpha_2\delta})^{-\frac{1}{\delta}}$. Together with (4.2.6) this yields the survival function of bivariate joint jumps

$$\overline{F}^{(\varepsilon)}(x,y) = \left(\frac{c_1^{-\delta}x^{\alpha_1\delta} + c_2^{-\delta}y^{\alpha_2\delta}}{c_1^{-\delta}\varepsilon^{\alpha_1\delta} + c_2^{-\delta}\varepsilon^{\alpha_2\delta}}\right)^{-\frac{1}{\delta}}, \quad x, y \geq \varepsilon,$$

with density given by

$$f^{(\varepsilon)}(x,y) = \frac{\alpha_1\alpha_2(1+\delta)(c_1^{-\delta}\varepsilon^{\alpha_1\delta}+c_2^{-\delta}\varepsilon^{\alpha_2\delta})^{\frac{1}{\delta}}}{(c_1c_2)^{\delta}} \frac{x^{\alpha_1\delta-1}y^{\alpha_2\delta-1}}{(c_1^{-\delta}x^{\alpha_1\delta}+c_2^{-\delta}y^{\alpha_2\delta})^{\frac{1}{\delta}+2}}. \tag{4.3.4}$$

This results in the log-likelihood function

$$\begin{aligned} l^{(\varepsilon)}(c_1,c_2,\alpha_1,\alpha_2,\delta;\boldsymbol{x}^{\|},\mathbf{y}^{\|}) \; = \; & -\lambda^{(\varepsilon)\|}t + n^{\|}\log(1+\delta) - n^{\|}\delta(\log c_1+\log c_2) \\ & +n^{\|}(\log\alpha_1+\log\alpha_2) + (\alpha_1\delta-1)\sum_{i=1}^{n^{\|}}\log x_i + (\alpha_2\delta-1)\sum_{i=1}^{n^{\|}}\log y_i \\ & -(\frac{1}{\delta}+2)\sum_{i=1}^{n^{\|}}\log\left(c_1^{-\delta}x_i^{\alpha_1\delta}+c_2^{-\delta}y_i^{\alpha_2\delta}\right), \end{aligned}$$

where $(\boldsymbol{x}^{\|},\mathbf{y}^{\|}) = ((x_1,y_1),\ldots,(x_{n^{\|}},y_{n^{\|}}))$.

Given the marginal parameter estimates from the first step, the score function with respect to the dependence parameter δ is given by

$$\begin{aligned} \frac{\partial l^{(\varepsilon)}(\delta \mid \widetilde{\gamma},\boldsymbol{x}^{\|},\mathbf{y}^{\|})}{\partial\delta} \; = \; & -\frac{\partial\lambda^{(\varepsilon)\|}}{\partial\delta}t + \frac{n^{\|}}{1+\delta} - n^{\|}(\log\widetilde{c}_1+\log\widetilde{c}_2) \\ & +\widetilde{\alpha}_1\sum_{i=1}^{n^{\|}}\log x_i + \widetilde{\alpha}_2\sum_{i=1}^{n^{\|}}\log y_i + \frac{1}{\delta^2}\sum_{i=1}^{n^{\|}}\log\left(\widetilde{c}_1^{\,-\delta}x_i^{\widetilde{\alpha}_1\delta}+\widetilde{c}_2^{\,-\delta}y_i^{\widetilde{\alpha}_2\delta}\right) \\ & -\left(\frac{1}{\delta}+2\right)\sum_{i=1}^{n^{\|}}\frac{\partial}{\partial\delta}\log\left(\widetilde{c}_1^{\,-\delta}x_i^{\widetilde{\alpha}_1\delta}+\widetilde{c}_2^{\,-\delta}y_i^{\widetilde{\alpha}_2\delta}\right). \end{aligned}$$

The parameter estimate $\widetilde{\delta}$ can be found numerically by solving the following equation for δ:

$$\frac{\partial l^{(\varepsilon)}(\delta \mid \widetilde{\gamma},\boldsymbol{x}^{\|},\mathbf{y}^{\|})}{\partial\delta} = 0.$$

Remark 4.3.3. *The vector of score functions in the two-step method is given by*

$$\begin{aligned} & \mathbf{J}^{(\varepsilon)}(\mathbf{X},\mathbf{Y};\eta) \; = \\ & \Big(\frac{\partial l_1^{(\varepsilon)}(\log c_1,\alpha_1;\mathbf{X})}{\partial\log c_1}, \frac{\partial l_1^{(\varepsilon)}(\log c_1,\alpha_1;\mathbf{X})}{\partial\alpha_1}, \frac{\partial l_2^{(\varepsilon)}(\log c_2,\alpha_2;\mathbf{Y})}{\partial\log c_2}, \frac{\partial l_2^{(\varepsilon)}(\log c_2,\alpha_2;\mathbf{Y})}{\partial\alpha_2}, \frac{\partial l^{(\varepsilon)}(\delta;\mathbf{X}^{\|},\mathbf{Y}^{\|})}{\partial\delta}\Big)^T, \end{aligned}$$

where $\eta = (\log c_1,\log c_2,\alpha_1,\alpha_2,\delta)^T$ *is the parameter vector,* $\mathbf{X} = (X_1,\ldots,X_{n_1})$*,* $\mathbf{Y} = (Y_1,\ldots,Y_{n_2})$ *and* $(\mathbf{X}^{\|},\mathbf{Y}^{\|}) = ((X_1,Y_1),\ldots,(X_{n^{\|}},Y_{n^{\|}}))$*.*

4.3.2 Two-step method for a bivariate α-stable Clayton subordinator with common marginal parameters

For an analysis of the two-step estimation procedure we simplify the model as follows. Let $\mathbf{S} = (S_1, S_2)$ be a bivariate α-stable subordinator as in Example 4.2.2 with common marginal parameters $\theta_1 = \theta_2 = (\alpha, c)$ and a Clayton Lévy copula parameter δ. Assume an observation scheme as explained in Section 2.4. Maximum likelihood estimation for the parameters of this model was discussed in Esmaeili and Klüppelberg [18] in detail. In this section we estimate the parameters with the two-step method.

Step 1 : The log-likelihood function (4.3.1), which ignores the dependence structure, is given by

$$\begin{aligned} l_{12}^{(\varepsilon)}(\log c, \alpha) &= l_1^{(\varepsilon)}(\log c, \alpha) + l_2^{(\varepsilon)}(\log c, \alpha) \\ &= -2ct\varepsilon^{-\alpha} + n(\log c + \log \alpha) - (\alpha + 1)\sum_{i=1}^{n} \log z_i, \end{aligned} \tag{4.3.5}$$

where $n := n_1 + n_2$ is Poisson distributed. Since n_1 and n_2 have both intensity $\lambda^{(\varepsilon)} := \lambda_1^{(\varepsilon)} = \lambda_2^{(\varepsilon)}$, n has intensity $2\lambda^{(\varepsilon)} = 2c\varepsilon^{-\alpha}$ and $(z_1, \ldots, z_n) = (x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_2})$. Note that the corresponding random variables $\log(\frac{Z_i}{\varepsilon})$, for $i = 1, \ldots, n$ are exponentially distributed with density $f(u) = \alpha e^{-\alpha u}$ for $u > 0$. The log-likelihood has score functions with respect to the marginal parameters $\log c$ and α as follows:

$$\begin{aligned} \frac{\partial l_{12}^{(\varepsilon)}(\log c, \alpha)}{\partial \log c} &= n - 2ct\varepsilon^{-\alpha} = n - 2\lambda^{(\varepsilon)} t \qquad (4.3.6) \\ \frac{\partial l_{12}^{(\varepsilon)}(\log c, \alpha)}{\partial \alpha} &= \frac{n}{\alpha} + 2ct\varepsilon^{-\alpha}\log\varepsilon - \sum_{i=1}^{n}\log z_i = -\sum_{i=1}^{n}\Big(\log\frac{z_i}{\varepsilon} - \frac{1}{\alpha}\Big) - (n - 2\lambda^{(\varepsilon)} t)\log\varepsilon\,. \end{aligned}$$

The common intensity parameter $\lambda^{(\varepsilon)} = c\varepsilon^{-\alpha}$ and the marginal parameters $\log c$ and α can be estimated by (4.3.3) as

$$\begin{aligned} \widetilde{\lambda}^{(\varepsilon)} &= \frac{n}{2t} \\ \widetilde{\alpha} &= \Big(\frac{1}{n}\sum_{i=1}^{n}(\log Z_i - \log\varepsilon) + \big(1 - \frac{\lambda^{(\varepsilon)}}{\widetilde{\lambda}^{(\varepsilon)}}\big)\log\varepsilon\Big)^{-1} \\ \log\widetilde{c} &= \log\widetilde{\lambda}^{(\varepsilon)} + \widetilde{\alpha}\log\varepsilon\,. \end{aligned}$$

Step 2 : As explained in Esmaeili and Klüppelberg [18], for simplifying the calculations of the second derivatives later we reparameterize the dependence to $\theta = \alpha\delta$. The joint density of bivariate jumps is a special case of (4.3.4) and has been calculated in (4.10) in Esmaeili and Klüppelberg [18]. From (4.2.7) we know that $\lambda^{(\varepsilon)\|} = c\varepsilon^{-\alpha}2^{-\frac{\alpha}{\theta}}$, which we use for abbreviation. Then the log-likelihood in (4.3.2) is

$$\begin{aligned} l^{(\varepsilon)}(\log c, \alpha, \theta) &= -\lambda^{(\varepsilon)\|}t + n^{\|}\log\alpha + n^{\|}\log(\alpha+\theta) + n^{\|}\log c \\ &\quad +(\theta-1)\sum_{i=1}^{n^{\|}}(\log x_i + \log y_i) - (2+\frac{\alpha}{\theta})\sum_{i=1}^{n^{\|}}\log(x_i^{\theta}+y_i^{\theta})\,. \end{aligned} \tag{4.3.7}$$

The score function with respect to the parameter θ is then given by (the derivatives of $\lambda^{(\varepsilon)\|}$ are calculated in Lemma 4.3.5 below)

$$\begin{aligned} \frac{\partial l^{(\varepsilon)}}{\partial\theta} &= -\frac{\partial\lambda^{(\varepsilon)\|}}{\partial\theta}t + \frac{n^{\|}}{\alpha+\theta} + \sum_{i=1}^{n^{\|}}(\log x_i + \log y_i) \\ &\quad +\frac{\alpha}{\theta^2}\sum_{i=1}^{n^{\|}}\log(x_i^{\theta}+y_i^{\theta}) - (2+\frac{\alpha}{\theta})\sum_{i=1}^{n^{\|}}\frac{\partial}{\partial\theta}\log(x_i^{\theta}+y_i^{\theta}). \end{aligned} \tag{4.3.8}$$

Given the estimates of the marginal parameters $\widetilde{c}$ and $\widetilde{\alpha}$ from the first step, the estimate of θ can be computed numerically as the argmax of the right hand side of (4.3.8).

