PhD Dissertation

Local Behavior and Graphical Models for Stochastic Processes

Jakob Dalsgaard Thøstesen

Department of Mathematics
Aarhus University
July 2023

Local Behavior and Graphical Models for Stochastic Processes

PhD dissertation by
Jakob Dalsgaard Thøstesen

Department of Mathematics, Aarhus University
Ny Munkegade 118, 8000 Aarhus C, Denmark

Supervised by
Associate Professor Jevgenijs Ivanovs

Submitted to Graduate School of Natural Sciences, Aarhus, July 2023

AARHUS
UNIVERSITY

Preface

This dissertation is the result of my PhD studies from September 2019 to July 2023 under the supervision of Associate Professor Jevgenijs Ivanovs at the Department of Mathematics, Aarhus University.

The dissertation consists of an introduction followed by four papers. The content of the papers is identical to that of the published or submitted versions but there will be minor differences such as typographical corrections and small revisions to some figures.

Paper I Discretization of the Lamperti representation of a positive self-similar Markov process.
Stochastic Processes and their Applications 137, 200-221.

Paper II Lévy processes conditioned to stay in a half-space with applications to directional extremes.
Modern Stochastics: Theory and Applications 10(1), 59-75.

Paper III Local behavior of diffusions at the supremum.
Submitted to Stochastic Analysis and Applications, January 2022.

Paper IV Graphical models for Lévy processes.
Not yet submitted.

Papers I and II were written jointly with Jevgenijs Ivanovs, with the research and writing being divided more or less equally between us. Large parts of these two papers were also included in the progress report for my qualifying exam. Paper III is the result of a solo project in the beginning of part B of my studies. Lastly, Paper IV has been written jointly with Sebastian Engelke and Jevgenijs Ivanovs. Sebastian Engelke is an associate professor at the University of Geneva in Switzerland and I visited him for nearly four months during the spring of 2022. For Paper IV, I contributed significantly to both the research and writing.

The introductory chapter is mostly there to set the scene and motivate the four projects. Each paper has its own section which discusses the general idea, main results and methodology. It also contains a brief conclusion on the project, including references to relevant work and some ideas for further research.

Here, at the end of my studies, I would like to thank my supervisor Jevgenijs Ivanovs for guiding me during the past four years. I have enjoyed not only our scientific collaboration but also the occasional hike, swim or bouldering session. Our trip through the snow to Les Rochers-de-Naye was particularly memorable.

Being a PhD student at the Department of Mathematics at Aarhus University has been great and my colleagues have played a big part in this. In particular, I would like to thank Jan Pedersen for our many discussions. The same applies to my office mates Helene Hauschultz, Ragnhild Laursen, Lota Copic and Kenneth Borup although our discussions have been less mathematical and more about things like politics, bouldering and Formula 1.

During my PhD I was fortunate to spend a few months in Geneva. Here I would like to thank Sebastian Engelke for letting me visit and the rest of the group for welcoming me. I had an excellent time with wine tasting, climbing, swimming in the lake and much more.

I thank also Lars Madsen for assisting me with LaTeX in the preparation of this dissertation. I only wish that I had asked for his input earlier as it has been very helpful.

Finally, my family and friends deserve my biggest thanks for always being interested in my work even though it may have been difficult to completely understand. In particular, I want to thank my girlfriend Malene. Your support means a lot and I am especially thankful that you took the time to come and visit me in Switzerland on several occasions.

Contents

Abstract

Stochastic processes are used to model quantities that exhibit random fluctuations over time, such as stock prices, temperatures, wind speeds, etc. Lévy processes make up a popular class of models due to their theoretical properties and applications in e.g. finance and physics.

From a theoretical point of view these processes evolve in continuous time but in practice only a finite number of observations are available. Understanding the implications of this discretization is essential and in many cases it requires knowledge of the local properties of the process. The first paper in this dissertation is concerned with a specific discretization scheme for positive self-similar Markov processes. To describe this we rely on knowledge about small-time fluctuations of Lévy processes. In another paper we examine the local properties of diffusions. At a fixed time point such a process behaves locally as a scaled Brownian motion and we prove a similar result for the small-time fluctuations at the supremum. The theory which describes a univariate Lévy process before and after its supremum is well-known and relies on the notion of a Lévy process conditioned to stay positive or negative. In a third paper we extend this to the multivariate setting, constructing the law of a Lévy process conditioned to stay in a half-space. This is related to splitting the process at its directional supremum and we further conjecture how it can be used to describe the local behavior of the process when it is farthest from the origin.

One of the big challenges in modern statistics is dealing with high-dimensional data. Classical models are faced with an increased risk of overfitting, large computational cost and low interpretability. The concept of sparsity addresses these issues by taking advantage of lower-dimensional structures in the data. The use of graphical models is one way of promoting sparsity and has recently been introduced in multivariate extreme value theory. The final paper in this dissertation introduces graphical models in the context of Lévy processes. To do this we exploit a subtle connection to extremes.

Resumé

Stokastiske processer anvendes til at modellere værdier, der udviser tilfældige udsving over tid, såsom aktiekurser, temperaturer, vindhastigheder osv. Lévy-processer udgør en populær klasse af modeller på grund af deres teoretiske egenskaber samt anvendelser inden for f.eks. finans og fysik.

Ud fra et teoretisk synspunkt udvikler disse processer sig i kontinuert tid, men i praksis er kun et endeligt antal observationer tilgængelige. Det er vigtigt at forstå konsekvenserne af denne diskretisering, og ofte kræver dette kendskab til processens lokale egenskaber. Den første artikel i denne afhandling drejer sig om et konkret diskretiseringsproblem for positive selvsimilære Markov-processer. Til at behandle dette benytter vi viden om Lévy-processers opførsel på kort tidsskala. I en anden artikel undersøger vi de lokale egenskaber for diffusionsprocesser. På et fast tidspunkt opfører en sådan proces sig lokalt som en skaleret Brownian motion, og vi viser et lignende resultat for fluktuationerne omkring supremum. Teorien, der beskriver en univariat Lévy-proces før og efter dens supremum, er velkendt og bygger på begrebet om en Lévy-proces betinget til at forblive positiv eller negativ. I en tredje artikel udvider vi dette til højere dimensioner, hvor vi konstruerer fordelingen af en Lévy-proces betinget til at forblive i et halvrum. Dette er relateret til at dele processen ved dens retningsbestemte supremum, og vi præsenterer en formodning om, hvordan dette kan bruges til at beskrive processens lokale opførsel, når den befinder sig længst væk fra origo.

En af de store udfordringer inden for moderne statistik er håndteringen af højdimensionelt data. Klassiske modeller står over for en øget risiko for overfitting, høje beregningsomkostninger samt lav fortolkningsevne. Konceptet om sparsitet adresserer disse problemer ved at udnytte lavdimensionelle strukturer i data. Brugen af grafiske modeller er en måde at fremme sparsitet, og disse er for nyligt blevet introduceret i multivariat ekstremværditeori. Den sidste artikel i denne afhandling introducerer grafiske modeller for Lévy-processer. Til at gøre dette udnytter vi en subtil forbindelse til ekstremværditeori.

Introduction

This chapter has two purposes, the first of which is to introduce the reader to certain areas which are essential for the papers in this dissertation. These are both relatively standard topics such as Lévy processes and extreme value theory but also more specialized concepts such as Lévy processes conditioned to stay positive and conditional independence for multivariate Pareto distributions. Some ideas and definitions may be reiterated in the introductory sections of each of the papers. The other purpose of this chapter is to provide an introduction to each of the four papers, including some motivation and connection to the literature.

Notation

The notation will not be entirely consistent across this introductory chapter and the papers. One thing which will differ is the way we write the value of a stochastic process at some time point. Depending on the context we will write either X_t or $X(t)$ for the value of a process X at time t. In connection with Paper IV we will also use bold letters to emphasize that something is a stochastic process and not just a random vector. That is, we will write $\mathbf{X}$ when we are talking about a process and $X(t)$ when we are talking about its value at time t.

1.1 Lévy processes

Lévy processes should be seen as the result of wanting to extend random walks to continuous time. The first steps in this direction were taken nearly 100 years ago and today there is a large number of books and research papers dealing with Lévy processes and related topics. Standard references include Bertoin (1996), Sato (1999) and Applebaum (2009). Over time other research areas have put the theory of Lévy processes to use. For example, Lévy driven models have become popular in financial disciplines such as option pricing and risk management, see e.g. Cont and Tankov (2004). The role of Lévy processes in other fields such as quantum mechanics is described by Barndorff-Nielsen et al. (2001).

A Lévy process is an $\mathbb{R}^d$-valued stochastic process $X = (X_t)_{t\geq 0}$, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which satisfies the following properties.

1. $X_0 = 0$ $\mathbb{P}$-a.s.

2. $X_t - X_s \overset{d}{=} X_{t-s}$ for any $0 \leq s \leq t$.

3. $X_t - X_s$ is independent of $\sigma(X_u \mid 0 \leq u \leq s)$ for any $0 \leq s \leq t$.

4. The path $t \mapsto X_t$ is càdlàg (right continuous with left limits) $\mathbb{P}$-a.s.

Another common definition replaces 4 by *continuity in probability*. That is, for any $\epsilon > 0$ and $t \geq 0$

$$\mathbb{P}(\|X_t - X_s\| > \epsilon) \to 0 \quad \text{as } t \to s,$$

where $\|\cdot\|$ is a norm on $\mathbb{R}^d$. In this case one may construct a version with almust surely càdlàg paths, see e.g. Kallenberg (2021, Thm. 16.2).

1.1.1 Connection to infinitely divisible distributions

An important consequence of points 1–3 above is that the distribution of X_t is infinitely divisible for any $t \geq 0$. That is, for any $n \geq 1$ there exists i.i.d. random variables $Y_1, \ldots, Y_n$ such that $X_t \overset{d}{=} Y_1 + \cdots + Y_n$. Indeed, we simply take $Y_i = X_{i/n \cdot t} - X_{(i-1)/n \cdot t}$. While this shows that every Lévy process induce many infinitely divisible distributions, the connection actually is much stronger. In fact, for every infinitely divisible distribution μ on $\mathbb{R}^d$ there exists a unique (in law) Lévy process X such that $X_1 \sim \mu$, see Sato (1999, Cor. 11.6).

It is well-known that the distribution of a random vector is characterized by its characteristic function. A consequence of the infinite divisibility is that the characteristic function of X_t can be written as

$$\mathbb{E}[e^{i\langle u, X_t\rangle}] = e^{t\psi(u)}, \qquad u \in \mathbb{R}^d,$$

where $\psi(u)$ is given by

$$\psi(u) = i\langle u, \gamma\rangle - \frac{1}{2}\langle u, \Sigma u\rangle + \int_{\mathbb{R}^d} e^{i\langle u, x\rangle} - 1 - i\langle u, x\rangle\, \mathbf{1}_{\{\|x\| \leq 1\}}\, \Lambda(\mathrm{d}x),$$

see Sato (1999, Thm. 8.1). Here $\gamma \in \mathbb{R}^d$, Σ is a positive semidefinite $d \times d$ matrix, and Λ is a measure on $\mathbb{R}^d$. The latter is typically referred to as the Lévy measure and it satisfies

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

This identity for the characteristic function is known as the Lévy–Khintchine formula and is one of the most fundamental results about infinitely divisible distributions. Typically, the process X is said to have has characteristic triplet $(\gamma, \Sigma, \Lambda)$ and it is unique when ψ is written like above. However, it is worth noting that the indicator $\mathbf{1}_{\{\|x\| \leq 1\}}$ may be replaced by any $c(x)$, where $c(x) = 1 + \mathrm{o}(\|x\|)$ as $\|x\| \to 0$ and $c(x) = \mathrm{O}(1/\|x\|)$ as $\|x\| \to \infty$, see Sato (1999, Rem. 8.4). This does not affect Σ and Λ but the vector γ must be modified accordingly.

1.1.2 The path behavior of a Lévy process

To work with Lévy processes it is important to understand how they behave. Point 4 in the definition does provide some immediate restrictions on what a sample path might look like. Since the sample paths are almost surely càdlàg it makes sense (except on a $\mathbb{P}$-null set) to consider the jump $\Delta X_t = X_t - X_{t-}$ at any time point $t > 0$. One can easily show that for any $\epsilon > 0$ and any bounded time interval $[T_1, T_2]$ there can be only finitely many time points $t \in [T_1, T_2]$ with $\|\Delta X_t\| \geq \epsilon$. This implies, in particular, that X has at most countably many jumps. However, some Lévy processes will actually have infinitely many infinitesimally small jumps during any time interval of positive length.