Remark 4.3.4. *The vector of score functions from Remark 4.3.3 reduces to*

$$\mathbf{J}^{(\varepsilon)}(\mathbf{X},\mathbf{Y};\eta) = \Big(\frac{\partial l_{12}^{(\varepsilon)}(\log c,\alpha;\mathbf{Z})}{\partial\log c}, \frac{\partial l_{12}^{(\varepsilon)}(\log c,\alpha;\mathbf{Z})}{\partial\alpha}, \frac{\partial l^{(\varepsilon)}(\log c,\alpha,\theta;\mathbf{X}^{\|},\mathbf{Y}^{\|})}{\partial\theta}\Big)^T, \tag{4.3.9}$$

where $\eta = (\log c, \alpha, \theta)^T$ *is the parameter vector,* $\mathbf{Z} = (X_1,\dots,X_{n_1},Y_1,\dots,Y_{n_2})$ *and* $(\mathbf{X}^{\|},\mathbf{Y}^{\|}) = (X_1,Y_1),\dots,(X_{n^{\|}},Y_{n^{\|}})$.

We shall need the following derivatives of $\lambda^{(\varepsilon)\|}$.

Lemma 4.3.5. *For* $\lambda^{(\varepsilon)\|} = c\varepsilon^{-\alpha}2^{-\frac{\alpha}{\theta}}$ *the partial derivatives are given by*

$$\frac{\partial\lambda^{(\varepsilon)\|}}{\partial\log c} = \lambda^{(\varepsilon)\|}\,, \quad \frac{\partial\lambda^{(\varepsilon)\|}}{\partial\alpha} = -\lambda^{(\varepsilon)\|}\big(\log\varepsilon + \frac{1}{\theta}\log 2\big)\,, \quad \frac{\partial\lambda^{(\varepsilon)\|}}{\partial\theta} = \lambda^{(\varepsilon)\|}\frac{\alpha\log 2}{\theta^2}.$$

The second derivatives can be calculated as

$$\begin{aligned}
\frac{\partial^2 \lambda^{(\varepsilon)\|}}{\partial\theta\partial\log c} &= \lambda^{(\varepsilon)\|}\frac{\alpha\log 2}{\theta^2}\,, \\
\frac{\partial^2 \lambda^{(\varepsilon)\|}}{\partial\theta\partial\alpha} &= -\lambda^{(\varepsilon)\|}\frac{\log 2}{\theta^2}\Big(\alpha\log\varepsilon + \frac{\alpha}{\theta}\log 2 - 1\Big)\,, \\
\frac{\partial^2 \lambda^{(\varepsilon)\|}}{\partial\theta^2} &= \lambda^{(\varepsilon)\|}\frac{\alpha\log 2}{\theta^2}\Big(\frac{\alpha\log 2}{\theta^2} - \frac{2}{\theta}\Big)\,.
\end{aligned}$$

4.4 Asymptotic properties of the two-step estimates

4.4.1 The Godambe information matrix

In the two-step estimation procedure, the Godambe information matrix plays the role of the Fisher information matrix in classical MLE. We explain this for our situation. Let $\mathbf{S} = (S_1, S_2)$ be a bivariate α-stable Clayton subordinator with parameter vector $\eta \in \mathbb{R}^k$ including marginal and dependence parameters. Assume further an observation scheme as explained in Section 2.4. In principle the two-step estimation procedure can be applied to both situations of Section 4.3.1 with $\eta \in \mathbb{R}^5$ or of Section 4.3.2 with $\eta \in \mathbb{R}^3$.

For the vector of score functions, denoted by $\mathbf{J}^{(\varepsilon)}(\mathbf{X}, \mathbf{Y}; \eta)$ as in Remarks 4.3.3 and 4.3.4, the so-called *Godambe information matrix* is calculated as

$$G \;:=\; D^T M^{-1} D, \tag{4.4.1}$$

where

$$D \;:=\; \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[-\frac{\partial \mathbf{J}^{(\varepsilon)}(\mathbf{X}, \mathbf{Y}; \eta)}{\partial\eta}\Big], \tag{4.4.2}$$

$$M \;:=\; \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[\mathbf{J}^{(\varepsilon)}(\mathbf{X}, \mathbf{Y}; \eta)\mathbf{J}^{(\varepsilon)}(\mathbf{X}, \mathbf{Y}; \eta)^T\Big] \tag{4.4.3}$$

are $k\times k$-matrices. Under some regularity conditions (see below) the asymptotic covariance matrix of $n^{-\frac{1}{2}}(\widetilde{\eta} - \eta)$ is equal to the inverse of G; cf. Joe [23], Section 10.1.1.

For the rest of this section we restrict the process $\mathbf{S}$ again to the model in Section 4.3.2, a bivariate α-stable subordinator with common marginal parameters $\log c$ and α, and dependence parameter θ. We denote by $l_{12}^{(\varepsilon)}$ the log-likelihood of the common marginal parameters $\gamma := (\log c, \alpha)$ as in (4.3.5), and by $l^{(\varepsilon)}$ the log-likelihood of the bivariate CPP in the second step as in (4.3.7). Assume further that $\eta_0 = (\log c_0, \alpha_0, \theta_0)$ is the true parameter vector. We prove consistency of the two-step estimators, and their joint asymptotic normality. We calculate the Godambe information matrix G as well as the asymptotic covariance matrix of the estimators. We do this step by step, calculating the matrices D and M from (4.4.2) and (4.4.3), respectively.

There is a minor difference between our approach and the classical in Joe [23], Section 10.1.1. Whereas he can work with a fixed number of multivariate data, our process structure with observations on an interval $[0, t]$ implies a random number of data points. Moreover, we have to deal with the problem of single and common jumps. However, this imposes no real difficulties, since we can invoke Slutzky's theorem.

Lemma 4.4.1. *Assume a bivariate α-stable Clayton subordinator with common marginal parameters $(\log c, \alpha)$ and dependence parameter $\theta = \alpha\delta$. Assume also the observation scheme given in Section 2.2. Recall that $\lambda^{(\varepsilon)} = \lambda_1^{(\varepsilon)} = \lambda_2^{(\varepsilon)} = c\varepsilon^{-\alpha}$ is the marginal intensity parameter, $\lambda^{(\varepsilon)\|} = c\varepsilon^{-\alpha}2^{-\frac{\alpha}{\theta}}$ is the joint jumps intensity parameter. Define $d = \frac{\lambda^{(\varepsilon)\|}}{2\lambda^{(\varepsilon)}} = 2^{-\frac{\alpha}{\theta}-1}$. Then the matrix D of (4.4.2) is given by*

$$D = \begin{pmatrix} 1 & -\log\varepsilon & 0 \\ -\log\varepsilon & \frac{1}{\alpha^2} + (\log\varepsilon)^2 & 0 \\ d\frac{\alpha\log 2}{\theta^2} & d\big(-\frac{\alpha\log 2}{\theta^2}\log\varepsilon + a\big) & d\,b \end{pmatrix}. \tag{4.4.4}$$

Furthemore, $a = a(\alpha, \theta)$ and $b = b(\alpha, \theta)$ are given in (4.4.7) and (4.4.8). They are deterministic functions of the parameters and do not depend on t or ε.

Proof. The score functions in (4.3.6) have derivatives

$$\begin{aligned}
\frac{\partial^2 l_{12}^{(\varepsilon)}(\log c, \alpha)}{\partial(\log c)^2} &= -2ct\varepsilon^{-\alpha} = -2\lambda^{(\varepsilon)}t \\
\frac{\partial^2 l_{12}^{(\varepsilon)}(\log c, \alpha)}{\partial\alpha\partial\log c} &= \frac{\partial^2 l_{12}^{(\varepsilon)}(\log c, \alpha)}{\partial\log c\,\partial\alpha} = 2ct\varepsilon^{-\alpha}\log\varepsilon = 2\lambda^{(\varepsilon)}t\log\varepsilon \\
\frac{\partial^2 l_{12}^{(\varepsilon)}(\log c, \alpha)}{\partial\alpha^2} &= -\frac{n}{\alpha^2} - 2ct\varepsilon^{-\alpha}(\log\varepsilon)^2 = -\frac{n}{\alpha^2} - 2\lambda^{(\varepsilon)}t(\log\varepsilon)^2.
\end{aligned} \tag{4.4.5}$$

This means that the upper left 2×2-matrix is the Fisher information matrix to the MLE of $(\log c, \alpha)$, calculated by Basawa and Brockwell [5] and presented in Esmaeili and Klüppelberg [18], Example 3.1 (up to a deterministic factor), since here all observations from both marginals are considered. Since the score functions in (4.3.6) are independent of the parameter θ, the matrix D has the structure as given in (4.4.4). It remains to calculate the last row of D. We calculate the derivatives of the score function in (4.3.8) as follows:

$$\frac{\partial^2 l^{(\varepsilon)}(\log c, \alpha, \theta)}{\partial \log c \partial \theta} = -\frac{\partial^2 \lambda^{(\varepsilon)\|}}{\partial \log c \partial \theta} t = -\lambda^{(\varepsilon)\|} t \frac{\alpha \log 2}{\theta^2}$$

$$\begin{aligned}
\frac{\partial^2 l^{(\varepsilon)}(\log c, \alpha, \theta)}{\partial \alpha \partial \theta} &= -\frac{\partial^2 \lambda^{(\varepsilon)\|}}{\partial \alpha \partial \theta} t - \frac{n^\|}{(\alpha+\theta)^2} + \frac{1}{\theta^2} \sum_{i=1}^{n^\|} \log\left(X_i^\theta + Y_i^\theta\right) \\
&\quad -\frac{1}{\theta} \sum_{i=1}^{n^\|} \frac{\partial}{\partial \theta} \log\left(X_i^\theta + Y_i^\theta\right) \qquad (4.4.6) \\
\frac{\partial^2 l^{(\varepsilon)}(\log c, \alpha, \theta)}{\partial \theta^2} &= -\frac{\partial^2 \lambda^{(\varepsilon)\|}}{\partial \theta^2} t - \frac{n^\|}{(\alpha+\theta)^2} - \frac{2\alpha}{\theta^3} \sum_{i=1}^{n^\|} \log\left(X_i^\theta + Y_i^\theta\right) \\
&\quad + \frac{2\alpha}{\theta^2} \sum_{i=1}^{n^\|} \frac{\partial}{\partial \theta} \log\left(X_i^\theta + Y_i^\theta\right) - \left(2 + \frac{\alpha}{\theta}\right) \sum_{i=1}^{n^\|} \frac{\partial^2}{\theta^2} \log\left(X_i^\theta + Y_i^\theta\right).
\end{aligned}$$

It remains to calculate the following expectations:

$$\begin{aligned}
d_{31} &= \frac{1}{2\lambda^{(\varepsilon)} t} \mathbb{E}\Big[- \frac{\partial^2 l^{(\varepsilon)}}{\partial \log c \partial \theta} \Big] = d \, \frac{\alpha \log 2}{\theta^2} \\
d_{32} &= \frac{1}{2\lambda^{(\varepsilon)} t} \mathbb{E}\Big[- \frac{\partial^2 l^{(\varepsilon)}}{\partial \alpha \partial \theta} \Big] = \frac{1}{2\lambda^{(\varepsilon)} t} \left(\frac{\partial^2 \lambda^{(\varepsilon)\|}}{\partial \theta \partial \alpha} t + \frac{\lambda^{(\varepsilon)\|} t}{(\alpha+\theta)^2} - \frac{\lambda^{(\varepsilon)\|} t}{\theta^2} \mathbb{E}\Big[\log(X_1^\theta + Y_1^\theta) \Big] \right. \\
&\quad \left. + \frac{\lambda^{(\varepsilon)\|} t}{\theta} \mathbb{E}\Big[\frac{\partial}{\partial \theta} \log\left(X_1^\theta + Y_1^\theta\right) \Big] \right) \\
&= d \Big(- \frac{\alpha \log 2}{\theta^2} \log \varepsilon - \frac{\alpha (\log 2)^2}{\theta^3} + \frac{\log 2}{\theta^2} + \frac{1}{(\alpha+\theta)^2} - \frac{1}{\theta^2} \mathbb{E}\Big[\log(X_1^\theta + Y_1^\theta) \Big] \\
&\quad + \frac{1}{\theta} \mathbb{E}\Big[\frac{\partial}{\partial \theta} \log\left(X_1^\theta + Y_1^\theta\right) \Big] \Big)
\end{aligned}$$

$$
\begin{aligned}
d_{33} &= \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[-\frac{\partial^2 l^{(\varepsilon)}}{\partial\theta^2}\Big] = \frac{1}{2\lambda^{(\varepsilon)}t}\Bigg(\frac{\partial^2\lambda^{(\varepsilon)\|}}{\partial\theta^2}t + \frac{\lambda^{(\varepsilon)\|}t}{(\alpha+\theta)^2} + \frac{2\alpha\lambda^{(\varepsilon)\|}t}{\theta^3}\mathbb{E}\Big[\log\left(X_1^\theta + Y_1^\theta\right)\Big] \\
&\quad -\frac{2\alpha\lambda^{(\varepsilon)\|}t}{\theta^2}\mathbb{E}\Big[\frac{\partial}{\partial\theta}\log\left(X_1^\theta + Y_1^\theta\right)\Big] + (2+\frac{\alpha}{\theta})\lambda^{(\varepsilon)\|}t\mathbb{E}\Big[\frac{\partial^2}{\partial\theta^2}\log\left(X_1^\theta + Y_1^\theta\right)\Big]\Bigg) \\
&= d\,\Bigg(\Big(\frac{\alpha\log 2}{\theta^2}\Big)^2 - \frac{2\alpha\log 2}{\theta^3} + \frac{1}{(\alpha+\theta)^2} + \frac{2\alpha}{\theta^3}\mathbb{E}\Big[\log\left(X_1^\theta + Y_1^\theta\right)\Big] \\
&\quad -\frac{2\alpha}{\theta^2}\mathbb{E}\Big[\frac{\partial}{\partial\theta}\log\left(X_1^\theta + Y_1^\theta\right)\Big] + (2+\frac{\alpha}{\theta})\mathbb{E}\Big[\frac{\partial^2}{\partial\theta^2}\log\left(X_1^\theta + Y_1^\theta\right)\Big]\Bigg)\,.
\end{aligned}
$$

Since only the first term of d_{32} in the bracket depends on ε, and d_{33} is independent of ε, the proof is complete by setting

$$
a(\alpha,\theta) = -\frac{\alpha(\log 2)^2}{\theta^3} + \frac{\log 2}{\theta^2} + \frac{1}{(\alpha+\theta)^2} - \frac{1}{\theta^2}\mathbb{E}\Big[\log(X_1^\theta + Y_1^\theta)\Big] + \frac{1}{\theta}\mathbb{E}\Big[\frac{\partial}{\partial\theta}\log(X_1^\theta + Y_1^\theta)\Big] \tag{4.4.7}
$$

$$
\begin{aligned}
b(\alpha,\theta) &= \Big(\frac{\alpha\log 2}{\theta^2}\Big)^2 - \frac{2\alpha\log 2}{\theta^3} + \frac{1}{(\alpha+\theta)^2} + \frac{2\alpha}{\theta^3}\mathbb{E}\Big[\log(X_1^\theta + Y_1^\theta)\Big] \\
&\quad -\frac{2\alpha}{\theta^2}\mathbb{E}\Big[\frac{\partial}{\partial\theta}\log(X_1^\theta + Y_1^\theta)\Big] + \frac{2\theta+\alpha}{\theta}\mathbb{E}\Big[\frac{\partial^2}{\partial\theta^2}\log(X_1^\theta + Y_1^\theta)\Big]
\end{aligned} \tag{4.4.8}
$$

□

Remark 4.4.2. *We shall need the following inverse:*

$$
D^{-1} = \begin{pmatrix} 1+\alpha^2(\log\varepsilon)^2 & \alpha^2\log\varepsilon & 0 \\ \alpha^2\log\varepsilon & \alpha^2 & 0 \\ -\frac{1}{b}\big(a\alpha^2\log\varepsilon + \frac{\alpha\log 2}{\theta^2}\big) & -\frac{a}{b}\alpha^2 & \frac{1}{d\,b} \end{pmatrix}, \tag{4.4.9}
$$

In order to calculate the matrix M below we shall need the following result on the dependence of n and $n^{\|}$.

Lemma 4.4.3. *Recall that* $n = n_1 + n_2 = 2n^{\|} + n_1^{\perp} + n_2^{\perp}$*. Then*

$$
\mathbb{E}[nn^{\|}] = 2\lambda^{(\varepsilon)\|}t(1+\lambda^{(\varepsilon)}t) \quad \textit{and} \quad \mathrm{cov}(n,n^{\|}) = 2\lambda^{(\varepsilon)\|}t\,.
$$

Proof. We calculate the expectation, the result for the covariance is then obvious. By independence of the Poisson processes of joint and single jumps,

$$\begin{aligned}\mathbb{E}\left[nn^{\|}\right] &= \mathbb{E}\left[(2n^{\|}+n_1^{\perp}+n_2^{\perp})n^{\|}\right] \\ &= 2\big(\mathrm{var}(n^{\|})+(\mathbb{E}[n^{\|}])^2\big)+\mathbb{E}[n_1^{\perp}+n_2^{\perp}]\mathbb{E}[n^{\|}] \\ &= 2(\lambda^{(\varepsilon)\|}t+(\lambda^{(\varepsilon)\|}t)^2)+(\lambda_1^{(\varepsilon)\perp}+\lambda_2^{(\varepsilon)\perp})\lambda^{(\varepsilon)\|}t^2 \\ &= 2\lambda^{(\varepsilon)\|}t(1+\lambda^{(\varepsilon)}t).\end{aligned}$$

Lemma 4.4.4. *Assume a bivariate α-stable Clayton subordinator with common marginal parameters* $(\log c, \alpha)$ *and dependence parameter $\theta = \alpha\delta$. Assume also the observation scheme given in Section 2.2. Define $d = \frac{\lambda^{(\varepsilon)\|}}{2\lambda^{(\varepsilon)}} = 2^{-\frac{\alpha}{\theta}-1}$ and*

$$T(x,y) := (\log x + \log y) + \frac{\alpha}{\theta^2}\log(x^\theta + y^\theta) - (2+\frac{\alpha}{\theta})\frac{\partial}{\partial\theta}\log(x^\theta+y^\theta)\,. \qquad (4.4.10)$$

Then the matrix M introduced in (4.4.3) *is given by*

$$M = \begin{pmatrix} 1 & -\log\varepsilon & 2d\frac{\alpha\log 2}{\theta^2} \\ -\log\varepsilon & \frac{1}{\alpha^2}+(\log\varepsilon)^2 & -d\Big(2\frac{\alpha\log 2}{\theta^2}\log\varepsilon + m\Big) \\ 2d\frac{\alpha\log 2}{\theta^2} & -d\Big(2\frac{\alpha\log 2}{\theta^2}\log\varepsilon + m\Big) & db \end{pmatrix}, \qquad (4.4.11)$$

where b is given by (4.4.8) *and*

$$m = 2\,\mathrm{cov}\Big(\log\frac{X_1}{\varepsilon}, T\Big(\frac{X_1}{\varepsilon},\frac{Y_1}{\varepsilon}\Big)\Big). \qquad (4.4.12)$$

Moreover, this covariance is independent of ε.

Proof. Recall the definition of the Z_i for $i = 1,\dots,n$ as in Step 1 of Section 4.2 and the fact that $\log\frac{Z_1}{\varepsilon},\dots,\log\frac{Z_n}{\varepsilon}$ are exponential random variables with expectation α^{-1}, and that in this first step they are treated as independent. The entries of the matrix $M = (m_{ij})_{1\le i,j\le 3}$ are calculated from the score functions (4.3.9) in Remark 4.3.4 as follows:

$$\begin{aligned} m_{11} &= \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[\Big(\frac{\partial l_{12}^{(\varepsilon)}}{\partial\log c}\Big)^2\Big] = \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[-\frac{\partial^2 l_{12}^{(\varepsilon)}}{\partial(\log c)^2}\Big] = d_{11} \\ m_{12} &= \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[\Big(\frac{\partial l_{12}^{(\varepsilon)}}{\partial\alpha}\Big)\Big(\frac{\partial l_{12}^{(\varepsilon)}}{\partial\log c}\Big)\Big] = \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[-\frac{\partial^2 l_{12}^{(\varepsilon)}}{\partial\alpha\partial\log c}\Big] = d_{12} = d_{21} = m_{21} \\ m_{22} &= \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[\Big(\frac{\partial l_{12}^{(\varepsilon)}}{\partial\alpha}\Big)^2\Big] = \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[-\frac{\partial^2 l_{12}^{(\varepsilon)}}{\partial\alpha^2}\Big] = d_{22}. \end{aligned}$$

We abbreviate $T_i := T(X_i, Y_i)$ and find from Lemma 4.4 in [18] that $\mu_T := \mathbb{E}(T_i) = \frac{\alpha \log 2}{\theta^2} - \frac{1}{\alpha+\theta}$. Then by (4.3.6) and (4.3.8) we find

$$\begin{aligned}
m_{13} &= \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[\Big(\frac{\partial l_{12}^{(\varepsilon)}}{\partial \log c}\Big)\Big(\frac{\partial l^{(\varepsilon)}}{\partial \theta}\Big)\Big] \\
&= \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[(n-2\lambda^{(\varepsilon)}t)\Big(-\lambda^{(\varepsilon)\|}t\frac{\alpha\log 2}{\theta^2} + \frac{n^{\|}}{\alpha+\theta} + \sum_{i=1}^{n^{\|}}(T_i-\mu_T) + n^{\|}(\frac{\alpha \log 2}{\theta^2} - \frac{1}{\alpha+\theta})\Big)\Big] \\
&= \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[(n-2\lambda^{(\varepsilon)}t)\Big(\frac{\alpha\log 2}{\theta^2}(n^{\|}-\lambda^{(\varepsilon)\|}t) + \sum_{i=1}^{n^{\|}}(T_i-\mu_T)\Big)\Big] \\
&= \frac{\alpha\log 2}{2\lambda^{(\varepsilon)}t\theta^2}\mathbb{E}\Big[n(n^{\|}-\lambda^{(\varepsilon)\|}t)\Big] + \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[\mathbb{E}\Big[n\sum_{i=1}^{n^{\|}}(T_i-\mu_T)\Big| n, n^{\|}\Big]\Big] \\
&= \frac{\alpha\log 2}{2\lambda^{(\varepsilon)}t\theta^2}\left(\mathbb{E}[nn^{\|}] - 2\lambda^{(\varepsilon)}\lambda^{(\varepsilon)\|}t^2\right) = \frac{\alpha\log 2}{2\lambda^{(\varepsilon)}t\theta^2}\mathrm{cov}(n,n^{\|}) = 2d\frac{\alpha\log 2}{\theta^2},
\end{aligned}$$

where we have used Lemma 4.4.3.