It is immediately clear from points 2 and 3 in the definition that a Lévy process is also a Markov process. Specifically, for any deterministic time $T \geq 0$ we see that the *restarted* process $(X_{t+T} - X_T)_{t \geq 0}$ is independent of $\sigma(X_u \mid 0 \leq u \leq T)$ and it has the same law as X.

A standard argument, relying on $t \mapsto X_t$ being almost surely right-continuous, extends the Markov property to also hold when T is a (finite) stopping time wrt. the filtration generated by X. This is commonly known as the *strong Markov property*.

Obvious examples of Lévy processes are a compound Poisson process, a Brownian motion, and a linear drift. If these are independent then their sum is also a Lévy process. This shows that the path of a Lévy process may contain both jumps, Brownian fluctuations and a drift. The Lévy–Itô decomposition states that for any Lévy process X there exists a Brownian motion B (possibly with a drift), a compound Poisson process C and a square integrable martingale M such that B, C and M are independent and X_t can be written as

$$X_t = B_t + C_t + M_t, \qquad t \geq 0,$$

see Applebaum (2009, Thm. 2.4.16). The process B is often referred to as the *Brownian part* of X while the sum $C + M$ is referred to as the *jump part*. Indeed, B is continuous so all jumps of X are contained in $C + M$.

The covariance matrix of B is the matrix Σ from the characteristic triplet of X. The processes C and M are not unique but one option is to let C contain all large jumps and M contain all small jumps. For example, C and M might be constructed to contain jumps with norm respectively larger and strictly smaller than 1. In this case C will have rate $\lambda = \Lambda(\{x \mid \|x\| \geq 1\})$ and jump distribution $\lambda^{-1}\Lambda(\,\cdot\, \cap \{x \mid \|x\| \geq 1\})$, while M is a Lévy process with no Brownian part and Lévy measure $\Lambda(\,\cdot\, \cap \{x \mid \|x\| < 1\})$.

1.2 Lévy processes conditioned to stay positive

A key concept in this dissertation is that of a Lévy process conditioned to stay positive. For many Lévy processes the probability of being positive at all times $t > 0$ is zero. Conditioning to stay positive should therefore not be understood in the usual sense. This section contains a brief overview of the properties which are central to this dissertation. The topic has been treated in much greater detail in numerous papers and books. Some of the typical references are Bertoin (1993), Chaumont (1996) and Chaumont and Doney (2005).

Throughout the section we let $\mathcal{D}$ denote the space of càdlàg functions $\omega\colon [0,\infty) \to \mathbb{R} \cup \{\dagger\}$, where $\dagger$ is an absorbing cemetery state. This is equipped with the Skorokhod topology and $\mathcal{F}$ is the resulting Borel σ-field. We further denote the coordinate process by $X = (X_t)$, i.e. $X_t(\omega) = \omega(t)$, and its natural filtration by $(\mathcal{F}_t)$. The *lifetime* of X is a random variable given by $\zeta = \inf\{t \geq 0 \mid X_t = \dagger\}$.

We let $\mathbb{P}$ be a measure on $(\mathcal{D}, \mathcal{F})$ such that X is a Lévy process (starting at 0) under $\mathbb{P}$. For any $x \in \mathbb{R}$ we use $\mathbb{P}_x$ to denote the law of $X + x$ under $\mathbb{P}$. As usual we will assume w.l.o.g. that for these and any other laws on $(\mathcal{D}, \mathcal{F})$ the filtration $(\mathcal{F}_t)$ is complete.

We make the typical assumption that X is not a compound Poisson process under $\mathbb{P}$. To deal with such processes one is usually able to use the tools and ideas from discrete time.

1.2.1 Bertoin's construction

There are different ways of constructing a Lévy process to stay positive. In this subsection we look at the one by Bertoin (1993). Another approach is given by Tanaka (1989). This was originally meant for discrete time, i.e. random walks, but it was extended to continuous time in Doney (2005).

Since the Lévy process X is also a semimartingale we may consider the semimartingale local time L of X at 0 which is given by the Meyer–Tanaka formula,

$$X_t^+ - X_0^+ = \int_0^t \mathbf{1}_{\{X_{s-}>0\}} \, \mathrm{d}X_s + \sum_{0<s\le t} \left(\mathbf{1}_{\{X_{s-}\le 0\}} \, X_s^+ + \mathbf{1}_{\{X_{s-}>0\}} \, X_s^- \right) + \tfrac{1}{2} L_t$$

for $t \ge 0$. Now, let Y^- and Y^+ be the processes given by

$$\begin{aligned} Y_t^- &= X_t - \sum_{0<s\le t} \left(\mathbf{1}_{\{X_s\le 0\}} \, X_{s-}^+ + \mathbf{1}_{\{X_s>0\}} \, X_{s-}^- \right) - \tfrac{1}{2} L_t, \\ Y_t^+ &= X_t + \sum_{0<s\le t} \left(\mathbf{1}_{\{X_s\le 0\}} \, X_{s-}^+ + \mathbf{1}_{\{X_s>0\}} \, X_{s-}^- \right) + \tfrac{1}{2} L_t. \end{aligned} \tag{1.1}$$

We further define two continuous and non-decreasing processes A^- and A^+ which track the time spent in $(-\infty, 0]$ and $(0, \infty)$ respectively. These are given by

$$A_t^- = \int_0^t \mathbf{1}_{\{X_s\le 0\}} \, \mathrm{d}s \quad \text{and} \quad A_t^+ = \int_0^t \mathbf{1}_{\{X_s>0\}} \, \mathrm{d}s$$

for $t \ge 0$. The right-continuous inverses of $A^\pm$ are denoted by $\alpha_t^\pm = \inf\{s \ge 0 \mid A_s^\pm > t\}$. It is now possible to define two processes $X^\downarrow$ and $X^\uparrow$ by

$$X_t^\downarrow = Y_{\alpha_t^-}^- \quad \text{and} \quad X_t^\uparrow = Y_{\alpha_t^+}^+. \tag{1.2}$$

If we ignore the local time in the definition of Y^- we see that $X^\downarrow$ is the juxtaposition of the excursions of X in $(-\infty, 0]$. That is, $X^\downarrow$ is obtained by 'gluing' together the parts of the path of X in the non-positive numbers including the jumps where $X_{s-} > 0$ and $X_s \le 0$. Similarly, $X^\uparrow$ comes from gluing together the excursions in $(0, \infty)$ (if we ignore the local time contribution).

The laws of $X^\downarrow$ and $X^\uparrow$ under $\mathbb{P}$ are denoted by $\mathbb{P}^\downarrow$ and $\mathbb{P}^\uparrow$ and we shall refer to these as the laws of X *conditioned to stay non-positive* and *conditioned to stay positive* respectively. The convergence in (1.7) below provides some justification of this terminology.

To understand why the processes $X^\downarrow$ and $X^\uparrow$ are interesting we will look at the infimum of X. Let $\underline{X}$ denote the running infimum of X. That is,

$$\underline{X}_t = \begin{cases} \inf_{0\le s\le t} X_s & \text{for } t < \zeta, \\ I & \text{for } t \ge \zeta, \end{cases}$$

where $I = \underline{X}_{\zeta-}$. The time of the ultimate infimum is denoted by $\tau = \sup\{0 \le t < \zeta \mid X_t \wedge X_{t-} = I\}$. We further define the pre-infimum and post-infimum processes by

$$\underleftarrow{X}_t = \begin{cases} X_{(\tau-t)-} - I & \text{for } t < \tau, \\ \dagger & \text{for } t \ge \tau, \end{cases} \quad \text{and} \quad \underrightarrow{X}_t = \begin{cases} X_{\tau+t} - I & \text{for } t < \zeta - \tau, \\ \dagger & \text{for } t \ge \zeta - \tau. \end{cases}$$

Note that the pre-infimum process is reversed in time such that it 'looks back' from the infimum.

For any fixed $T > 0$ we denote by $\mathbb{P}^T$ the law of X killed at time T under $\mathbb{P}$. Then

$$(-\underleftarrow{X}, \underrightarrow{X}) \overset{d}{=} (X^\downarrow, X^\uparrow) \quad \text{under } \mathbb{P}^T, \tag{1.3}$$

see Bertoin (1993, Thm. 3.1). This fact can be seen as the primary motivation behind studying the laws $\mathbb{P}^\downarrow$ and $\mathbb{P}^\uparrow$. We will see an application of this in §1.3 below.

1.2.2 Properties

The processes $X^\downarrow$ and $X^\uparrow$ defined in (1.2) are Markov processes under $\mathbb{P}$, see Bertoin (1993, Thm. 3.4). In the following we will focus on $X^\uparrow$ and simply remark that $X^\downarrow = -(-X)^\uparrow$ a.s. In order to write down the semigroup of $X^\uparrow$ we first recall that the *reflected process* $\underline{X} - X$ is a Markov process, see e.g. Bertoin (1996, Prop. VI.1). Let $\underline{L}$ denote a Markov process local time at 0 for $\underline{X} - X$ and define

$$h(x) = \mathbb{E}\Big[\int_{[0,\infty)} \mathbf{1}_{\{\underline{X}_t \geq -x\}} \,\mathrm{d}\underline{L}_t\Big], \qquad x > 0. \tag{1.4}$$

Recall that the local time $\underline{L}$ is only unique up to multiplication by a positive constant and that scaling of the local time will result in the same scaling of h. However, in the following it should be clear that this scaling is irrelevant as we always consider ratios such as $h(y)/h(x)$.

Generally, we have the inequality

$$\mathbb{E}_x[h(X_t)\,\mathbf{1}_{\{\underline{X}_t > 0\}}] \leq h(x), \tag{1.5}$$

and when X does not drift to $-\infty$ it is an equality, see Chaumont and Doney (2005, Lem. 1). We define

$$p_t^\uparrow(x, \mathrm{d}y) = \frac{h(y)}{h(x)} \mathbb{P}_x(X_t \in \mathrm{d}y, \underline{X}_t > 0), \qquad x, y, t > 0. \tag{1.6}$$

The collection $(p_t^\uparrow)$ is the semigroup of $X^\uparrow$ in $(0, \infty)$ under $\mathbb{P}$. We denote the law of the Markov process starting at $x > 0$ and having semigroup $(p_t^\uparrow)$ by $\mathbb{P}_x^\uparrow$.

Note that $p_t^\uparrow(x, \cdot\,)$ need not be a probability measure since the inequality in (1.5) may be strict. In that case X has finite lifetime under $\mathbb{P}_x^\uparrow$. In the literature $\mathbb{P}^\uparrow$ is typically only referred to as the law of X conditioned to stay positive when (1.5) is an equality for all $t > 0$. In this case the construction of $p_t^\uparrow$ in (1.6) can be viewed as a so-called Doob h-transform. The latter is a general technique which allows one to condition a Markov process to stay in a certain set even if that event has a probability of zero. Other examples include conditioning the Lévy process to hit an interval continuously, see Döring and Weissmann (2020), and conditioning it to avoid an interval, see Döring et al. (2019).

While the terminology 'conditioned to stay positive' can be explained via the Doob h-transform, there is a much more intuitive way to justify this. If $x > 0$ and e_q is an independent exponential variable with rate q then

$$\lim_{q \to 0} \mathbb{P}_x(\Lambda, T < e_q \mid \underline{X}_{e_q} > 0) = \mathbb{P}_x^\uparrow(\Lambda, T < \zeta) \tag{1.7}$$

for any $(\mathcal{F}_t)$-stopping time T and $\Lambda \in \mathcal{F}_T$, see Chaumont and Doney (2005, Prop. 1). In other words, to obtain the law $\mathbb{P}_x^\uparrow$ we take X under $\mathbb{P}_x$, condition it to stay positive (at least) until e_q, and finally let $\mathbb{E}[e_q] = 1/q \to \infty$.

1.2.3 The case of self-similarity

Recall that X is called strictly α-stable under $\mathbb{P}$ if there exists a constant $\alpha \in (0, 2]$ such that $(X_{ct}) \overset{d}{=} c^{1/\alpha} X$ for all $c > 0$ under $\mathbb{P}$. This is the same as saying that X is $1/\alpha$-self-similar. If $\alpha = 2$ then X is a Brownian motion and if $\alpha \in (0, 2)$ then X has no Brownian component. When X is strictly α-stable the function h from (1.4) takes the simple form $h(x) = x^{\alpha\rho}$ where $\rho = \mathbb{P}(X_1 < 0)$, see Caballero and Chaumont (2006, §3.2).