$$\begin{aligned}
m_{23} &= \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[\Big(\frac{\partial l_{12}^{(\varepsilon)}}{\partial \alpha}\Big)\Big(\frac{\partial l^{(\varepsilon)}}{\partial \theta}\Big)\Big] \\
&= -\frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[\Big(\sum_{i=1}^{n}(\log\frac{Z_i}{\varepsilon} - \frac{1}{\alpha}) + \log\varepsilon(n-2\lambda^{(\varepsilon)}t)\Big)\Big(\frac{\alpha\log 2}{\theta^2}(n^{\|}-\lambda^{(\varepsilon)\|}t) + \sum_{i=1}^{n^{\|}}(T_i-\mu_T)\Big)\Big] \\
&= -\frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[\Big(\sum_{i=1}^{n}(\log\frac{Z_i}{\varepsilon} - \frac{1}{\alpha})\Big)\Big(\sum_{i=1}^{n^{\|}}(T_i-\mu_T)\Big)\Big] \\
&\quad -\frac{\alpha\log 2}{2\lambda^{(\varepsilon)}t\theta^2}\log\varepsilon\,\mathbb{E}\Big[(n-2\lambda^{(\varepsilon)}t)(n^{\|}-\lambda^{(\varepsilon)\|}t)\Big]
\end{aligned}$$

Now note that the jumps $(X_i, Y_i)_{i=1,\dots,n^{\|}}$ are independent and independent of all single jumps in either component. Recall that $T_i = T(X_i, Y_i) = T(X_i/\varepsilon, Y_i/\varepsilon)$, where the last equality is easily checked. Hence, the right hand side above reduces to

$$\begin{aligned}
&= -\frac{1}{2\lambda^{(\varepsilon)}t}\,\mathbb{E}\Big[\sum_{i=1}^{n^{\|}}(\log\frac{X_i}{\varepsilon} + \log\frac{Y_i}{\varepsilon} - \frac{2}{\alpha})(T_i-\mu_T)\Big] - \frac{\alpha\log 2}{2\lambda^{(\varepsilon)}t\theta^2}\log\varepsilon\,\mathrm{cov}(n,n^{\|}) \\
&= -\frac{\lambda^{(\varepsilon)\|}t}{2\lambda^{(\varepsilon)}t}\,\mathbb{E}\Big[\Big(\log\frac{X_1}{\varepsilon} + \log\frac{Y_1}{\varepsilon} - \frac{2}{\alpha}\Big)\Big(T_i-\mu_T\Big)\Big] - \frac{2\lambda^{(\varepsilon)\|}t}{2\lambda^{(\varepsilon)}t}\frac{\alpha\log 2}{\theta^2}\log\varepsilon \\
&= -d\Big(m + \frac{2\alpha\log 2}{\theta^2}\log\varepsilon\Big).
\end{aligned}$$

Finally,

$$m_{33} = \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[\Big(\frac{\partial l^{(\varepsilon)}}{\partial\theta}\Big)^2\Big] = -\frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\Big[\frac{\partial^2 l^{(\varepsilon)}}{\partial\theta^2}\Big] = d_{33}. \quad \square$$

Now we can calculate the Godambe information matrix as in (4.4.1).

4.4.2 Consistency and asymptotic normality of the two-step estimators

Assume that the log-likelihood $l_{12}^{(\varepsilon)}(\log c, \alpha)$ in (4.3.5) is used for estimating the marginal parameters $\log c$ and α in the first step and the log-likelihood $l^{(\varepsilon)}(\log c, \alpha, \theta)$ in (4.3.7) for estimating the dependence parameter θ in the second step; i.e. we work with $\mathbf{J}^{(\varepsilon)}(\mathbf{X}, \mathbf{Y}; \eta)$ as given in Remark 4.3.4. As before we denote the resulting estimates by $\widetilde{\gamma} = (\log \widetilde{c}, \widetilde{\alpha})$ and $\widetilde{\eta} = (\log \widetilde{c}, \widetilde{\alpha}, \widetilde{\theta})$. Assume further that $\gamma_0 = (\log c_0, \alpha_0)$ and $\eta_0 = (\log c_0, \alpha_0, \theta_0)$ are the true parameter vectors.

A Taylor expansion of the score functions in (4.3.6) and (4.3.8) implies

$$
\begin{aligned}
\frac{\partial l_{12}^{(\varepsilon)}(\log c,\alpha)}{\partial \log c}\bigg|_{\gamma=\widetilde{\gamma}} &= \frac{\partial l_{12}^{(\varepsilon)}(\gamma)}{\partial \log c}\bigg|_{\gamma=\gamma_0} + (\log \widetilde{c} - \log c_0)\, \frac{\partial^2 l_{12}^{(\varepsilon)}(\gamma)}{\partial (\log c)^2}\bigg|_{\gamma=\gamma^*} + (\widetilde{\alpha} - \alpha_0)\, \frac{\partial^2 l_{12}^{(\varepsilon)}(\gamma)}{\partial \alpha \partial \log c}\bigg|_{\gamma=\gamma^*} \\
\frac{\partial l_{12}^{(\varepsilon)}(\log c,\alpha)}{\partial \alpha}\bigg|_{\gamma=\widetilde{\gamma}} &= \frac{\partial l_{12}^{(\varepsilon)}(\gamma)}{\partial \alpha}\bigg|_{\gamma=\gamma_0} + (\log \widetilde{c} - \log c_0)\, \frac{\partial^2 l_{12}^{(\varepsilon)}(\gamma)}{\partial \log c \partial \alpha}\bigg|_{\gamma=\gamma^*} + (\widetilde{\alpha} - \alpha_0)\, \frac{\partial^2 l_{12}^{(\varepsilon)}(\gamma)}{\partial \alpha^2}\bigg|_{\gamma=\gamma^*} \\
\frac{\partial l^{(\varepsilon)}(\log c,\alpha,\theta)}{\partial \theta}\bigg|_{\eta=\widetilde{\eta}} &= \frac{\partial l^{(\varepsilon)}(\eta)}{\partial \theta}\bigg|_{\eta=\eta_0} + (\log \widetilde{c} - \log c_0)\frac{\partial^2 l^{(\varepsilon)}(\eta)}{\partial \log c\ \partial \theta}\bigg|_{\eta=\eta^{**}} + (\widetilde{\alpha} - \alpha_0)\frac{\partial^2 l^{(\varepsilon)}(\eta)}{\partial \alpha \partial \theta}\bigg|_{\eta=\eta^{**}} \\
&\quad + (\widetilde{\theta} - \theta_0)\, \frac{\partial^2 l^{(\varepsilon)}(\eta)}{\partial \theta^2}\bigg|_{\eta=\eta^{**}}
\end{aligned}
\tag{4.4.13}
$$

where the vector $\gamma^* = (\log c^*, \alpha^*)$ is between $\widetilde{\gamma} = (\log \widetilde{c}, \widetilde{\alpha})$ and $\gamma_0 = (\log c_0, \alpha_0)$ and the vector $\eta^{**} = (\log c^{**}, \alpha^{**}, \theta^{**})$ is between $\widetilde{\eta}$ and η_0, componentwise.

Since the left hand side of the equations in (4.4.13) are zero, so are the equations on the right hand side. Therefore, these equations can be written in matrix form as follows

$$\begin{pmatrix} \frac{\partial^2 l_{12}^{(\varepsilon)}(\log c,\alpha)}{\partial(\log c)^2}\Big|_{\gamma=\gamma^*} & \frac{\partial^2 l_{12}^{(\varepsilon)}(\log c,\alpha)}{\partial\alpha\partial\log c}\Big|_{\gamma=\gamma^*} & 0 \\ \frac{\partial^2 l_{12}^{(\varepsilon)}(\log c,\alpha)}{\partial\log c\ \partial\alpha}\Big|_{\gamma=\gamma^*} & \frac{\partial^2 l_{12}^{(\varepsilon)}(\log c,\alpha)}{\partial\alpha^2}\Big|_{\gamma=\gamma^*} & 0 \\ \frac{\partial^2 l^{(\varepsilon)}(\log c,\alpha,\theta)}{\partial\log c\partial\theta}\Big|_{\eta=\eta^{**}} & \frac{\partial^2 l^{(\varepsilon)}(\log c,\alpha,\theta)}{\partial\alpha\partial\theta}\Big|_{\eta=\eta^{**}} & \frac{\partial^2 l^{(\varepsilon)}(\log c,\alpha,\theta)}{\partial\theta^2}\Big|_{\eta=\eta^{**}} \end{pmatrix} \times \begin{pmatrix} \log\widetilde{c} - \log c_0 \\ \widetilde{\alpha} - \alpha_0 \\ \widetilde{\theta} - \theta_0 \end{pmatrix}$$

$$= - \begin{pmatrix} \frac{\partial l_{12}^{(\varepsilon)}(\log c,\alpha)}{\partial\log c}\Big|_{\gamma=\gamma_0} \\ \frac{\partial l_{12}^{(\varepsilon)}(\log c,\alpha)}{\partial\alpha}\Big|_{\gamma=\gamma_0} \\ \frac{\partial l^{(\varepsilon)}(\log c,\alpha,\theta)}{\partial\theta}\Big|_{\eta=\eta_0} \end{pmatrix}. \tag{4.4.14}$$

We denote by $H^{(\varepsilon)}$ the matrix of the second-order derivatives of the log-likelihoods in (4.4.14) at $\eta = (\log c, \alpha, \theta)$ and by $H^{(\varepsilon)*}$ the matrix of $H^{(\varepsilon)}$ at γ^* and η^{**} as in (4.4.14). (Note that $D = \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}[-H^{(\varepsilon)}]$.) Moreover, we recall the vector of the score functions $\mathbf{J}^{(\varepsilon)}(\eta)$ from Remark 4.3.4; then the right hand side of equation (4.4.14) is equal to $-\mathbf{J}^{(\varepsilon)}(\eta)\big|_{\eta=\eta_0}$.

Furthermore, equation (4.4.14) becomes

$$H^{(\varepsilon)*}\ (\widetilde{\eta} - \eta_0) = -\mathbf{J}^{(\varepsilon)}(\eta)\Big|_{\eta=\eta_0}. \tag{4.4.15}$$

The common marginal log-likelihood in (4.3.5) and the log-likelihood in (4.3.7) have the score functions (4.3.6) and (4.3.8), respectively. They can be rewritten as

$$\frac{\partial l_{12}^{(\varepsilon)}(\log c, \alpha)}{\partial \log c} = 2\lambda^{(\varepsilon)}t\,\Big(\frac{\widehat{\lambda}^{(\varepsilon)}}{\lambda^{(\varepsilon)}} - 1\Big) \tag{4.4.16}$$

$$\frac{\partial l_{12}^{(\varepsilon)}(\log c, \alpha)}{\partial \alpha} = 2\lambda^{(\varepsilon)}t\,\log\varepsilon\,\Big(\frac{\widehat{\lambda}^{(\varepsilon)}}{\lambda^{(\varepsilon)}} - 1\Big) - \sum_{i=1}^{n}\Big(\log(\frac{Z_i}{\varepsilon}) - \frac{1}{\alpha}\Big) \tag{4.4.17}$$

$$\begin{aligned} \frac{\partial l^{(\varepsilon)}(\log c, \alpha, \theta)}{\partial \theta} &= \sum_{i=1}^{n^{\|}} T_i + \frac{n^{\|}}{\alpha+\theta} - \lambda^{(\varepsilon)\|}t\,\frac{\alpha\log 2}{\theta^2} \\ &= \sum_{i=1}^{n^{\|}}(T_i - \mu_T) + \frac{\alpha\log 2}{\theta^2}\lambda^{(\varepsilon)\|}t\Big(\frac{\widehat{\lambda}^{(\varepsilon)\|}}{\lambda^{(\varepsilon)\|}} - 1\Big). \end{aligned} \tag{4.4.18}$$

The next result shows the consistency of the estimator $\widetilde{\eta}$.

Proposition 4.4.5. *Assume the bivariate α-stable Clayton subordinator with common marginal parameters $(\log c, \alpha)$ and dependence parameter $\theta = \alpha\delta$. Assume also the observation scheme as described in Section 2.4. Let b be defined as in (4.4.8). If $b(\alpha_0, \theta_0) \neq 0$, then the two-step estimator $\widetilde{\eta}$ is a consistent estimator; i.e., as $n^{\parallel} \to \infty$ (then also $n \to \infty$) for a fixed ε,*

$$\widetilde{\eta} \xrightarrow{P} \eta_0.$$

Proof. We divide (4.4.15) by n and obtain

$$\frac{1}{n} H^{(\varepsilon)*}(\widetilde{\eta} - \eta_0) = -\frac{1}{n}\mathbf{J}^{(\varepsilon)}(\eta)\Big|_{\eta=\eta_0}. \tag{4.4.19}$$

From the equations in (4.4.16), (4.4.17) and (4.4.18) the vector on the right hand side of (4.4.19) is given by (note that $\widehat{\lambda}^{(\varepsilon)} = \frac{n}{2t}$ estimates $\lambda^{(\varepsilon)}$)

$$\frac{1}{n}\mathbf{J}^{(\varepsilon)}(\eta)\Big|_{\eta=\eta_0} = (1 + o_P(1)) \begin{pmatrix} \frac{\widehat{\lambda}^{(\varepsilon)}}{\lambda^{(\varepsilon)}} - 1 \\ \log\varepsilon\left(\frac{\widehat{\lambda}^{(\varepsilon)}}{\lambda^{(\varepsilon)}} - 1\right) - \frac{1}{2\lambda^{(\varepsilon)}t}\sum_{i=1}^{n}\left(\log(\frac{Z_i}{\varepsilon}) - \frac{1}{\alpha}\right) \\ \frac{1}{2\lambda^{(\varepsilon)}t}\sum_{i=1}^{n^{\parallel}}(T_i - \mu_T) + \frac{\alpha\log 2}{\theta^2}\frac{\lambda^{(\varepsilon)\parallel}}{2\lambda^{(\varepsilon)}}\left(\frac{\widehat{\lambda}^{\varepsilon)\parallel}}{\lambda^{(\varepsilon)\parallel}} - 1\right) \end{pmatrix}_{\eta=\eta_0},$$

where we have used Slutzky's theorem together with the LLN, which implies that $\frac{n}{2\lambda^{(\varepsilon)}t} \xrightarrow{P} 1$. We shall show that all vector components tend to 0 as $n^{\parallel} \to \infty$ in probability. It is a well-known fact that $\sqrt{n}(\frac{\widehat{\lambda}^{(\varepsilon)}}{\lambda^{(\varepsilon)}} - 1) \xrightarrow{d} N(0,1)$. By Slutzky's theorem again, $\log\varepsilon(\frac{\widehat{\lambda}^{(\varepsilon)}}{\lambda^{(\varepsilon)}} - 1) = (1 + o_P(1))\frac{\log\varepsilon}{\sqrt{2tc\varepsilon^{-\alpha}}}\sqrt{n}(\frac{\widehat{\lambda}^{(\varepsilon)}}{\lambda^{(\varepsilon)}} - 1) \xrightarrow{P} 0$. This implies that the first component as well as the first term in the second component tend to 0. Since $\mathbb{E}\left[\log(\frac{Z_i}{\varepsilon})\right] = \frac{1}{\alpha}$, also the second term tends to 0 by the LLN. For the third component it suffices to note that $\frac{\lambda^{(\varepsilon)\parallel}}{2\lambda^{(\varepsilon)}} = 2^{-\frac{\alpha}{\theta}-1}$, then it tends to 0 also as a consequence of the LLN and the fact that $\frac{\widehat{\lambda}^{(\varepsilon)\parallel}}{\lambda^{(\varepsilon)\parallel}} \xrightarrow{P} 1$.

We show now that the matrix $\frac{1}{n}H^{(\varepsilon)*}$ is an invertible matrix and does not converge to a zero-matrix as $n^{\parallel} \to \infty$, which proves consistency of $\widetilde{\eta}$. We show, first, that the limit of $\frac{1}{n}H^{(\varepsilon)}$ is deterministic and independent of t as $n^{\parallel} \to \infty$. From (4.4.5) we can read off the upper left 2×2-matrix, and from (4.4.6) we obtain the last line of $H^{(\varepsilon)}$. We denote again $d = \frac{\lambda^{(\varepsilon)\parallel}}{2\lambda^{(\varepsilon)}} = 2^{-\frac{\alpha}{\theta}-1}$, then

$$H^{(\varepsilon)} = 2\lambda^{(\varepsilon)}t \begin{pmatrix} -1 & \log\varepsilon & 0 \\ \log\varepsilon & -\frac{n}{\alpha^2 2\lambda^{(\varepsilon)}t} - (\log\varepsilon)^2 & 0 \\ -d\,\frac{\alpha\log 2}{\theta^2} & d\big(\frac{\alpha\log 2}{\theta^2}\log\varepsilon - A\big) & -dB \end{pmatrix} \tag{4.4.20}$$

where

$$\begin{aligned} A &:= -\frac{\alpha(\log 2)^2}{\theta^3} + \frac{\log 2}{\theta^2} + \frac{n^{\|}}{\lambda^{(\varepsilon)\|}t(\alpha+\theta)^2} - \frac{1}{\lambda^{(\varepsilon)\|}t\theta^2}\sum_{i=1}^{n^{\|}}\log(X_i^\theta + Y_i^\theta) \\ &\quad + \frac{1}{\lambda^{(\varepsilon)\|}t\theta}\sum_{i=1}^{n^{\|}}\frac{\partial}{\partial\theta}\log(X_i^\theta + Y_i^\theta), \\ B &:= \Big(\frac{\alpha\log 2}{\theta^2}\Big)^2 - \frac{2\alpha\log 2}{\theta^3} + \frac{n^{\|}}{\lambda^{(\varepsilon)\|}t(\alpha+\theta)^2} + \frac{2\alpha}{\lambda^{(\varepsilon)\|}t\theta^3}\sum_{i=1}^{n^{\|}}\log(X_i^\theta + Y_i^\theta) \\ &\quad - \frac{2\alpha}{\lambda^{(\varepsilon)\|}t\theta^2}\sum_{i=1}^{n^{\|}}\frac{\partial}{\partial\theta}\log(X_i^\theta + Y_i^\theta) + \frac{2\theta+\alpha}{\lambda^{(\varepsilon)\|}t\theta}\sum_{i=1}^{n^{\|}}\frac{\partial^2}{\partial\theta^2}\log(X_i^\theta + Y_i^\theta). \end{aligned}$$

Note that the distributions of A and B do not depend on t and ε (replacing X_i and Y_i by X_i/ε and Y_i/ε does not change A and B, respectively). It remains to show that the matrix $\frac{1}{n}H^{(\varepsilon)*}$ is invertible.

From (4.4.20) the determinant of matrix $\frac{1}{n}H^{(\varepsilon)}$ for an arbitrary parameter η is given by

$$\det\Big(\frac{1}{n}H^{(\varepsilon)}\Big) = (1 + o_P(1))\Big(-\frac{2^{-\alpha/\theta-1}B}{\alpha^2}\Big).$$

Since by the LLN $\frac{n}{2\lambda^{(\varepsilon)}t} \xrightarrow{P} 1$ and $B \xrightarrow{P} b(\alpha,\theta) \neq 0$, the determinant does not converge to zero as $n \to \infty$, i.e. $t \to \infty$. That is, the matrix $\frac{1}{n}H^{(\varepsilon)*}$ is invertible and does not converge to a zero matrix and this completes the proof. □

Remark 4.4.6. *Note that* $\mathbb{E}[A] = a = a(\alpha,\theta)$ *and* $\mathbb{E}[B] = b = b(\alpha,\theta)$ *as defined in* (4.4.7) *and* (4.4.8). *As a consequence of the LLN in combination with Slutzky's theorem we have that* $A \xrightarrow{P} a$ *and* $B \xrightarrow{P} b$.

We are now ready to formulate the main result of our paper.

Theorem 4.4.7. *Assume a bivariate α-stable Clayton subordinator with common marginal parameters $(\log c, \alpha)$ and dependence parameter $\theta = \alpha\delta$. Assume also the observation scheme as described in Section 2.4. Let a and b be defined as in (4.4.7) and (4.4.8), respectively, and let m be as in (4.4.12). If $b(\alpha, \theta) \neq 0$, then as $\varepsilon \to 0$,*

$$\sqrt{2c\varepsilon^{-\alpha}t}\begin{pmatrix} \frac{\log \widetilde{c} - \log c}{\log \varepsilon} \\ \widetilde{\alpha} - \alpha \\ \widetilde{\theta} - \theta \end{pmatrix} \xrightarrow{d} N_3\left(\mathbf{0}, V\right), \tag{4.4.21}$$

where

$$V = \begin{pmatrix} \alpha^2 & \alpha^2 & -\frac{\alpha^2(a+m)}{b} \\ \alpha^2 & \alpha^2 & -\frac{\alpha^2(a+m)}{b} \\ -\frac{\alpha^2(a+m)}{b} & -\frac{\alpha^2(a+m)}{b} & \frac{1}{bd} - \frac{3\alpha^2(\log 2)^2}{b^2\theta^4} + \frac{a\alpha^2(a+2m)}{b^2} \end{pmatrix}. \tag{4.4.22}$$

Proof. We start with the left hand side of equation in (4.4.19), where η_0 is the true parameter vector and we write η^* for the "combination" of γ^* and η^{**}. Then

$$\begin{aligned} &\frac{1}{n}H^{(\varepsilon)*}(\widetilde{\eta} - \eta_0) \\ &= \frac{2\lambda^{(\varepsilon)^*}t}{n}\begin{pmatrix} -1 & \log\varepsilon & 0 \\ \log\varepsilon & -\frac{n}{\alpha^2 2\lambda^{(\varepsilon)}t} - (\log\varepsilon)^2 & 0 \\ -d\,\frac{\alpha\log 2}{\theta^2} & d\big(\frac{\alpha\log 2}{\theta^2}\log\varepsilon - A\big) & -dB \end{pmatrix}\Bigg|_{\eta=\eta^*} \times \begin{pmatrix} \log\widetilde{c} - \log c_0 \\ \widetilde{\alpha} - \alpha_0 \\ \widetilde{\theta} - \theta_0 \end{pmatrix} \\ &= \frac{2\lambda^{(\varepsilon)^*}t}{n}\begin{pmatrix} -\log\varepsilon & \log\varepsilon & 0 \\ (\log\varepsilon)^2 & -\frac{n}{\alpha^2 2\lambda^{(\varepsilon)}t} - (\log\varepsilon)^2 & 0 \\ -\frac{d\alpha\log 2}{\theta^2}\log\varepsilon & d\big(\frac{\alpha\log 2}{\theta^2}\log\varepsilon - A\big) & -dB \end{pmatrix}\Bigg|_{\eta=\eta^*} \times \begin{pmatrix} \frac{\log\widetilde{c} - \log c_0}{\log\varepsilon} \\ \widetilde{\alpha} - \alpha_0 \\ \widetilde{\theta} - \theta_0 \end{pmatrix}. \end{aligned}$$

where $\lambda^{(\varepsilon)^*} := \lambda^{(\varepsilon)}|_{\eta^*}$.