Assume that X is strictly α-stable and that neither X nor $-X$ is a subordinator. Then X is oscillating, see e.g. Kyprianou and Pardo (2022, §3.4), meaning that $\limsup_{t\to\infty} X_t = \infty$

a.s. and $\liminf_{t\to\infty} X_t = -\infty$ a.s. Since X does not drift to $-\infty$ we have that $(p_t^\uparrow)$ defined in (1.6) is the semigroup of a Markov process with infinite lifetime. Furthermore, it follows that this process is $1/\alpha$-self-similar. More precisely, for any $x > 0$ and $c > 0$ the process (X_{ct}) under $\mathbb{P}_x^\uparrow$ has the same law as $c^{1/\alpha}X$ under $\mathbb{P}^\uparrow_{c^{-1/\alpha}x}$.

For $x > 0$ the process X is a positive self-similar Markov process (pssMp) under $\mathbb{P}_x^\uparrow$. It was shown by Lamperti (1972) that there exists a Lévy process ξ such that

$$X_t = x\exp(\xi_{\tau(tx^{-\alpha})}), \qquad \tau(r) = \inf\{s > 0 \mid I_s \geq r\}, \qquad I_s = \int_0^s \exp(\alpha\xi_u)\,\mathrm{d}u.$$

This representation is called the *Lamperti representation* of X. The particular case where the pssMp is a Lévy process conditioned to stay positive has been further studied by Caballero and Chaumont (2006).

To end the section on an example, assume that X is a standard Brownian motion under $\mathbb{P}$. Then X is strictly α-stable with $\alpha = 2$ and $\rho = 1/2$. Hence, $h(x) = x$ and one may deduce that $X^\uparrow$ is, in fact, a Bessel-3 process. In this case the Lévy process ξ in the Lamperti representation is a Brownian motion with unit variance and drift $1/2$, see e.g. Caballero and Chaumont (2006, p. 969).

1.3 Zooming in on a Lévy process

Zooming out is a classical idea in the theory of stochastic processes. For example, Donsker's theorem states that zooming out from a symmetric random walk results in a Brownian motion. In this section we briefly discuss the idea of zooming in on a Lévy process. Throughout we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Lévy process X defined on this space.

1.3.1 Zooming in at the origin

We say that X has a zooming-in limit if there exists a random variable $\widehat{X}_1$ and a positive scaling function $a\colon (0,\infty) \to (0,\infty)$ such that $\mathbb{P}(\widehat{X}_1 = 0) < 1$ and

$$a(\epsilon)X_\epsilon \xrightarrow{d} \widehat{X}_1 \quad \text{as } \epsilon \downarrow 0. \tag{1.8}$$

The limiting random variable $\widehat{X}_1$ is necessarily infinitely divisible so we may view it as a Lévy process $\widehat{X}$ evaluated at time 1. Furthermore, the convergence in (1.8) is equivalent to the functional convergence

$$(a(\epsilon)X_{\epsilon t}) \xrightarrow{d} \widehat{X} \quad \text{as } \epsilon \downarrow 0,$$

see Jacod and Shiryaev (1987, Cor. VII.3.6). Using standard 'convergence to types' arguments it follows that the limiting process $\widehat{X}$ must be strictly stable, see Ivanovs (2018, Thm. 1).

If (1.8) holds we say that X is in the domain of attraction of $\widehat{X}$. If X has a non-zero Brownian component then X is necessarily attracted to a (driftless) Brownian motion. If not, one has to look at the Lévy measure close to the origin to determine the domain of attraction (and if it exists). This is characterized in detail by Ivanovs (2018, Thm. 2).

1.3.2 Zooming in at the infimum and supremum

We assume that the Lévy process X is attracted to a strictly α-stable Lévy process $\widehat{X}$ with scaling function a. As in §1.2.1 we let $\underleftarrow{X}$ and $\underrightarrow{X}$ denote the pre-infimum and post-infimum processes for some fixed time interval $[0, T]$. As shown by Ivanovs (2018, Thm. 4) the identity in law (1.3) and the zooming-in assumption can be combined to obtain the join convergence

$$\big((a(\epsilon)\underleftarrow{X}_{\epsilon t}), (a(\epsilon)\underrightarrow{X}_{\epsilon t})\big) \xrightarrow{d} \big(-\widehat{X}^\downarrow, \widehat{X}^\uparrow\big) \quad \text{as } \epsilon \downarrow 0, \tag{1.9}$$

where $\widehat{X}^{\downarrow}, \widehat{X}^{\uparrow}$ are created from $\widehat{X}$ as described in §1.2.1. Note that Ivanovs (2018) considers the supremum rather than the infimum but by duality this is no different.

In Bisewski and Ivanovs (2020) the convergence in (1.9) was used to study threshold exceedance of a Lévy process X. If one defines $M = \sup_{t\in[0,1]} X_t$ and $M^{(n)} = \max_{0\le i\le n} X_{i/n}$ then $p^{(n)} = \mathbb{P}(M > x, M^{(n)} \le x)$ denotes the probability of failing to observe exceedance above the level $x > 0$ given the observations $X_{i/n}$ where $i = 0, 1, \ldots, n$. It turns out that the asymptotic behavior of $p^{(n)}$ (as $n \to \infty$) can be described using the scaling function a and the limit pair $(-\widehat{X}^{\downarrow}, \widehat{X}^{\uparrow})$.

Another application of the result concerns estimation of the supremum M given the observations $X_{i/n}$, $i = 0, 1, \ldots, n$. The obvious estimator is the maximum $M^{(n)}$ and Ivanovs (2018, Thm. 5) gives the convergence rate of the difference $M - M^{(n)}$. Again this can be formulated using a and $(-\widehat{X}^{\downarrow}, \widehat{X}^{\uparrow})$. This is explored further by Ivanovs and Podolskij (2022).

1.4 Graphical models for extremes

Graphical models and extreme value theory are two rather classical topics. More recently there has been advances in the area of graphical models within multivariate extreme value theory. This section aims to introduce some of these ideas.

1.4.1 Multivariate extreme value theory

We consider a sequence $Z(1), Z(2), \ldots$ of i.i.d. d-dimensional random vectors with distribution function F. One approach in multivariate extremes studies the vector of component-wise maxima $M(n)$ where $M_i(n) = \max_{k=1,\ldots,n} Z_i(k)$. We are then interested in any multivariate distribution function G which has non-degenerate marginals and can arise as the limit

$$\begin{aligned} G(x) &= \lim_{n\to\infty} \mathbb{P}\Big(\frac{M_1(n) - b_1(n)}{a_1(n)} \le x_1, \ldots, \frac{M_d(n) - b_d(n)}{a_d(n)} \le x_d\Big) \\ &= \lim_{n\to\infty} F^n(a_1(n)x_1 + b_1(n), \ldots, a_d(n)x_d + b_d(n)) \end{aligned}$$

for $x \in \mathbb{R}^d$, where $a(n) \in (0,\infty)^d$ and $b(n) \in \mathbb{R}^d$ for all n. In this case we say that F is in the domain of attraction of G. The collection of these limit distributions is called the class of *multivariate extreme value distributions* and it turns out to coincide with the class of *max-stable distributions*, see e.g. Resnick (2008, Prop. 5.9). A distribution function G is called max-stable if, for any $t > 0$, there exist $\alpha(t) \in (0,\infty)^d$ and $\beta(t) \in \mathbb{R}^d$ such that

$$G^t(x) = G(\alpha_1(t)x_1 + \beta_1(t), \ldots, \alpha_d(t)x_d + \beta_d(t))$$

for $x \in \mathbb{R}^d$.

It is common practice to standardize the marginals of G through a transformation of the random vectors as this allows one to focus on the dependence between components. For each $k \in \mathbb{N}$ and $i \in \{1, \ldots, d\}$ we apply the transformation $x \mapsto (1 - F_i(x))^{-1}$ to $Z_i(k)$, where F_i is the distribution function of $Z_i(1)$. We denote by $\tilde{F}$ the distribution function for the standardized vectors. Then, if F is in the domain of attraction of a max-stable distribution we have the convergence

$$\tilde{F}^n(nx) \to \tilde{G}(x) \quad \text{as } n \to \infty \tag{1.10}$$

for $x \in \mathbb{R}^d$, see Resnick (2008, Prop. 5.10). The limit $\tilde{G}$ is a max-stable distribution function with marginals $\tilde{G}_i(x) = \exp(-x^{-1})\,\mathbf{1}_{(0,\infty)}(x)$ for all i. A consequence of having standardized marginals is that

$$\tilde{G}^t(tx) = \tilde{G}(x) \quad \text{for all } t > 0, x \in \mathbb{R}^d,$$

see Resnick (2008, 5.24). This implies, in particular, that $\tilde{G}$ is max-infinitely divisible, i.e. for every $n \in \mathbb{N}$ there exists i.i.d. random vectors $U(1), \ldots, U(n)$ such that the vector of component-wise maxima has distribution function $\tilde{G}$.

The distribution function $\tilde{G}$ is characterized by its so-called *exponent measure* Λ. This is a measure on the cone $\mathcal{E} = [0, \infty)^d \setminus \{0\}$ and is related to $\tilde{G}$ by the formula

$$\tilde{G}(x) = \exp(-\Lambda(\mathcal{E} \setminus [0, x])),$$

where $[0, x] = [0, x_1] \times \cdots \times [0, x_d]$ for any $x \in \mathcal{E}$. Furthermore, Λ is homogeneous in the sense that $\Lambda(tB) = t^{-1}\Lambda(B)$ for any $t > 0$ and any Borel set $B \subseteq \mathcal{E}$.

The vector of component-wise maxima is just one object of interest in multivariate extremes. Another popular idea is the so-called *peaks-over-threshold* approach which considers the distribution of a random vector $\tilde{Z}$ conditioned on the event that at least one component is large. We assume that $\tilde{Z}$ has been normalized as above such that its distribution function $\tilde{F}$ satisfies (1.10). Then we have the convergence

$$n(1 - \tilde{F}(nx)) \to \Lambda(\mathcal{E} \setminus [0, x]) \quad \text{as } n \to \infty$$

for any $x \in \mathcal{E}$, see Resnick (2008, Prop. 5.17). We can now look at $n^{-1}\tilde{Z}$ conditioned on the event $\{\|\tilde{Z}\|_\infty > n\}$. For $x \in \mathcal{E}$ we find that

$$\begin{aligned}
\mathbb{P}(n^{-1}\tilde{Z} \leq x \mid \|\tilde{Z}\|_\infty > n) &= \frac{\mathbb{P}(\{n^{-1}\tilde{Z} \leq x\} \setminus \{n^{-1}\tilde{Z} \leq 1_d\})}{1 - \tilde{F}(n1_d)} \\
&= \frac{\mathbb{P}(\{n^{-1}\tilde{Z} \leq x\} \setminus \{n^{-1}\tilde{Z} \leq 1_d \wedge x\})}{1 - \mathbb{P}(n^{-1}\tilde{Z} \leq 1_d)} \\
&= \frac{1 - \tilde{F}(n(1_d \wedge x)) - (1 - \tilde{F}(nx))}{1 - \tilde{F}(n1_d)} \\
&\to \frac{\Lambda(\mathcal{E} \setminus [0, 1_d \wedge x]) - \Lambda(\mathcal{E} \setminus [0, x])}{\Lambda(\mathcal{E} \setminus [0, 1_d])}
\end{aligned}$$

as $n \to \infty$, where $1_d \in \mathbb{R}^d$ has 1 in each component and $1_d \wedge x$ is the component-wise minimum of 1_d and x. The limit is the distribution function of a so-called *multivariate Pareto distribution*. If Y has this distribution function then the support of Y is contained in the L-shaped space $\mathcal{L} = \{x \in \mathcal{E} \mid \|x\|_\infty > 1\}$, and for any measurable $A \subseteq \mathcal{L}$ we have that

$$\mathbb{P}(Y \in A) = \frac{\Lambda(A)}{\Lambda(\mathcal{L})}.$$

1.4.2 Conditional independence and graphical models

For $d \in \mathbb{N}$ let $V = \{1, \ldots, d\}$ and consider a random vector Z taking values in $\mathbb{R}^d$. For any non-empty $A \subseteq V$ we denote by Z_A the A-component of Z. For disjoint subsets $A, B, C \subseteq V$ we can ask if Z_A is conditionally independent of Z_B given Z_C, typically written as $Z_A \perp\!\!\!\perp Z_B \mid Z_C$. If $C = \emptyset$ this is just regular independence and we write $Z_A \perp\!\!\!\perp Z_B$.