Multiplying both sides of (4.4.19) by $\sqrt{n}$ and using $\frac{n}{2\lambda^{(\varepsilon)}t} \xrightarrow{P} 1$ yields

$$\begin{aligned} &\sqrt{n}\left(\frac{\log\widetilde{c} - \log c_0}{\log\varepsilon}, \widetilde{\alpha} - \alpha_0, \widetilde{\theta} - \theta_0\right)^\top \\ &= -\big(\frac{n}{2\lambda^{(\varepsilon)^*}t}\big)\begin{pmatrix} -\log\varepsilon & \log\varepsilon & 0 \\ (\log\varepsilon)^2 & -\frac{n}{\alpha^2 2\lambda^{(\varepsilon)}t} - (\log\varepsilon)^2 & 0 \\ -\frac{d\alpha\log 2}{\theta^2}\log\varepsilon & d\big(\frac{\alpha\log 2}{\theta^2}\log\varepsilon - A\big) & -dB \end{pmatrix}^{-1}_{\eta=\eta^*} \times \frac{1}{\sqrt{n}}\mathbf{J}^{(\varepsilon)}(\eta)\Big|_{\eta=\eta_0} \end{aligned}$$

$$
\begin{aligned}
&= (\frac{n}{2\lambda^{(\varepsilon)*}t}) \begin{pmatrix} \frac{1}{\log\varepsilon} + (\frac{2\lambda^{(\varepsilon)}t}{n})\alpha^2\log\varepsilon & (\frac{2\lambda^{(\varepsilon)}t}{n})\alpha^2 & 0 \\ (\frac{2\lambda^{(\varepsilon)}t}{n})\alpha^2\log\varepsilon & (\frac{2\lambda^{(\varepsilon)}t}{n})\alpha^2 & 0 \\ -\frac{A\alpha^2}{B}(\frac{2\lambda^{(\varepsilon)}t}{n})\log\varepsilon - \frac{\alpha\log 2}{\theta^2 B} & -\frac{A\alpha^2}{B}(\frac{2\lambda^{(\varepsilon)}t}{n}) & \frac{1}{dB} \end{pmatrix}_{\eta=\eta^*} \times \frac{1}{\sqrt{n}}\mathbf{J}^{(\varepsilon)}(\eta)\Big|_{\eta=\eta_0} \\
&= \begin{pmatrix} \frac{1}{\log\varepsilon}(\frac{n}{2\lambda^{(\varepsilon)}t}) + \alpha^2\log\varepsilon & \alpha^2 & 0 \\ \alpha^2\log\varepsilon & \alpha^2 & 0 \\ -\frac{A\alpha^2}{B}\log\varepsilon - \frac{\alpha\log 2}{\theta^2 B}(\frac{n}{2\lambda^{(\varepsilon)}t}) & -\frac{A\alpha^2}{B} & \frac{1}{dB} \end{pmatrix}_{\eta=\eta^*} \times \frac{1}{\sqrt{n}}\mathbf{J}^{(\varepsilon)}(\eta)\Big|_{\eta=\eta_0} \\
&= (1+o_P(1)) \begin{pmatrix} \frac{1}{\log\varepsilon} + \alpha^2\log\varepsilon & \alpha^2 & 0 \\ \alpha^2\log\varepsilon & \alpha^2 & 0 \\ -\frac{A\alpha^2}{B}\log\varepsilon - \frac{\alpha\log 2}{\theta^2 B} & -\frac{A\alpha^2}{B} & \frac{1}{dB} \end{pmatrix}_{\eta=\eta^*} \times \frac{1}{\sqrt{n}}\mathbf{J}^{(\varepsilon)}(\eta)\Big|_{\eta=\eta_0} \\
&= H_1^{(\varepsilon)^{-1}}\Big|_{\eta=\eta^*} \times \frac{1}{\sqrt{n}}\mathbf{J}^{(\varepsilon)}(\eta)\Big|_{\eta=\eta_0}. \qquad (4.4.23)
\end{aligned}
$$

The vector $\frac{1}{\sqrt{n}}\mathbf{J}^{(\varepsilon)}(\eta)$ is a zero-mean vector with covariance matrix $M = \frac{1}{2\lambda^{(\varepsilon)}t}\mathbb{E}\big[\frac{\partial l^{(\varepsilon)}(\eta)}{\partial\eta}\frac{\partial l^{(\varepsilon)}(\eta)}{\partial\eta}^T\big]$ calculated in Lemma 4.4.4. Since γ^* lies between $\widetilde{\gamma}$ and γ_0 and η^{**} between $\widetilde{\eta}$ and η_0, from the consistency of the estimators in Proposition 4.4.5, η^* can in the limit be replaced by η_0. Then, according to equation (10.6) in Joe [23], for $n\to\infty$ the asymptotic covariance matrix of $\sqrt{n}\left(\frac{\log\widetilde{c}-\log c_0}{\log\varepsilon}, \widetilde{\alpha}-\alpha_0, \widetilde{\theta}-\theta_0\right)^\top$ is given by

$$
\begin{aligned}
G_\varepsilon^{-1} &= D_1^{-1}M(D_1^{-1})^\top\Big|_{\eta=\eta_0} \\
&= \begin{pmatrix} \frac{1}{(\log\varepsilon)^2} + \alpha_0^2 & \alpha_0^2 & -\frac{(a_0+m_0)\alpha_0^2}{b_0} + \frac{\alpha_0\log 2}{b_0\theta_0^2\log\varepsilon} \\ \alpha_0^2 & \alpha_0^2 & -\frac{(a_0+m_0)\alpha_0^2}{b_0} \\ -\frac{(a_0+m_0)\alpha_0^2}{b_0} + \frac{\alpha_0\log 2}{b_0\theta_0^2\log\varepsilon} & -\frac{(a_0+m_0)\alpha_0^2}{b_0} & \frac{1}{b_0d_0} - \frac{3\alpha_0^2(\log 2)^2}{b_0^2\theta_0^4} + \frac{a_0\alpha_0^2(a_0+2m_0)}{b_0^2} \end{pmatrix}.
\end{aligned}
$$

where D_1 is the componentwise mean of the matrix $H_1^{(\varepsilon)}$ presented in (4.4.23).

Now note that by the LLN $\frac{n}{2\lambda^{(\varepsilon)}t} = \frac{n}{2c\varepsilon^{-\alpha}t} \xrightarrow{P} 1$ as $n\to\infty$ (either $\varepsilon\downarrow 0$ or $t\to\infty$), hence the rate $\sqrt{n}$ can be replaced by $\sqrt{2ct\varepsilon^{-\alpha}}$. Finally, $G_\varepsilon^{-1}\to V_0$ as $\varepsilon\to 0$ and this completes the proof. □

Remark 4.4.8. *(i) Note that for $t\to\infty$ and fixed $\varepsilon>0$ the asymptotic covariance matrix in* (4.4.21) *is given by G_ε^{-1}.*

(ii) The above theorem implies that the normal limit vector has representation

$$(N_1, N_1, N_2)^\top,$$

where N_1 *has variance* α^2, N_2 *has variance* $\frac{1}{bd} + \frac{\alpha^2}{b^2}\Big(-\frac{3(\log 2)^2}{\theta^4} + a(a+2m)\Big)$, *and the correlation between* N_1 *and* N_2 *is given by*

$$\operatorname{corr}(N_1, N_2) = -\frac{a+m}{\sqrt{\frac{1}{\alpha^2 d} - \frac{3(\log 2)^2}{b\theta^4} + \frac{a}{b}(a+2m)}}.$$

Example 4.4.9. *[Asymptotic covariance matrix for a bivariate* α*-stable Clayton subordinator]*

Let $\mathbf{S} = (\mathbf{S}(t))_{t\geq 0}$ *be a bivariate* α*-stable subordinator with a Clayton Lévy copula as introduced in Example 4.2.2. Assume further its parameters* $c_1 = c_2 = c$, $\alpha_1 = \alpha_2 = \alpha$ *and* $\theta = \alpha\delta$ *are estimated by a two-step method as in Section 4.2. The asymptotic covariance matrix as* $\varepsilon \to 0$ *for the model with parameter values* $c = 1$, $\alpha = 0.5$, $\theta = 1$ *can be computed numerically similar to the calculation at the end of Section 5 in [18]. Note that it involves the numerical integration of certain integrals. We find the asymptotic covariance matrix of* $\widetilde{\eta} = (\log \widetilde{c}, \widetilde{\alpha}, \widetilde{\theta})$ *as*

$$V = \begin{bmatrix} 0.25 & 0.25 & 0.1042 \\ 0.25 & 0.25 & 0.1042 \\ 0.1042 & 0.1042 & 2.6273 \end{bmatrix}.$$

Alternatively, this matrix can also be estimated replacing the numerical integration by a Monte Carlo simulation. For this the expectations in (4.4.7) *and* (4.4.8) *and the covariance* m *from* (4.4.12) *are empirically estimated by generating bivariate observations from* (4.3.4) *with parameter values mentioned above. Based on* 10^6 *bivariate observations, this yields the same asymptotic covariance matrix as above (4 leading decimals coindice).*

From this, we calculate $\operatorname{corr}(N_1, N_2) = 0.1286$.

Remark 4.4.10. *For the bivariate* α*-stable Clayton Lévy process with equal marginal processes we have been able to calculate the Godambe information matrix analytically. However, for most models this is too complicated. It requires derivatives of first and second* *order, the integration of some compound functions and the inverses and multiplications of possibly high dimensional matrices. As an alternative, a jackknife resampling method has been suggested and can also be applied in this context for arbitrary Lévy processs; cf. Joe [23], Section 10.1, and references given there.*

4.5 Maximum likelihood estimation of the full model

We compare the two-step procedure presented in Section 2.4 with two alternatives. Firstly, we consider the estimation method presented in [18] based on only common jumps. Secondly, we also compare this method with the full likelihood, based on single and common jumps. For this reason we present here the likelihood function of the full model. The observation scheme is as explained in Section 2.4, where $n^{\parallel} + n_1^{\perp} + n_2^{\perp}$ is the number of observations.

From (4.2.6) the Lévy densities of $\Pi_1^{(\varepsilon)\perp}$, $\Pi_2^{(\varepsilon)\perp}$ and $\Pi^{(\varepsilon)\parallel}$ are given by

$$\begin{aligned}
\nu_1^{\perp}(x) &= c\alpha x^{-\alpha-1}\left[1-\left(1+(x/\varepsilon)^{-\alpha\delta}\right)^{-1/\delta-1}\right], \quad x>\varepsilon\\
\nu_2^{\perp}(y) &= c\alpha y^{-\alpha-1}\left[1-\left(1+(y/\varepsilon)^{-\alpha\delta}\right)^{-1/\delta-1}\right], \quad y>\varepsilon\,,\\
\nu^{\parallel}(x,y) &= c\alpha^2(1+\delta)(xy)^{\alpha\delta-1}\left(x^{\alpha\delta}+y^{\alpha\delta}\right)^{-1/\delta-2}.
\end{aligned}$$

As intensities we obtain from (4.2.7) $\lambda^{(\varepsilon)\parallel} = c2^{-1/\delta}\varepsilon^{-\alpha}$. Moreover, the marginal intensities are $\lambda_1^{(\varepsilon)} = \lambda_2^{(\varepsilon)} = c\varepsilon^{-\alpha}$, so that $\lambda_1^{(\varepsilon)\perp} = \lambda_2^{(\varepsilon)\perp} = c\varepsilon^{-\alpha}(1-2^{-1/\delta})$. This implies the intensity of the bivariate CPP $\rho^{(\varepsilon)} := \lambda^{(\varepsilon)\parallel} + \lambda_1^{(\varepsilon)\perp} + \lambda_2^{(\varepsilon)\perp} = c\varepsilon^{-\alpha}(2-2^{-1/\delta})$.

For simplicity we reparameterize the model again as in Section 4.2 by setting $\alpha\delta = \theta$ and take $\log c$ instead of c as second marginal parameter. Now we recall Th. 4.1 of [17] for a bivariate CPP and obtain the likelihood function; here $(x_i, y_i)_{i=1,\dots,n^{\parallel}}$ denote the common jumps in both components and $\widetilde{x}_i$ for $i = 1,\dots,n_1^{\perp}$ and $\widetilde{y}_i$ for $i = 1,\dots,n_2^{\perp}$

denote the single jumps. The likelihood function of the bivariate CPP is then given by

$$\begin{aligned}
L^{(\varepsilon)}(\log c,\alpha,\theta) &= \Big(e^{-\rho^{(\varepsilon)}t}\prod_{i=1}^{n^{\|}}\nu^{\|}(x_i,y_i)\Big)\times\Big(e^{-\lambda_1^{(\varepsilon)\perp}t}\prod_{i=1}^{n_1^{\perp}}\nu_1^{\perp}(\widetilde{x}_i)\Big)\times\Big(e^{-\lambda_2^{(\varepsilon)\perp}t}\prod_{i=1}^{n_2^{\perp}}\nu(\widetilde{y}_i)\Big)\\
&= e^{-ct\varepsilon^{-\alpha}(2-2^{-\alpha/\theta})}(\alpha+\theta)^{n^{\|}}(\alpha c)^{n^{\|}+n_1^{\perp}+n_2^{\perp}}\prod_{i=1}^{n^{\|}}\left[(x_iy_i)^{\theta-1}(x_i^{\theta}+y_i^{\theta})^{-\alpha/\theta-2}\right]\\
&\quad\times\prod_{i=1}^{n_1^{\perp}}\left[\widetilde{x}_i^{-\alpha-1}\Big(1-\big(1+(\widetilde{x}_i/\varepsilon)^{-\theta}\big)^{-\alpha/\theta-1}\Big)\right] \qquad (4.5.1)\\
&\quad\times\prod_{i=1}^{n_2^{\perp}}\left[\widetilde{y}_i^{-\alpha-1}\Big(1-\big(1+(\widetilde{y}_i/\varepsilon)^{-\theta}\big)^{-\alpha/\theta-1}\Big)\right].
\end{aligned}$$

The log-likelihood is given by

$$\begin{aligned}
l^{(\varepsilon)}(\log c,\alpha,\theta) &= -ct\varepsilon^{-\alpha}(2-2^{-\alpha/\theta})+n^{\|}\log(\alpha+\theta)+(n^{\|}+n_1^{\perp}+n_2^{\perp})(\log\alpha+\log c)\\
&\quad+(\theta-1)\sum_{i=1}^{n^{\|}}(\log x_i+\log y_i)-(2+\frac{\alpha}{\theta})\sum_{i=1}^{n^{\|}}\log(x_i^{\theta}+y_i^{\theta})\\
&\quad-(\alpha+1)\sum_{i=1}^{n_1^{\perp}}\log\widetilde{x}_i+\sum_{i=1}^{n_1^{\perp}}\log\left[1-\big(1+(\widetilde{x}_i/\varepsilon)^{-\theta}\big)^{-\alpha/\theta-1}\right]\\
&\quad-(\alpha+1)\sum_{i=1}^{n_2^{\perp}}\log\widetilde{y}_i+\sum_{i=1}^{n_2^{\perp}}\log\left[1-\big(1+(\widetilde{y}_i/\varepsilon)^{-\theta}\big)^{-\alpha/\theta-1}\right].
\end{aligned}$$

For the score functions we obtain

$$\begin{aligned}
\frac{\partial l^{(\varepsilon)}}{\partial\log c} &= -ct\varepsilon^{-\alpha}(2-2^{-\alpha/\theta})+\frac{n^{\|}+n_1^{\perp}+n_2^{\perp}}{c}\\
\frac{\partial l^{(\varepsilon)}}{\partial\alpha} &= ct\varepsilon^{-\alpha}\Big(2\log\varepsilon-2^{-\alpha/\theta}\log\varepsilon-\frac{2^{-\alpha/\theta}\log 2}{\theta}\Big)+\frac{n^{\|}}{\alpha+\theta}+\frac{n^{\|}+n_1^{\perp}+n_2^{\perp}}{\alpha}\\
&\quad-\frac{1}{\theta}\sum_{i=1}^{n^{\|}}\log(x_i^{\theta}+y_i^{\theta})-\sum_{i=1}^{n_1^{\perp}}\log\widetilde{x}_i+\sum_{i=1}^{n_1^{\perp}}\frac{\partial}{\partial\alpha}\log\left[1-\big(1+(\widetilde{x}_i/\varepsilon)^{-\theta}\big)^{-\alpha/\theta-1}\right]\\
&\quad-\sum_{i=1}^{n_2^{\perp}}\log\widetilde{y}_i+\sum_{i=1}^{n_2^{\perp}}\frac{\partial}{\partial\alpha}\log\left[1-\big(1+(\widetilde{y}_i/\varepsilon)^{-\theta}\big)^{-\alpha/\theta-1}\right]
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial l^{(\varepsilon)}}{\partial \theta} &= \frac{ct\alpha\varepsilon^{-\alpha}2^{-\alpha/\theta}\log 2}{\theta^2} + \frac{n^{\|}}{\alpha+\theta} + \sum_{i=1}^{n^{\|}}(\log x_i + \log y_i) + \frac{\alpha}{\theta^2}\sum_{i=1}^{n^{\|}}\log(x_i^\theta + y_i^\theta) \\
&\quad -(2+\frac{\alpha}{\theta})\sum_{i=1}^{n^{\|}}\frac{\partial}{\partial\theta}\log(x_i^\theta + y_i^\theta) + \sum_{i=1}^{n_1^{\perp}}\frac{\partial}{\partial\theta}\log\left[1-\left(1+(\widetilde{x}_i/\varepsilon)^{-\theta}\right)^{-\alpha/\theta-1}\right] \\
&\quad +\sum_{i=1}^{n_2^{\perp}}\frac{\partial}{\partial\theta}\log\left[1-\left(1+(\widetilde{y}_i/\varepsilon)^{-\theta}\right)^{-\alpha/\theta-1}\right].
\end{aligned}
$$

The three parameters are obtained by numerical optimization.

It is possible to prove joint asymptotic normality of $(\log c, \alpha, \theta)$ similar to our calculations in Esmaeili and Klüppelberg [18] and in Section 4.4 of the present paper. However, for the observation scheme of the present paper this is even more complicated than in [18]. We refrain from this tedious analytic exercise and, instead, present the results of a simulation study in the next section, where we compare all three methods presented.

4.6 Comparison of the estimation procedures

In this section we compare the quality of the MLEs $\widehat{\eta} = (\log\widehat{c}, \widehat{\alpha}, \widehat{\theta})$ of the full model of Section 2.6 with the estimates $\widetilde{\eta} = (\log\widetilde{c}, \widetilde{\alpha}, \widetilde{\theta})$ obtained by the two-step method in Section 2.5. Moreover, we also include in our comparison those estimates obtained from bivariate jumps larger than ε only as derived in Th. 4.6 of [18]. Since this last method means to base the statistical analysis on less data, we expect that this method is less efficient than the MLE based on all available data. More precisely, for the first two parameters $\log c$ and α, the rate has simply changed from $\sqrt{c2^{-\alpha/\theta}\varepsilon^{-\alpha}t}$ to $\sqrt{c2\varepsilon^{-\alpha}t}$.

The simulation study

We simulate sample paths of the bivariate α-stable Clayton subordinator with equal marginals and parameters given by $c = 1$ ($\log c = 0$), $\alpha = 1/2$ and $\delta = 2$ ($\theta = 1$). We generate sample paths of this process over a time span $[0, t]$, where we choose $t = 1$ for simplicity. Recall from our observation scheme introduced in Section 2.2 that we observe all jumps larger than ε either in one component or in both. Obviously, we cannot simulate a trajectory of a stable process, since we are restricted to the simulation of a finite number of jumps. For simulation purposes we choose a threshold ξ (which should be much smaller than ε) and simulate jumps larger than ξ in one component, and arbitrary in the second component. To this end we invoke Algorithm 6.15 in Cont and Tankov [13].

The simulation of a bivariate α-stable Clayton subordinator is explained in detail in Example 6.18 of [13]. The algorithm starts by fixing a number τ determined by the required precision. This number coincides with the jump intensity $\lambda_1^{(\xi)}$, which fixes the average number of terms in the approximating CPP. More details can be found in [18].

For the estimation we first consider $\varepsilon = 0.001$, i.e. a relatively large truncation point. Not surprisingly, the MLEs based on the full model discussed in Section 2.6 are definitely better than the other estimates in Table 4.1. We find it, however, remarkable that the two-step method outperforms the MLE based on joint jumps only. The reason for this is presumably that the MLE's based only on joint jumps use only such data with Lévy measure on $[\varepsilon, \infty)^2$. The two-step method, however, uses also data, which are only in one component larger than ε in its first step. The marginal parameters are based on substantially more data.

When we consider also smaller jumps; i.e., if we choose $\varepsilon = 10^{-5}$, the estimates will be more precise with less variation and smaller bias. In Table 4.1, the results in the lower part of each estimation method show this fact. It can also be seen from this table that the MLEs from a full model have the least mean relative bias (MRB) and mean square errors (MSE) as expected. In Figure 4.2 we visualize the situation for the this jump truncation point of $\varepsilon = 10^{-5}$ based on 1000 simulated sample paths. Again all three estimation methods are performed for each sample path.