Pairing V with an edge set $E \subseteq V \times V$ results in the graph $G = (V, E)$. The latter is assumed to be undirected, meaning that we do not distinguish between the edges (i, j) and (j, i) for $i, j \in V$. The distribution of Z is said to satisfy the *pairwise Markov property* wrt. G if $Z_i \perp\!\!\!\perp Z_j \mid Z_{V \setminus \{i,j\}}$ for any $(i, j) \notin E$. Furthermore, the distribution of Z satisfies the stronger *global Markov property* wrt. G if $Z_A \perp\!\!\!\perp Z_B \mid Z_C$ for any disjoint $A, B, C \subseteq V$ such that C separates A and B in G. Generally, the pairwise Markov property does not imply the global Markov property. However, a sufficient condition for equivalence of the two properties is that Z has a positive and continuous density wrt. a product measure on $\mathbb{R}^d$, see Lauritzen (1996, Thm. 3.9).

1.4.3 Conditional independence in extremes

Studying the vector of component-wise maxima of n i.i.d. d-dimensional random vectors is, in the limit, the same as studying an associated max-stable distribution as we saw in the previous subsection. If Z is a d-dimensional max-stable vector and we further assume it has a positive and continuous Lebesgue density on $(0,\infty)^d$ then, in fact, there is the implication

$$Z_A \perp\!\!\!\perp Z_B \mid Z_{V\setminus(A\cup B)} \quad \Rightarrow \quad Z_A \perp\!\!\!\perp Z_B,$$

see Papastathopoulos and Strokorb (2016, Thm. 1). A consequence of this is that the graphical structure of such a max-stable distribution is somewhat trivial. Indeed, Z will satisfy the global Markov property wrt. a graph which consists entirely of isolated cliques.

Conditional independence is less straightforward when we consider a multivariate Pareto distributed vector Y with exponent measure Λ. Even independence of e.g. Y_1 and Y_2 becomes an issue since the support of Y might not be a product set. To work around this problem one can look at Y^k, a vector with the distribution of Y conditioned on the event $\{Y_k > 1\}$. Since the only restriction on the support of Y^k is that it must be contained in the set $\{x \in [0,\infty)^d \mid x_k > 1\}$ it is, in particular, allowed to be on product form. Thus, it is more natural to look at independence and conditional independence for Y^k. If $A, B, C \subseteq V$ are disjoint and we have $Y^k_A \perp\!\!\!\perp Y^k_B \mid Y^k_C$ for all $k \in V$ then Engelke and Hitz (2020) say that Y_A is conditionally independent of Y_B given Y_C. This is written $Y_A \perp_e Y_B \mid Y_C$. The distribution of Y is then called an extremal graphical model relative to a graph G if it satisfies the pairwise Markov property wrt. G. The pairwise Markov property implies the global Markov property when Y has a positive and continuous Lebesgue density on $\mathcal{L}$, see Engelke and Hitz (2020, Thm. 1).

1.5 Paper I

Positive self-similar Markov processes play an important role in studying, for example, the supremum or infimum of a strictly stable Lévy process, with a couple of examples already mentioned in §1.3. In applied probability theory it is sometimes difficult to do calculations and typically one must resort to simulation methods. But how can we simulate a Lévy process conditioned to stay positive? This question was the inspiration for Paper I as the literature did not provide any good answers.

Initially one might suggest using Bertoin's construction from §1.2.1. That is, we simulate the Lévy process on a fine grid and glue together the parts of the path that are positive. With this approach, simulation of the conditioned process until some time $T > 0$ requires simulation of the Lévy process until time α^+_T, which is the time point when the process has been positive for a total time of T. However, $\mathbb{E}[\alpha^+_T] = \infty$ for many Lévy processes, such as the standard Brownian motion. Therefore, this approach is not suited for Monte Carlo methods as it will take a significant amount of time to run many simulations.

The idea in the paper is to use the Lamperti representation. Recall that a pssMp X, starting at $x > 0$, can be represented as

$$X_t = x \exp(\xi_{\tau(tx^{-\alpha})}),$$

where ξ is a Lévy process, $1/\alpha$ is the self-similarity index and

$$\tau(r) = \inf\{s > 0 \mid I_s \geq r\}, \qquad I_s = \int_0^s \exp(\alpha\xi_u)\,\mathrm{d}u.$$

It is assumed that the process ξ is available at times i/n, $i \in \mathbb{N}$ and in the paper we suggest a simple approximation scheme. With $\xi^{(n)}$ being the discretization of ξ, i.e. $\xi^{(n)}_t = \xi_{[tn]/n}$, we define

$$X_t^{(n)} = x \exp(\xi^{(n)}_{\tau_n(tx^{-\alpha})}),$$

where

$$\tau_n(r) = \inf\{s > 0 \mid I_s^{(n)} \geq r\}, \qquad I_s^{(n)} = \int_0^s \exp(\alpha \xi_u^{(n)}) \, \mathrm{d}u.$$

The main quantity of interest is the discretization error $X_t - X_t^{(n)}$.

Some kind of regularity of the Lévy process ξ is essential and the zooming-in property in (1.8) turns out to be what we need. Therefore, we assume throughout that there exist a scaling function a and a random variable $\widehat{\xi}_1$ (which is not a.s. zero) such that $a(\epsilon)\xi_\epsilon \xrightarrow{d} \widehat{\xi}_1$ as $\epsilon \downarrow 0$. One of the first main results provides a convergence rate for the scaled relative error. That is, for $t > 0$ we establish bounds $\ell^{(n)}(t), u^{(n)}(t)$ such that

$$\ell^{(n)}(t) \leq a(n^{-1}) \frac{X_t - X_t^{(n)}}{X_t} \leq u^{(n)}(t),$$

and it is further shown that the pair $(\ell^{(n)}(t), u^{(n)}(t))$ converges in distribution to a non-trivial limit as $n \to \infty$. A limit theorem for the scaled relative error is also proved but doing this requires an additional assumption. Namely that $(\tau(r), \xi_{\tau(r)})$ is absolutely continuous for every small $r > 0$.

While it does not seem immediately related to the original problem the paper also briefly considers the zooming-in property for a pssMp X. It turns out that X has such a property if and only if the Lévy process ξ satisfies the zooming-in assumption. This result has a rather useful consequence; if X^0 is a strictly α-stable Lévy process and X has the law of X^0 conditioned to stay positive then the Lévy process ξ satisfies the zooming-in assumption with $\widehat{\xi}_1 = X_1^0$ and $a(\epsilon) = \epsilon^{-1/\alpha}$.

The paper gets somewhat technical in certain places but outlining the general ideas is not too difficult. Firstly, one can rewrite the scaled relative error using the mean value theorem. The result is the identity

$$a(n) \frac{X_t - X_t^{(n)}}{X_t} = a(n^{-1})(\xi_{\tau(r)} - \xi_{[\tau_n(r)n]/n})(1 + o_{\mathbb{P}}(1)),$$

where $r = tx^{-\alpha}$ and $o_{\mathbb{P}}(1)$ is a term converging to zero in probability. Then, if we further write $[\tau_n(r)n] = \tau(r) + ([\tau_n(r)n] - \tau(r)n)/n$ we see that we are essentially zooming in on ξ at time $\tau(r)$, and it becomes clear that the zooming-in assumption is natural.

In order to proceed we must study the quantity $[\tau_n(r)n] - \tau(r)n$ as $n \to \infty$. Here we rely on a result by Jacod et al. (2003) which provides a functional limit theorem for the discretized integration error

$$(\Delta_t^{(n)})_{t \geq 0} = n(I_{[tn]/n} - I^{(n)}_{[tn]/n})_{t \geq 0}.$$

Initially, this lets us derive a convergence result for $n(\tau_n(r) - \tau(r))$. Writing the quantity of interest as

$$[\tau_n(r)n] - \tau(r)n = [\{\tau(r)n\} + n(\tau_n(r) - \tau(r))] - \{\tau(r)n\}$$

suggests that we should control $\{\tau(r)n\}$. The previously mentioned assumption about absolute continuity of $\tau(r)$ ensures that $\{\tau(r)n\}$ converges in distribution to a standard uniform random variable.

An important point is that the sequence $(\Delta_t^{(n)})_{t \geq 0}$ actually converges *stably*. This means that it converges in distribution jointly with any $\sigma(\xi)$-measurable random variable, see Podolskij and Vetter (2010) for more details. This additional property is used frequently throughout the paper and most convergence results are again formulated with stable convergence.

The paper does not solve the original simulation problem but it reduces it to simulation of the Lévy process ξ. The latter is a separate issue but it is worth noting that the characteristic triplet of ξ is provided in Caballero and Chaumont (2006, Cor. 2). It might be worth investigating if the approximate simulation method treated by Asmussen and Rosiński (2001) could be applied here.

An interesting question is if it is possible to infer the characteristics of ξ given high-frequency observations of the pssMp X. It is possible to obtain ξ from X using the fact that

$$\tau(t) = \int_0^{x^\alpha t} X_s^{-\alpha} \, \mathrm{d}s \quad \mathbb{P}_x\text{-a.s.},$$

see Caballero and Chaumont (2006). However, obtaining results similar to those in this paper requires a different approach. For example, the convergence result for $n(\tau_n(r) - \tau(r))$ is essential and the proof relies heavily on the result by Jacod et al. (2003). However, the latter does not apply in this case so one must deal with it differently.

The idea in this paper could possibly be applied in other areas as identities similar to the Lamperti representation exist elsewhere. For instance, a continuous-state branching processes can be represented as a time-changed Lévy process, see e.g. Kyprianou (2006, Thm. 10.2). One technical difference is that the underlying Lévy process no longer has infinite lifetime.

1.6 Paper II

Conditioning a univariate Lévy process to stay positive is a useful concept because of its relation to the post-infimum process. The goal of Paper II was to come up with a multivariate generalization of this particular type of conditioning as this did not exist in the literature at the time.

The seemingly natural multivariate analogue of conditioning to be positive is to condition the process to stay in a certain half-space. For a fixed normal vector $\eta \in \mathbb{R}^d$ we associate the half-space $S = \{x \in \mathbb{R}^d \mid \langle x, \eta \rangle > 0\}$. Conditioning a d-dimensional Lévy process X to stay in S corresponds to conditioning the projected process $Z(\,\cdot\,) = \langle X(\,\cdot\,), \eta \rangle$ to stay positive. Since Z is a univariate Lévy process it would be natural define a function h as in (1.4) (with Z playing the role of X) and then study the semi-group

$$p_t^{\uparrow}(x, \mathrm{d}y) = \frac{h(\langle y, \eta \rangle)}{h(\langle x, \eta \rangle)} \mathbb{P}_x(X_t \in \mathrm{d}y, \underline{Z}_t > 0), \qquad t > 0, x, y \in S. \tag{1.11}$$

However, what we really want is a construction like the one given by Bertoin (1993) since it provides the connection to the post-infimum process.

To generalize Bertoin's construction one must define a multivariate version of the process Y^+ from (1.1). Right away this looks like it involves coming up with an appropriate multivariate generalization of the local time L. Note that the local time at 0 of the multivariate process X cannot be used since it is still a univariate process. To get around this we use ideas from the proof of Bertoin (1993, Thm. 3.1) to define

$$Y_t^+ = -\int_{[-t,0]} \mathbf{1}_{\{\langle \tilde{X}_{s-}, \eta \rangle > 0\}} \, \mathrm{d}\tilde{X}_s, \qquad t \geq 0,$$

where $\tilde{X}_t = X_{(-t)-}$. In a similar fashion we construct Y^- and through time-changes (like in the univariate case) we can define two processes $X^{\uparrow}$ and $X^{\downarrow}$. This leads to the desired distributional identity

$$(\underset{\leftarrow}{X}, \underset{\rightarrow}{X}) \overset{d}{=} (-X^{\downarrow}, X^{\uparrow})$$

under the measure $\mathbb{P}^T$. Here $\underset{\leftarrow}{\underline{X}}$ is the reversed *directional* pre-infimum process and $\underset{\rightarrow}{\underline{X}}$ is the *directional* post-infimum process, i.e. the processes looking back and forward respectively from the time of the infimum of the projected process Z.