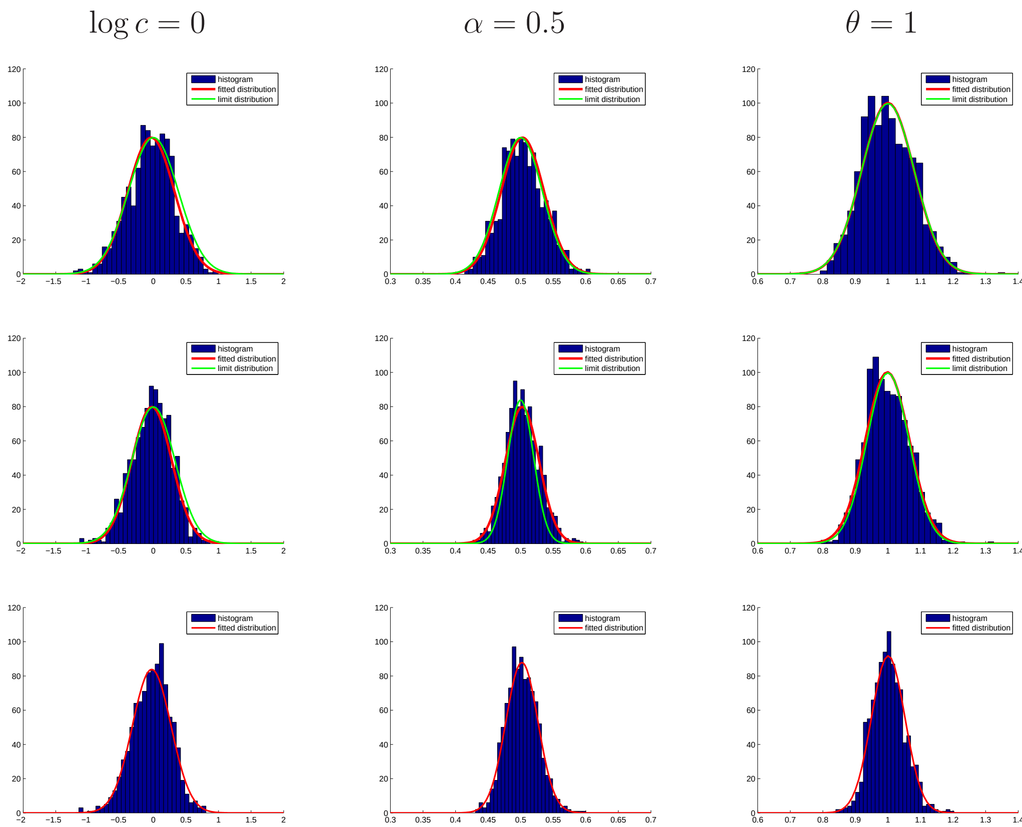

Figure 4.2: Histogram with statistically fitted normal density (red) and theoretical limit distribution (green) for 1000 parameter estimates of a bivariate Clayton stable Lévy process. The parameter values are $c = 1$, $\alpha = 0.5$ and $\delta = 2$ and the jump-truncated point is $\varepsilon = 0.00001$. The estimation procedures are MLEs based on joint jumps only (first row, limit distribution derived in Theorem 4.6 of [18]), the two-step method (second row, limit distribution derived in Theorem 4.4.7 above) and MLEs based on all jumps (third row, without theoretical limit law).

Method of estimation	Truncation point		$c = 1$	$\alpha = 0.5$	$\delta = 2$
MLE (only bivariate jumps) as in Section 3.4	$\varepsilon = 0.001$	Mean	1.0678	0.5289	2.1489
		$\sqrt{MSE}$	0.6344	0.1206	0.9511
		MRB	0.0517	0.0525	0.0842
	$\varepsilon = 0.00001$	Mean	1.0460	0.5020	2.0301
		$\sqrt{MSE}$	0.3677	0.0349	0.2488
		MRB	0.0413	0.0044	0.0144
MLE (full model) as in Section 4.5	$\varepsilon = 0.001$	Mean	1.0177	0.5216	2.0129
		$\sqrt{MSE}$	0.5248	0.0777	0.4337
		MRB	0.0072	0.0423	0.0119
	$\varepsilon = 0.00001$	Mean	1.0175	0.5021	2.0091
		$\sqrt{MSE}$	0.2808	0.0239	0.1253
		MRB	0.0142	0.0045	0.0042
Two-step method as in Section 4.3	$\varepsilon = 0.001$	Mean	1.0453	0.5231	2.0762
		$\sqrt{MSE}$	0.5535	0.0859	0.6764
		MRB	0.0264	0.0471	0.0379
	$\varepsilon = 0.00001$	Mean	1.0301	0.5021	2.0149
		$\sqrt{MSE}$	0.3003	0.0257	0.1696
		MRB	0.0249	0.0048	0.0065

Table 4.1: Comparison of estimates for a bivariate $\frac{1}{2}$-stable Clayton process with common marginal parameters. We simulated 100 sample paths and estimated all parameters 100 times. Each of the 100 estimates was based on one sample path, on which all three methods were performed. From each sample path we truncated the small jumps based on the two truncation points ($\varepsilon = 0.001$ and $\varepsilon = 0.00001$), respectively. Each sample path of the process was simulated as a continuous time realization of a CPP in one unit of time, $0 \le t < 1$, for $\tau = 1000$, equivalent to truncation of the small jumps at the cut-off point $\xi = \overline{\Pi}^{\leftarrow}(\tau) = 10^{-6}$.

Bibliography

[1] Applebaum, D. (2003) *Lévy Processes and Stochastic Calculus.* Cambridge University Press, Cambridge.

[2] Arnold, B.C., Castillo, E. and Sarabia, J.M. (1999) *Conditional Specification of Statistical Models.* Springer, New York.

[3] Barndorff-Nielsen, O. and Lindner, A. (2007) Lévy copulas: dynamics and transforms of Upsilon type. *Scand. J. Stat.* **34**, No. 2, 298-316(19).

[4] Basawa, I.V. and Brockwell, P.J. (1978) Inference for gamma and stable processes. *Biometrika* **65**, No. 1, 129-133.

[5] Basawa, I.V. and Brockwell, P.J. (1980) A note on estimation for gamma and stable processes. *Biometrika* **67**, No. 1, 234-236.

[6] Basawa, I.V. and Prakasa Rao, B.L.S. (1980) *Statistical Inference for Stochastic Processes.* Academic Press, New York.

[7] Bertoin, J. (1996) *Lévy processes.* Cambridge University Press, Cambridge.

[8] Bregman, Y. and Klüppelberg, C. (2005) Ruin estimation in multivariate models with Clayton dependence structure. *Scand. Act. J.* 2005 (6), 462-480.

[9] Böcker, K. and Klüppelberg, C. (2010) Multivariate models for operational risk. *Quantitative Finance* **10**(8), 855-869.

[10] Böcker, K. and Klüppelberg, C. (2008) Modelling and measuring multivariate Operational Risk with Lévy copulas. *J. Operational Risk* **3**(2), 3-27.

[11] Böcker, K. and Klüppelberg, C. (2009) First order approximations to Operational Risk - dependence and consequences. To appear in: G.N. Gregoriou (ed.), *Operational Risk Toward Basel III, Best Practices and Issues in Modeling, Management and Regulation.* Wiley, New York.

[12] Cherubini, U., Luciano, E. and Vecchiato, W. (2004) *Copula Methods in Finance.* Wiley, London.

[13] Cont, R. and Tankov, P. (2004) *Financial Modelling with Jump Processes.* Chapman & Hall/CRC, Boca Raton.

[14] Eder, I. and Klüppelberg, C. (2009) The first passage event for sums of dependent Lévy processes with applications to insurance risk. *Ann. Appl. Probab.* **19**(6), 2047-2079.

[15] Eder, I. and Klüppelberg, C. (2009)Pareto Lévy measures and multivariate regular variation. Submited for publication.

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

[17] Esmaeili, H. and Klüppelberg, C. (2010) Parameter estimation of a bivariate compound Poisson process. *Insurance: Mathematics and Economics* **47**, 224-233.

[18] Esmaeili, H. and Klüppelberg, K. (2011) Parametric estimation of a bivariate stable Lévy process. *J. Mult. Anal.* **102**, 918-930.

[19] Esmaeili, H. and Klüppelberg, K. (2010) Two-step estimation of a multivariate Lévy process. Submitted for publication.

[20] Godambe, V.P. (1991) *Estimating Functions.* Oxfprd University Press, Oxford.

[21] Höpfner, R. (1999) On statistical models for d-dimensional stable processes, and some generalisations. *Scand. J. Stat.* **26**, 611-620.

[22] Höpfner, R. and Jacod, J. (1994) Some remarks on the joint estimation of the index and the scale parameter for stable processes. In: Mandl, P. and Huskova, M. (Eds.) *Asymptotic Statistics. Proceedings of the Fifth Prague Symposium 1993*, pp. 273-284. Physica Verlag, Heidelberg.

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

[24] Joe, H. and Xu, J.J. (1996) The estimation method of inference functions for margins of multivariate models. Technical Report No. 166. Department of Statistics, University of British Columbia. Available at {*http://www.stat.ubc.ca/~harry/ifm.pdf*}

[25] Kallsen, J. and Tankov, P. (2006) Characterization of dependence of multidimentional Lévy processes using Lévy copulas. *J. Mult. Anal.* **97**, 1551-1572.

[26] Küchler, U. and Sørensen, M. (1997) *Exponential Families of Stochastic Processes.* Springer, New York.

[27] Nelsen, R. B. (1997) *An Introduction to Copulas.* Springer, New York.

[28] Nolan, J.P., Panorska, A. and McCulloch, J.H. (2001) Estimation of stable spectral measures. *Mathematical and Computer Modelling* **34**, 1113-1122.

[29] Prakasa Rao, B.L.S (1987) *Asymptotic Theory of Statistical Inference.* Wiley, New York.

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

[31] Samorodnitsky, G. and Taqque, M. S. (1994) *Stable non-Gaussian random processes, stochastic models with infinite variance.* Chapman & Hall, New York.

[32] Sato, K.-I. (1999) *Lévy Processes and Infinitely Divisible Distributions.* Cambridge University Press, Cambridge.

[33] Sklar, A. (1959) Functions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8**, 229-231.

[34] Tankov, P. (2003) Dependence structure of spectrally positive multidimensional Lévy processes. Available at {*http://www.camp.polytechnique.fr/*}

[35] Ueltzhöfer, F. and Klüppelberg, C. (2010) An oracle inequality for penalised projection estimation of Lévy densities from high-frequency observations. submitted for publication.

[36] Xu, J.J. (1996) Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response data. *Ph.d. thesis, University of British Columbia, Department of Statistics.*

Index

List of Abbreviations and Symbols

Abbreviation or Symbol	**Explanation**
a.s.	almost surely
$\mathcal{B}(\cdot)$	Borel σ-algebra
$\lvert\cdot\rvert, \#$	cardinality of a set
C	distributional copula
$\overline{C}$	survival copula
$\mathfrak{C}$	Lévy copula
CPP	compound Poisson process
$\mathbb{E}[\cdot]$	expectation operator
ε	cut-off point
$F, \overline{F}$	distribution and survival function of jumps
(γ, A, Π)	characteristic triplet
$\mathcal{I}(x)$	$(-\infty, x)$ for $x < 0$ and $[x, -\infty)$ for $x \geq 0$
i.i.d.	independent and identically distributed
1_B	indicator function of set B
$(\cdot\,,\cdot)$	inner product
λ	intensity of jumps for a CPP
$\lambda^{\perp}, \lambda^{\parallel}$	intensity of single and joint jumps
l.h.s.	left hand side
n	number of jumps for a CPP
Π, Π_k	multi-/one-dimensional Lévy measure
$\Pi^{(\varepsilon)}$	Lévy measure of a jump-truncated process

$\overline{\Pi}$	tail integral
$\overline{\Pi}^{(\varepsilon)}$	tail integral of a jump-truncated process
$\mathbb{R}$	real line
$\mathbb{R}_+^d$	$[\mathbf{0}, \infty) \setminus \{\mathbf{0}\}$
$\overline{\mathbb{R}}$	$[-\infty, \infty]$
Ran	range
r.h.s.	right hand side
r.v.	random variable
$\mathbf{S}, S$	multi-/one-dimensional Lévy process
$\mathbf{S}^{(\varepsilon)}, S^{(\varepsilon)}$	jump-truncated multi-/one-dimensional Lévy process
$S_k^{\perp}$	process constructed by means of single jumps
$S^{\parallel}$	process constructed by means of joint jumps
$\mathbf{Z} = (X, Y)$	jump size of a bivariate Lévy process
$\mathbf{0}$	$(0, \ldots, 0)$