We establish some of the fundamental properties of the processes $X^\uparrow$ and $X^\downarrow$. As in the univariate case these are independent Markov processes and we show that the semigroup of the former is indeed given by (1.11). The convergence in (1.7) showed that the terminology *conditioned to stay positive* is sensible since this law can be obtained by conditioning the process to stay positive until an independent exponential time where the rate vanishes. In the multivariate case we show that a similar result holds.

We further show how the construction of $X^\uparrow$ behaves under linear transformations. This is used to study the example where X is a driftless Brownian motion. By multiplying with appropriate matrices we show that this reduces to the case where η is the first standard basis vector and $\mathrm{Var}(X)$ is the identity matrix.

In §1.3 we discussed how zooming in at the infimum of a Lévy process is described using a limiting Lévy process conditioned to stay positive. To illustrate how this looks in the multivariate setting we assume that the Brownian part B of X is such that $\langle B_1, \eta \rangle$ is not a.s. zero. We have the convergence

$$\sqrt{n}(\underset{\leftarrow}{\underline{X}}_{\cdot/n}, \underset{\rightarrow}{\underline{X}}_{\cdot/n}) : \mathbb{P}^1 \xrightarrow{d} (-B^\downarrow, B^\uparrow),$$

where the notation $: \mathbb{P}^1$ means that the left-hand side should be viewed under $\mathbb{P}^1$, i.e. where X is killed at time 1.

It seems interesting to go beyond studying the directional infimum, i.e. the value of X at the time of the infimum of the projected process Z, for some time interval $[0, T]$. What we want to consider instead is the value of X when it is farthest from the origin (also for a bounded time interval). Again we can define associated pre- and post-maximum processes $\overleftarrow{X}$ and $\overrightarrow{X}$ but analyzing these turns out to be significantly more difficult since it is not clear how to obtain a result similar to the rather essential distributional identity in (1.3). Instead we provide a conjecture for zooming in at the point farthest from the origin. The proposed limiting object is of the form $(-B^\downarrow, B^\uparrow)$ where the direction η is now a random variable. We simulate an example in dimension $d = 2$ and the results look promising.

As mentioned above it is completely straightforward to obtain a multivariate analogue of conditioning a Lévy process to stay positive since one can just define a semigroup as in (1.11) and proceed from there. However, conditioning just for the sake of conditioning was not our goal. Instead we succeeded in providing a multivariate version of Bertoin's construction along with an identity like the one in (1.3).

1.7 Paper III

Knowledge about zooming-in properties of stochastic processes is useful. This is evident from Paper I below and also from e.g. Ivanovs (2018), Bisewski and Ivanovs (2020) and Ivanovs and Podolskij (2022). These papers all consider Lévy processes and the idea behind Paper III was instead to analyze the zooming-in properties of diffusions. This particular class was chosen because a diffusion process behaves locally as a scaled Brownian motion, and for the latter we understand the zooming-in properties quite well. Hence, the hope was to make use of this knowledge and intuition.

The paper considers a weak solution (X, W) to the SDE

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t \quad \text{and} \quad X_0 = x_0,$$

where $x_0 \in \mathbb{R}$ and W is a standard Brownian motion. Assuming certain regularity conditions on the functions μ and σ the paper presents two zooming-in results. These are distributional limit theorems and just like in Paper I we even have stable convergence. The first result concerns zooming in at a fixed time point $T > 0$ and the result provides a distributional limit for the process $(\epsilon^{-1/2}(X_{T+\epsilon t} - X_T))_{t\in\mathbb{R}}$. The limit is a two-sided Brownian motion scaled by $\sigma(X_T)$ where the two sides, corresponding to $t < 0$ and $t > 0$, are conditionally independent given the scaling $\sigma(X_T)$. The second result establishes a zooming-in limit at the supremum over a bounded time interval. We consider the two-sided process $(\epsilon^{-1/2}(X_{m+\epsilon t} - \overline{X}))_{t\in\mathbb{R}}$ where m is the time of supremum and $\overline{X}$ is the value of the supremum. The limiting object is constructed using two independent Bessel-3 processes, one for $t < 0$ and the other for $t > 0$, which are both scaled by $-\sigma(\overline{X})$. As an application of the second result the paper studies the problem of estimating the supremum based on equidistant discrete observations. This is similar to the problem considered in Ivanovs (2018, §6.1).

As mentioned we rely on zooming-in knowledge for the Brownian motion, both at fixed times but also at the supremum. Another key ingredient in the proofs is a certain way of representing the diffusion X. If $x_0 = 0$ and $\mu \equiv 0$ then X may be represented as a time-changed Brownian motion. To be precise, $X_t = \tilde{W}_{[X]_t}$, where $[X]$ denotes the quadratic variation of X and $\tilde{W}$ is a Brownian motion. Zooming in is, in some sense, similar to differentiating and this allows us to employ a 'chain rule' approach. It remains to argue that the assumption $\mu \equiv 0$ can be dropped. Intuitively, this is rather easy since the drift vanishes linearly with ϵ whereas we only scale by $\epsilon^{-1/2}$.

The result for zooming in at a fixed time is essentially just a precise formulation of the intuitive understanding of a diffusion process; locally it behaves as a scaled Brownian motion. This should be rather straightforward to prove for $t > 0$, but including $t < 0$ makes it a bit more difficult. An alternative approach would be to consider time-reversal but for this we would need that the time-reversed process, say $(X_{1-t})_{t\in[0,1]}$, is a diffusion. This is not necessarily satisfied. For example, to prove that this holds Haussmann and Pardoux (1985) need assume that μ and σ are locally Lipschitz continuous.

Zooming in at the supremum of a Lévy process has already been characterized by Ivanovs (2018) and Paper III deals with diffusions. In addition to these there seem to be other classes of processes which could be suitable for further study. Two examples are already mentioned in Paper III; positive self-similar Markov processes and continuous-state branching processes. Both can be represented as time-changed Lévy processes, suggesting that one might be able to employ ideas like the one mentioned above. Solutions to Lévy driven SDEs is another direction that could be explored. Zooming in at a fixed time is studied by Reker (2023), although only $t > 0$ is considered. One can imagine that results for zooming in at the supremum are also obtainable in this setting using the existing results for Lévy processes.

1.8 Paper IV

Recently Engelke and Hitz (2020) defined a notion of conditional independence for the components of a multivariate Pareto distributed vector Y using the vectors Y^k introduced in §1.4. This idea was extended by Engelke et al. (2022), resulting in a generalized notion of conditional independence for measures on $\mathbb{R}^d$. The interesting case is to consider measures that may have infinite mass, such as exponent measures or Lévy measures. For disjoint sets $A, B, C \subseteq V = \{1, \dots, d\}$ they define what it means for A and B to be conditionally independent given C with respect to a measure Λ, and in this case they write $A \perp B \mid C\,[\Lambda]$.

For a d-dimensional Lévy process we know that the dependence between components is determined by the covariance matrix Σ and the Lévy measure Λ. The motivation behind Paper IV was to study the above mentioned conditional independence in the context of Lévy measures in order to understand what it means in terms of the distribution of the process.

We recall that a Lévy process $\mathbf{X}$ can be written as the sum $\mathbf{X} = \mathbf{B}+\mathbf{J}$ of its Brownian and jump parts. We first show that the conditional independence $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C$ is equivalent to having the same conditional independence for both $\mathbf{B}$ and $\mathbf{J}$. This implies that we can study conditional independence for these parts separately. In the case of the Brownian part it is well-understood how the dependence between components is characterized by the covariance Σ. For example, if Σ is invertible then $\mathbf{B}_i$ and $\mathbf{B}_j$ are conditionally independent given all the other components if $(\Sigma^{-1})_{ij} = 0$. Then one might ask how the conditional independence structure of the jump part is characterized by the Lévy measure. Our main result provides an answer to this question. We show that

$$\mathbf{J}_A \perp\!\!\!\perp \mathbf{J}_B \mid \mathbf{J}_C \quad \Leftrightarrow \quad A \perp B \mid C\,[\Lambda].$$

For this to hold we require a certain somewhat technical assumption which ensures that we are always working with either no jumps or with infinite jump activity. If the process is α-stable then this assumption is automatically satisfied due to homogeneity of the Lévy measure. For the class of stable processes we study graphical models, where the underlying graph is given by a tree. This includes theoretical properties of the Lévy measure but also a method for consistent estimation of the tree given discrete observations. This is illustrated with both simulations and stock price data.

In order to prove the results we employ ideas from the typical proof of the Lévy–Itô decomposition. More precisely, we use the fact that the Brownian and jump parts can be constructed from the full path of the Lévy process. This is important when we want to condition on $\mathbf{X}_C$ for some $C \subseteq \{1, \ldots, d\}$. Indeed, for an integrable random variable Z it lets us write

$$\mathbb{E}[Z \mid \mathbf{X}_C] = \mathbb{E}[Z \mid \mathbf{B}_C, \mathbf{J}_C]$$

almost surely. To prove results specifically for stable processes we rely heavily on homogeneity of the measure. In fact, we can borrow ideas from Engelke and Volgushev (2022) since many of these can be extended to any homogeneous measure. There is, however, a difference when we want to estimate the underlying graph from discrete observations. In extremes one uses attraction to a multivariate Pareto distribution but for Lévy processes we must do something different. In the end we use a result which follows from Sato (1999, Cor. 8.9). If $E \in \mathcal{B}(\mathbb{R}^d)$ is bounded away from the origin then

$$\Lambda(E) = \lim_{t \to 0} t^{-1}\mathbb{P}(X(t) \in E).$$

Hence, as we increase the sampling frequency we get closer to 'observing' the Lévy measure. This lets us learn the conditional independence structure of Λ and therefore also that of the process.

It is our understanding that spatial conditional independence and graphical models for Lévy processes are rather unexplored topics. The introduction in Paper IV already contains an overview of related works but we will give brief recap here. For time series there is the famous concept of Granger causality, see Granger (1969), and for continuous time processes this idea has been generalized to various notions of *local independence*, see e.g. Didelez (2008). The general idea is to study how the past of one component influences influences the next 'step' of another component. However, this is not particularly interesting for Lévy processes due to the Markov property. Misra and Kuruoglu (2016) define what they call *α-stable* *graphical models*. This is, in fact, related to the work in Paper IV, see Engelke et al. (2022, §7.3), but the models they obtain are very simple since the resulting Lévy measures are forced to have a particular structure.

References

D. Applebaum. *Lévy processes and stochastic calculus*, volume 116 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2009. ISBN 978-0-521-73865-1.

S. Asmussen and J. Rosiński. Approximations of small jumps of Lévy processes with a view towards simulation. *J. Appl. Probab.*, 38(2):482–493, 2001. ISSN 0021-9002.

O. E. Barndorff-Nielsen, T. Mikosch, and S. I. Resnick, editors. *Lévy processes*. Birkhäuser Boston, Inc., Boston, MA, 2001. ISBN 0-8176-4167-X. Theory and applications.

J. Bertoin. Splitting at the infimum and excursions in half-lines for random walks and Lévy processes. *Stochastic Process. Appl.*, 47(1):17–35, 1993. ISSN 0304-4149.

J. Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996. ISBN 0-521-56243-0.

K. Bisewski and J. Ivanovs. Zooming-in on a Lévy process: failure to observe threshold exceedance over a dense grid. *Electron. J. Probab.*, 25:Paper No. 113, 33, 2020.

M. E. Caballero and L. Chaumont. Conditioned stable Lévy processes and the Lamperti representation. *J. Appl. Probab.*, 43(4):967–983, 2006. ISSN 0021-9002.

L. Chaumont. Conditionings and path decompositions for Lévy processes. *Stochastic Process. Appl.*, 64(1):39–54, 1996. ISSN 0304-4149.

L. Chaumont and R. A. Doney. On Lévy processes conditioned to stay positive. *Electron. J. Probab.*, 10:no. 28, 948–961, 2005. ISSN 1083-6489.

R. Cont and P. Tankov. *Financial modelling with jump processes*. Chapman & Hall/CRC Financial Mathematics Series. Chapman & Hall/CRC, Boca Raton, FL, 2004. ISBN 1-5848-8413-4.

V. Didelez. Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 70(1):245–264, 2008. ISSN 13697412, 14679868.

R. A. Doney. Tanaka's construction for random walks and Lévy processes. In *Séminaire de Probabilités XXXVIII*, volume 1857 of *Lecture Notes in Math.*, pages 1–4. Springer, Berlin, 2005.

L. Döring and P. Weissmann. Stable processes conditioned to hit an interval continuously from the outside. *Bernoulli*, 26(2):980–1015, 2020. ISSN 1350-7265.

L. Döring, A. R. Watson, and P. Weissmann. Lévy processes with finite variance conditioned to avoid an interval. *Electron. J. Probab.*, 24:Paper No. 55, 32, 2019.

S. Engelke and A. S. Hitz. Graphical models for extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 82(4):871–932, 2020. ISSN 1369-7412. With discussions.

S. Engelke and S. Volgushev. Structure learning for extremal tree models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 84(5):2055–2087, 2022. ISSN 1369-7412,1467-9868.

S. Engelke, J. Ivanovs, and K. Strokorb. Graphical models for infinite measures with applications to extremes and Lévy processes. *arXiv preprint arXiv:2211.15769*, 2022.

C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969. ISSN 00129682, 14680262.

U. G. Haussmann and E. Pardoux. Time reversal of diffusion processes. In *Stochastic differential systems (Marseille-Luminy, 1984)*, volume 69 of *Lect. Notes Control Inf. Sci.*, pages 176–182. Springer, Berlin, 1985. ISBN 3-540-15176-1.

J. Ivanovs. Zooming in on a Lévy process at its supremum. *Ann. Appl. Probab.*, 28(2): 912–940, 2018. ISSN 1050-5164.

J. Ivanovs and M. Podolskij. Optimal estimation of the supremum and occupation times of a self-similar Lévy process. *Electron. J. Stat.*, 16(1):892–934, 2022.

J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1987. ISBN 3-540-17882-1.

J. Jacod, A. Jakubowski, and J. Mémin. On asymptotic errors in discretization of processes. *Ann. Probab.*, 31(2):592–608, 2003. ISSN 0091-1798.

O. Kallenberg. *Foundations of modern probability*, volume 99 of *Probability Theory and Stochastic Modelling*. Springer, Cham, third edition, 2021. ISBN 978-3-030-61871-1; 978-3-030-61870-4.

A. E. Kyprianou. *Introductory lectures on fluctuations of Lévy processes with applications*. Universitext. Springer-Verlag, Berlin, 2006. ISBN 978-3-540-31342-7; 3-540-31342-7.

A. E. Kyprianou and J. C. Pardo. *Stable Lévy Processes via Lamperti-Type Representations*. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2022.

J. Lamperti. Semi-stable Markov processes. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 22:205–225, 1972.

S. L. Lauritzen. *Graphical models*, volume 17 of *Oxford Statistical Science Series*. The Clarendon Press, Oxford University Press, New York, 1996. ISBN 0-19-852219-3. Oxford Science Publications.

N. Misra and E. E. Kuruoglu. Stable graphical models. *Journal of Machine Learning Research*, 17(168):1–36, 2016.

I. Papastathopoulos and K. Strokorb. Conditional independence among max-stable laws. *Statist. Probab. Lett.*, 108:9–15, 2016. ISSN 0167-7152.

M. Podolskij and M. Vetter. Understanding limit theorems for semimartingales: a short survey. *Stat. Neerl.*, 64(3):329–351, 2010. ISSN 0039-0402.

J. Reker. Short-time behavior of solutions to Lévy-driven stochastic differential equations. *Journal of Applied Probability*, page 1–16, 2023.

S. I. Resnick. *Extreme values, regular variation and point processes.* Springer Series in Operations Research and Financial Engineering. Springer, New York, 2008. ISBN 978-0-387-75952-4. Reprint of the 1987 original.

K.-i. Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 1999. ISBN 0-521-55302-4. Translated from the 1990 Japanese original, Revised by the author.

H. Tanaka. Time reversal of random walks in one-dimension. *Tokyo J. Math.*, 12(1):159–174, 1989. ISSN 0387-3870.

Paper I

Discretization of the Lamperti representation of a positive self-similar Markov process

Jevgenijs Ivanovs
Department of Mathematics, Aarhus University

Jakob D. Thøstesen
Department of Mathematics, Aarhus University

Abstract. This paper considers discretization of the Lévy process appearing in the Lamperti representation of a strictly positive self-similar Markov process. Limit theorems for the resulting approximation are established under some regularity assumptions on the given Lévy process. Additionally, the scaling limit of a positive self-similar Markov process at small times is provided. Finally, we present an application to simulation of self-similar Lévy processes conditioned to stay positive.

Keywords: Exponential functional, Lamperti representation, positive self-similar Markov process, small time behavior, stable Lévy process conditioned to stay positive

2010 MSC: 60F17; 60G18; 60G51

I.1 Introduction

Positive self-similar Markov processes (pssMp) have received a lot of attention in recent years, see Baurdoux et al. (2016), Chaumont et al. (2012) and the survey by Pardo and Rivero (2013). One class of examples is given by self-similar Lévy processes 'conditioned' to stay positive, which arise in various limit theorems concerned with extremes, first passage times and Skorokhod reflection (Asmussen and Ivanovs, 2018; Ivanovs, 2018; Ivanovs and Podolskij, 2020). Recall that $X = (X_t)_{t\geq 0}$ is a pssMp if it is a positive strong Markov process with the self-similarity property: $(X_{tc})_{t\geq 0}$ with $X_0 = x > 0$ has the law of $(c^{1/\alpha}X_t)_{t\geq 0}$ with $X_0 = c^{-1/\alpha}x$ for any $c > 0$, where $1/\alpha > 0$ is sometimes called the Hurst index. Throughout this work we restrict our attention to strictly positive X.

The fundamental result of Lamperti (1972) states that every pssMp X (not hitting 0) can be represented via a Lévy process ξ as follows:

$$X_t = x\exp(\xi_{\tau(tx^{-\alpha})}), \qquad \tau(r) := \inf\{s > 0 \mid I_s \geq r\}, \qquad I_s := \int_0^s \exp(\alpha\xi_u)\,\mathrm{d}u \tag{I.1}$$

where $\limsup_{t\to\infty} \xi_t = \infty$ a.s., and $x > 0$ is a given starting position. Moreover, this relation can be inverted to obtain ξ in terms of X. The Lamperti representation is key for deriving various properties of pssMp (Pardo, 2009). Furthermore, it also provides a way to simulate from the law of X, which is important in application of the above mentioned limit theory and beyond.

The purpose of this paper is to investigate the basic discretization scheme, where the path of the Lévy process ξ is sampled at equidistant times $i/n, i \in \mathbb{N}$. Throughout this work we assume that the increment $\xi_{1/n}$ can be sampled exactly and efficiently. This allows to approximate the integral function I, which is then used to construct an approximation $X^{(n)}$ of X at the times of interest. Our main result is the limit theorem for the scaled error $a_n(X_t - X_t^{(n)})$ as $n \to \infty$, as well as its multidimensional version concerning a finite set of times, see Corollary I.6. This result crucially depends on the limit theory for the integrated process error in Jacod et al. (2003), which is extended to include zooming-in on ξ (Ivanovs, 2018) at inverse times.

A result of independent interest is presented in Theorem I.11, which complements the classical law of the iterated logarithm for a pssMp at small times (Lamperti, 1972, Thm. 7.1). We show that $a_n(X_{t/n} - x)_{t\geq 0}$ has a non-trivial weak limit as $n \to \infty$ under the obvious regularity condition that there is weak convergence to a non-zero limit for some fixed $x, t > 0$ and some positive function a_n. Furthermore, this assumption is equivalent to the regularity of the underlying Lévy process, which we assume in the above discussed approximation theory.

This work has been initially motivated by the problem of simulating a strictly stable Lévy processes conditioned to stay positive, see Engelke and Ivanovs (2016, §4) for various available representations. Importantly, the most obvious methods result in infinite expected running times. One of the reasons is that for an oscillating Lévy process the first passage time over a fixed level has infinite expectation. In this regard we note that González Cázares et al. (2019) recently provided an ε-strong simulation algorithm for the convex minorants of stable meanders, which are closely related to conditioned processes. Our method amounts to discretization of the Lévy process ξ in (I.1) which is applicable to a much broader class of pssMps. We supplement this method with a limit theorem for the relative error showing that it decays with the rate $n^{-1/\alpha}$ in the case of a strictly α-stable Lévy processes conditioned to stay positive.

Even though there is a large body of literature on high-frequency statistics and discretization of stochastic processes, see the monograph by Jacod and Protter (2011), discretization of the Lamperti transform has not yet been considered neither in the context of pssMps nor for continuous-state branching processes. Our results build upon existing limit theory, but also employ various novel ideas and methods. One of the main technical challenges was to incorporate the convergence of the fraction part of $\tau(r)n$ into the main limit result in Theorem I.5.

The structure of this paper is as follows. We start with the definitions, assumptions and necessary basic theory in §I.2. The limit theory for the approximation is derived in §I.3 relying on the joint stable convergence of some fundamental objects which is proven later in §I.4. The scaling limit of a pssMp is studied in §I.5 relying on a basic convergence result for Lévy processes which may be of an independent interest. In §I.6 these results are applied to self-similar Lévy processes conditioned to stay positive, where we also provide a numerical illustration in the simplest setting of a standard Brownian motion. We conclude with §I.7 providing comments about the density assumption and the trapezoidal approximation of the integral.

I.2 Definitions and prerequisites

I.2.1 Fundamentals

We work with càdlàg processes on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ and use the Skorokhod J_1 topology. Let $\xi = (\xi_t)_{t\geq 0}$ be a Lévy process, that is, an adapted càdlàg process starting at the origin with the property that $\xi_{t+s} - \xi_t$ is independent of $\mathcal{F}_t$ and has the law of ξ_s for any $t, s \geq 0$. Furthermore, as indicated above we assume that

$$\limsup_{t\to\infty} \xi_t = \infty \quad \text{a.s.}, \tag{I.2}$$

which is satisfied, for example, if ξ_1 is integrable and $\mathbb{E}\xi_1 \geq 0$, excluding the trivial 0 process.

To make the results slightly cleaner we shall extend ξ to the real line. We do so by letting $(-\xi_{(-t)-})_{t\geq 0}$ be an independent copy of the Lévy process $(\xi_t)_{t\geq 0}$, where the left limit is needed to get a càdlàg path over the real line. Note that the increments are still stationary and independent. Furthermore,

$$\xi_T \stackrel{d}{=} \operatorname{sign}(T)\xi_{|T|} \tag{I.3}$$

for any random $T \in \mathbb{R}$ independent of ξ, because Lévy processes do not jump at fixed times.

The concept of stable convergence (Aldous and Eagleson, 1978; Rényi, 1963) is fundamental in discretization of processes (Jacod and Protter, 2011; Podolskij and Vetter, 2010). Consider a sequence of random variables Z_n defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in some Polish space. The sequence Z_n is said to converge stably to Z ($Z_n \stackrel{st}{\to} Z$) defined on an extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ if

$$\mathbb{E}[f(Z_n)Y] \to \widetilde{\mathbb{E}}[f(Z)Y] \tag{I.4}$$

for all bounded continuous functions f and all bounded $\mathcal{F}$-measurable Y, see also (I.16) below for further intuition. The standard example concerns Z being independent of $\mathcal{F}$, and then the term mixing convergence is sometimes used.

I.2.2 Approximation

Consider the discretized process $\xi^{(n)}$ given by $\xi^{(n)}_t = \xi_{[tn]/n}$, where $[x]$ denotes the integer part of x. Later we also use the fractional part $\{x\} = x - [x]$. The basic approximation of the integrated process I_t in (I.1) is given by the left Riemann sum

$$I^{(n)}_t := \int_0^t \exp(\alpha\xi^{(n)}_s)\,\mathrm{d}s = \frac{1}{n}\sum_{k=1}^{[tn]} \exp(\alpha\xi_{(k-1)/n}) + \frac{\{tn\}}{n}\exp(\alpha\xi_{[tn]/n}).$$

In §I.7.1 below we also comment on the use of the trapezoid rule.

Note that the integrals I_t and $I^{(n)}_t$ are continuous and strictly increasing from 0 to ∞ a.s., which is an easy consequence of (I.2). Since ξ has countably many jumps, we see that $I^{(n)}$ converges to I pointwise a.s. Define the respective inverse

$$\tau_n(r) := \inf\{s > 0 \mid I^{(n)}_s \geq r\}, \qquad r \geq 0,$$

and observe that a.s. both $\tau(r)$ and $\tau_n(r)$ are finite and

$$\tau_n(r) \to \tau(r).$$

Finally, we use the approximation

$$X^{(n)}_t := x\exp(\xi^{(n)}_{\tau_n(tx^{-\alpha})}) = x\exp(\xi_{[\tau_n(tx^{-\alpha})n]/n}),$$

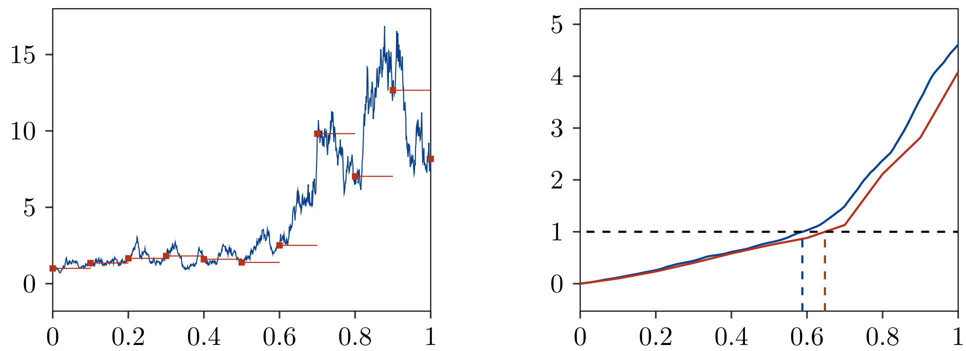
since ξ is sampled at i/n only.

Let us note that $X_t^{(n)} \to X_t$ a.s., because of the continuity of ξ at $\tau(r)$. The latter readily follows from quasi left-continuity of ξ (Bertoin, 1998, Prop. I.7) and the fact that I_t is continuous and strictly increasing. Hence the main question concerns the speed of convergence. Finally, observe that $X_t^{(n)}$ coincides with $X_{T^{(n)}}$ for some $T^{(n)} \to t$ a.s., that is, sampling is exact up to time perturbation. More precisely, such $T^{(n)}$ is given by

$$T^{(n)} = x^\alpha I_{[\tau_n(tx^{-\alpha})n]/n}, \tag{I.5}$$

so that $\tau(T^{(n)}x^{-\alpha}) = [\tau_n(tx^{-\alpha})n]/n$. The corresponding limit result is also given in the following.

Figure I.1a below illustrates the discretization of ξ in the case where $n = 10$, $\alpha = 2$ and ξ is a Brownian motion with unit variance and drift $1/2$ (corresponding to X being a standard Brownian motion conditioned to stay positive, see §I.6). In Figure I.1b we see the integrals I and $I^{(n)}$ and their inverses at $r = 1$.



(a) The processes $\exp(\alpha\xi_t)$ in blue and $\exp(\alpha\xi_t^{(n)})$ in red with $n = 10$.

(b) Integrals of the processes in Figure I.1a. The colored dashed lines mark $\tau(1)$ and $\tau_n(1)$.

Figure I.1: An illustration of the discretization and how it affects calculation of $\tau(1)$.

I.2.3 Integrated process error

Integrated discretization error for Itô semimartingales has been studied in Jacod et al. (2003), see also Jacod and Protter (2011, Ch. 6). In our case the function of interest is $f(x) = \exp(\alpha x)$. Let us first describe the limiting process defined on an extension of the original probability space:

$$\begin{aligned}\Delta_t = {}& \frac{\sigma^2}{\sqrt{12}} \int_0^t f'(\xi_s)\,\mathrm{d}W_s' \\ &+ \sum_{m:T_m \le t} (f(\xi_{T_m}) - f(\xi_{T_m-}))(\kappa_m - \tfrac{1}{2}) + \tfrac{1}{2}(f(\xi_t) - f(0)).\end{aligned} \tag{I.6}$$

Here W' is a standard Brownian motion and $(\kappa_m)_{m\ge1}$ is an i.i.d. sequence of standard uniforms, independent of each other and of $\mathcal{F}$. Furthermore, $(T_m)_{m\ge1}$ denotes a weakly exhausting sequence (Jacod and Protter, 2011, p. 100) of the jump times of ξ, and σ^2 is the variance of the Brownian component of ξ. The filtered extension is taken to be 'very good' (Jacod and Protter, 2011, p. 36) so that, in particular, Δ_t is adapted to $\widetilde{\mathcal{F}}_t$.

Theorem I.1 (Jacod and Protter (2011, Thm. 6.1.2))**.** *There is the convergence*

$$(\Delta_t^{(n)})_{t\geq 0} := n(I_{[tn]/n} - I^{(n)}_{[tn]/n})_{t\geq 0} \xrightarrow{st} (\Delta_t)_{t\geq 0}, \tag{I.7}$$

where Δ_t is defined in (I.6)*.*

The result is stated for the difference of the integral and its approximation up to the last epoch $[tn]/n$ rather than time t. In fact, there is no functional convergence in Skorokhod's J_1-topology in the latter setting unless ξ is continuous, see also Jacod et al. (2003). The problem here is that the jumps enter into the limit expression, whereas the pre-limit evolves continuously approximating these jumps by steep (almost linear) curves. Intuitively, this can be remedied by switching to Skorokhod's weaker M_1-topology, where the completed graphs of paths are compared. We do not pursue this question in the present paper, though.

I.2.4 Regularity of the Lévy process

Not surprisingly, our limit result requires certain regularity of the process ξ. Following Ivanovs (2018) we assume that there exists a positive scaling function $a_n > 0$ and a random variable $\widehat{\xi}_1 \neq 0$ such that

$$a_n \xi_{\frac{1}{n}} \xrightarrow{d} \widehat{\xi}_1 \quad \text{as } \mathbb{R} \ni n \to \infty. \tag*{[I.A1]}$$

Each $a_n \xi_{\frac{1}{n}}$ is infinitely divisible and consequently so is $\widehat{\xi}_1$. Importantly, this convergence extends to the weak convergence for processes:

$$(a_n \xi_{t/n})_{t\geq 0} \xrightarrow{d} (\widehat{\xi}_t)_{t\geq 0}. \tag{I.8}$$

Intuitively, this is understood as zooming in on ξ at the origin. The Lévy process $\widehat{\xi}$ is necessarily self-similar with index $1/\beta$ for some $\beta \in (0, 2]$, whereas a_n is regularly varying at infinity with the same index $1/\beta$, see Ivanovs (2018). This follows by a standard argument relying on the 'convergence to types' lemma. It must be noted that [I.A1] can be formulated in terms of the Lévy triplet of ξ, yielding the parameters of $\widehat{\xi}$ and the scaling function a_n (Ivanovs, 2018, Thm. 2). See also Bisewski and Ivanovs (2020) for further examples and simple sufficient conditions. Finally, it will be shown in Lemma I.7 that the convergence in (I.8) is, in fact, stable and the limiting $\widehat{\xi}$ is independent of $\mathcal{F}$. Again, to make everything a little cleaner we extend $\widehat{\xi}$ to the real line.

Our limit theory will also require convergence of $\{\tau(r)n\}$. The classic result of Kosulajeff (1937) states that such a sequence converges to a standard uniform random variable if the distribution of $\tau(r)$ is absolutely continuous, see also Tukey (1938) for sufficient and necessary conditions. Again the convergence is stable and the limiting uniform is independent of $\tau(r)$, see Jacod and Protter (2011). We impose a slightly stronger assumption:

$$\text{The law of } (\tau(r), \xi_{\tau(r)}) \text{ is absolutely continuous for every (small) } r > 0. \tag*{[I.A2]}$$

This assumption on the inverse can be replaced by an assumption on the integral I. More precisely, in §I.7.2 we show that it is sufficient to assume that the pair

$$\Big(\int_0^t \exp(\alpha\xi_s)\,\mathrm{d}s, \exp(\alpha\xi_t)\Big)$$

has a density $g_t(x, y)$ which is jointly continuous in $t, x, y > 0$. The latter question concerns the exponential functional and has been studied in a number of papers, see Salminen and Vostrikova (2018), Carmona et al. (2001) and Pardo et al. (2013). Verification of this condition, however, is still non-trivial and thus we avoid assuming [I.A2] in various places, including §I.3.2 which establishes the rate of convergence of our approximation.

I.3 Approximation results

Throughout this paper we assume (I.2). The assumptions [I.A1] and [I.A2] are needed only for some results, and this is stated at the corresponding places.

Our main aim is to establish a limit result (as $n \to \infty$) for the scaled relative error, which according to (I.1) is given by

$$a_n \frac{X_t - X_t^{(n)}}{X_t} = a_n(\xi_{\tau(r)} - \xi_{[\tau_n(r)n]/n})(1 + o_{\mathbb{P}}(1)), \qquad r = tx^{-\alpha}, \tag{I.9}$$

where we also use the mean value theorem and the fact that ξ is continuous at $\tau(r)$. The notation $o_{\mathbb{P}}(1)$ is used to denote a term which converges to zero in probability (as $n \to \infty$). The scaling sequence $a_n > 0$ will be chosen according to [I.A1] in the following. Letting

$$\widehat{\xi}^{(n)} = a_n(\xi_{\tau(r)+s/n} - \xi_{\tau(r)})_{s\in\mathbb{R}}$$

be the two-sided process arising upon zooming in on ξ at $(\tau(r), \xi_{\tau(r)})$, we find that

$$a_n(\xi_{\tau(r)} - \xi_{[\tau_n(r)n]/n}) = -\widehat{\xi}^{(n)}_{[\tau_n(r)n]-\tau(r)n}. \tag{I.10}$$

Hence we need to establish the joint limit of the two-sided process $\widehat{\xi}^{(n)}$ and the scaled time difference $\tau(r)n - [\tau_n(r)n]$, and to further extend it to the multivariate setting with $0 < t_1 < \cdots < t_d$. It will be shown that the scaled time differences $n(\tau(r_i) - \tau_n(r_i))$ are not affected by infinitesimally small time intervals, whereas the zoomed-in processes are given by the local behavior of ξ at $\tau(r_i)$ and in the limit result in independent copies of $\widehat{\xi}$.

I.3.1 Time variable and the inverse

Our first result concerns the error in the inverse $\tau_n(r)$. The limiting random variable $L(r)$, defined below, will play an important role in the following.

Proposition I.2. *For any $r > 0$ it holds that*

$$n(\tau(r) - \tau_n(r)) \xrightarrow{st} L(r) := -\Delta_{\tau(r)} \exp(-\alpha\xi_{\tau(r)}), \tag{I.11}$$

where the process Δ_t is defined in (I.6).

Proof. First, we show that

$$n(I_{\tau(r)} - I^{(n)}_{\tau(r)}) \xrightarrow{st} \Delta_{\tau(r)}. \tag{I.12}$$

Recall that ξ is continuous at $\tau(r)$ a.s., and note that the same is true for the Δ process. Hence by Theorem I.1 and the continuous mapping theorem we have the stated convergence, where $\tau(r)$ in the left-hand side of (I.12) is replaced by $t_n(r) := [\tau(r)n]/n$. It is left to show that the remaining term vanishes. Looking at this difference, we see that

$$\begin{aligned} &n(I_{\tau(r)} - I^{(n)}_{\tau(r)}) - n(I_{t_n(r)} - I^{(n)}_{t_n(r)}) \\ &\quad = n \int_{t_n(r)}^{\tau(r)} \exp(\alpha\xi_s) - \exp(\alpha\xi_{t_n(r)})\, \mathrm{d}s \\ &\quad \le \exp(\alpha \sup\{\xi_t \mid t \in [t_n(r), \tau(r)]\}) - \exp(\alpha\xi_{t_n(r)}), \end{aligned}$$

since $n(\tau(r) - t_n(r)) \le 1$. Note that the right-hand side converges to 0 a.s. by the continuity of ξ at $\tau(r)$. The lower bound is treated analogously.

Next, observe that

$$n\int_{\tau(r)}^{\tau_n(r)} \exp(\alpha\xi_t^{(n)})\,\mathrm{d}t = n(r - I_{\tau(r)}^{(n)}) = n(I_{\tau(r)} - I_{\tau(r)}^{(n)}) \xrightarrow{st} \Delta_{\tau(r)}.$$

Similar bounds to above show that the left-hand side is given by

$$n(\tau_n(r) - \tau(r))\exp(\alpha\xi_{\tau(r)})$$

times a term converging to 1 a.s. The result readily follows. □

Observe that Proposition I.2 above easily extends to a multivariate version with $0 < r_1 < \cdots < r_d$.

I.3.2 Rate of convergence

In order to proceed we need to supplement the convergence in Theorem I.1 by zooming in on ξ at the times $\tau(r_i)$.

Theorem I.3. *Assume [I.A1]. For any $0 < r_1 < \cdots < r_d$ and $r_i^n \to r_i$ there is the stable convergence*

$$\big((\Delta_t^{(n)})_{t\geq 0}, (a_n(\xi_{\tau(r_i^n)+t/n} - \xi_{\tau(r_i^n)})_{t\in\mathbb{R}})_{i=1,\dots,d}\big) \xrightarrow{st} \big((\Delta_t)_{t\geq 0}, ((\widehat{\xi}^i)_{t\in\mathbb{R}})_{i=1,\dots,d}\big),$$

where $\widehat{\xi}^i$ are independent copies of $\widehat{\xi}$, also independent of everything else.

The proof of this result is postponed to §I.4.1. We will use r_i^n dependent on n in the proof of the multivariate version of Theorem I.5. It is very important that the time t is allowed to be negative, which is a non-trivial extension of the case $t \geq 0$. This is needed, because the discretized epoch $[\tau_n(r)n]/n$ may be smaller than $\tau(r)$. Now the arguments underlying Proposition I.2 readily yield the joint stable convergence:

$$\big(n(\tau(r_i) - \tau_n(r_i)), a_n(\xi_{\tau(r_i)+t/n} - \xi_{\tau(r_i)})_{t\in\mathbb{R}}\big)_{i=1,\dots,d} \xrightarrow{st} \big(L(r_i), (\widehat{\xi}^i)_{t\in\mathbb{R}}\big)_{i=1,\dots,d}. \quad \text{(I.13)}$$

Next, we turn our attention to the pssMp and reconsider (I.9) and (I.10). Note that (I.13) readily yields the result for the error in approximation of X where $\tau_n(r)$ is used instead of $[\tau_n(r)n]/n$, but we do not observe $\xi_{\tau_n(r)}$. Our main limit theorem presented in §I.3.3 requires further work and assumptions, whereas here we establish the rate of convergence in our pssMp approximation up to a bounded stochastic term.

Consider (I.9) and the respective upper bound:

$$a_n(\xi_{\tau(r)} - \xi_{[\tau_n(r)n]/n}) \leq a_n(\xi_{\tau(r)} - \inf_{t\in[0,1]} \xi_{\tau_n(r)-t/n}) =: \overline{B}^{(n)}(r),$$

and analogous lower bound $\underline{B}^{(n)}(r)$ when using sup. According to (I.13) we have

$$\overline{B}^{(n)}(r) = -\inf_{t\in[0,1]} \widehat{\xi}^{(n)}_{-n(\tau(r)-\tau_n(r))-t} \xrightarrow{st} -\inf_{t\in[0,1]} \widehat{\xi}_{-L(r)-t},$$

because $\widehat{\xi}$ does not jump at fixed times. Since $(-\widehat{\xi}_{(-t)-})_{t\in\mathbb{R}}$ has the same law as $(\widehat{\xi}_t)_{t\in\mathbb{R}}$, we conclude that

$$\big(\overline{B}^{(n)}(r_i), \underline{B}^{(n)}(r_i)\big)_{i=1,\dots,d} \xrightarrow{st} \Big(\sup_{t\in[0,1]} \widehat{\xi}_{L(r_i)+t}, \inf_{t\in[0,1]} \widehat{\xi}_{L(r_i)+t}\Big)_{i=1,\dots,d}. \quad \text{(I.14)}$$

The following result is now immediate from (I.9). It establishes the rate of convergence a_n^{-1} and provides explicit limiting bounds.

Corollary I.4. *Assuming [I.A1], for any $x > 0$ and $0 < t_1 < \cdots < t_d$ it holds that*

$$\underline{B}^{(n)}(t_i x^{-\alpha}) + o_{\mathbb{P}}(1) \le a_n \left(\frac{X_{t_i} - X^{(n)}_{t_i}}{X_{t_i}} \right) \le \overline{B}^{(n)}(t_i x^{-\alpha}) + o_{\mathbb{P}}(1), \qquad i = 1, \ldots, d,$$

where the joint limit for the bounds is given in (I.14).

I.3.3 Discretization error in pssMp

More precise analysis requires further work and it hinges on the assumption [I.A2] implying, in particular, that $\{\tau(r)n\}$ converges to the standard uniform distribution. We have the following generalization of Theorem I.3.

Theorem I.5. *Consider $0 < r_1 < \cdots < r_d$ and assume [I.A1] and [I.A2]. Then*

$$\big((\Delta^{(n)}_t)_{t\ge 0}, (\{\tau(r_i)n\}, a_n(\xi_{\tau(r_i)+t/n} - \xi_{\tau(r_i)})_{t\in\mathbb{R}})_{i=1,\ldots,d}\big)$$
$$\xrightarrow{st} \big((\Delta_t)_{t\ge 0}, (U_i, (\widehat{\xi}^i_t)_{t\in\mathbb{R}})_{i=1,\ldots,d}\big),$$

where U_i are independent standard uniforms, also independent of the rest.

The proof of this result is given in §I.4 below. This readily yields an extension of (I.13) including the variables $\{\tau(r_i)n\}$ and their uniform limits. Thus we arrive at our main result.

Corollary I.6. *Assume [I.A1] and [I.A2]. Then for any $x > 0$ and $0 < t_1 < \cdots < t_d$ we have*

$$\left(a_n \frac{X_{t_i} - X^{(n)}_{t_i}}{X_{t_i}} \right)_{i=1,\ldots,d} \xrightarrow{st} (\widehat{\xi}^i{}_{L(t_i x^{-\alpha})+U_i})_{i=1,\ldots,d},$$

where $L(\cdot)$ is defined in (I.11). *The standard uniforms U_i and $\widehat{\xi}^i \overset{d}{=} \widehat{\xi}$ are mutually independent and independent of the rest.*

Proof. Using the identity

$$a - [b] = \{a\} - [\{a\} - (a - b)]$$

and Proposition I.2 we observe that

$$n(\tau(r_i) - [\tau_n(r_i)n]/n) = \tau(r_i)n - [\tau_n(r_i)n] \xrightarrow{st} U_i - [U_i - L(r_i)] =: L(r_i) + U_i', \qquad \text{(I.15)}$$

because $U_i - L(r_i)$ has no mass at integers and thus continuous mapping can be applied. It is easy to verify that $U_i' = \{U_i - L(r_i)\}$ are again standard uniforms independent of $L(r_i)$ and the rest (excluding the respective U_i). Furthermore, jointly with the above we also have the zooming-in limits $\widehat{\xi}^i$, and so the representations (I.9) and (I.10) yield the limit $(-\widehat{\xi}^i{}_{-L(r_i)-U_i'})_{i=1,\ldots,d}$. The result follows from the definition of $(\widehat{\xi}^i)_{t\in\mathbb{R}}$. □

It is noted that the limiting vector has dependent components and its realization depends on the realization of ξ via L. Recall that $\widehat{\xi}^i$ is $1/\beta$-self-similar, which together with (I.3) and its independence of the rest yields an alternative representation of the limit components in Corollary I.6:

$$\operatorname{sign}(L(t_i x^{-\alpha}) + U_i) \cdot |L(t_i x^{-\alpha}) + U_i|^{1/\beta} \widehat{\xi}^i_1.$$

Finally, we also have the limit result for the time shift defined in (I.5):

$$n(t - T^{(n)}) = x^\alpha n(I_{\tau(tx^{-\alpha})} - I_{[\tau_n(tx^{-\alpha})n]/n}) \xrightarrow{st} (L(tx^{-\alpha}) + U)X^\alpha_t,$$

by means of (I.15), where U is a standard uniform, independent of the rest. That is, our procedure yields the samples of X_t up to a time shift of order n^{-1}.

I.4 Proof of the joint convergence

Reconsider the definition of stable convergence in (I.4). In this paper Z_n is derived from the Lévy process ξ, and the limit Z is constructed from ξ and some additional random variables independent of $\mathcal{F}$. Thus it is sufficient to take $\sigma(\xi)$-measurable Y in (I.4) to ensure the stable convergence, see also Jacod and Protter (2011, p. 110). Furthermore, it is sufficient to show

$$(Z_n, \xi_{t_1}, \dots, \xi_{t_k}) \xrightarrow{d} (Z, \xi_{t_1}, \dots, \xi_{t_k}) \tag{I.16}$$

for an arbitrary finite set of times $t_1, \dots, t_k > 0$, which can be seen using the monotone class theorem as in Kallenberg (2002, Prop. 3.2).

I.4.1 Reinforcement of convergence results

This subsection consists of sequential reinforcement of convergence results stated in (I.8) and in Theorem I.1, and culminates with the proof of Theorem I.3.

Lemma I.7. *Assume [I.A1] and let τ_n be a sequence of finite stopping times. Then*

$$(\widehat{\xi}_t^{(n)})_{t\geq 0} := a_n(\xi_{\tau_n+t/n} - \xi_{\tau_n})_{t\geq 0} \xrightarrow{st} (\widehat{\xi}_t)_{t\geq 0},$$

where $\widehat{\xi}$ is independent of $\mathcal{F}$.

Proof. It is sufficient to consider the process $\widehat{\xi}^{(n)}$ on some time interval $[0, T]$ jointly with ξ at some times $t_1 < \dots < t_k$, see (I.16). That is, we need to show that

$$\big((\widehat{\xi}_t^{(n)})_{t\in[0,T]}, (\xi_{t_i})_{i=1,\dots,k}\big) \xrightarrow{d} \big((\widehat{\xi}_t)_{t\in[0,T]}, (\xi_{t_i})_{i=1,\dots,k}\big)$$

with an independent $\widehat{\xi}_t$. Write $\xi_{t_i} = X_i^{(n)} + Y_i^{(n)}$, where $Y_i^{(n)}$ are independent of $\widehat{\xi}_t^{(n)}, t \in [0, T]$ and $X_i^{(n)} \xrightarrow{\mathbb{P}} 0$, which can be achieved by considering independent increments over time intervals separated by τ_n, $\tau_n + T/n$ and t_i. Specifically we can use

$$X_i^{(n)} = \xi_{t_i\wedge(\tau_n+T/n)} - \xi_{t_i\wedge\tau_n} \quad \text{and} \quad Y_i^{(n)} = \xi_{t_i\wedge\tau_n} + (\xi_{t_i\vee(\tau_n+T/n)} - \xi_{\tau_n+T/n}).$$

Note how $X_i^{(n)}$ is an increment over (part of) the interval $[\tau_n, \tau_n + T/n]$ and thus is negligible in the limit, while $Y_i^{(n)}$ satisfies the required independence property.

Since we may ignore $X_i^{(n)}$, the stated convergence is immediate from independence and the weak convergence $(\widehat{\xi}_t^{(n)})_{t\in[0,T]} \xrightarrow{d} (\widehat{\xi}_t)_{t\in[0,T]}$. Note that the limit process $\widehat{\xi}$ does not jump at T a.s. and hence the latter is a consequence of (I.8). □

Lemma I.8. *Assume [I.A1] and let τ_n be a sequence of finite stopping times. It holds as $n \to \infty$ that*

$$\big((\Delta_t^{(n)})_{t\geq 0}, (\widehat{\xi}_t^{(n)})_{t\geq 0}\big) \xrightarrow{st} \big((\Delta_t)_{t\geq 0}, (\widehat{\xi}_t)_{t\geq 0}\big),$$

where $\widehat{\xi}$ is independent of everything else.

Moreover, if $0 \leq \tau_n^1 < \dots < \tau_n^d < \infty$ are stopping times for each n and such that $n(\tau_n^{i+1} - \tau_n^i) \xrightarrow{\mathbb{P}} \infty$ for all $i = 1, \dots, d-1$ then the multivariate version holds with the corresponding limits $\widehat{\xi}^i$ being independent copies of $\widehat{\xi}$, also independent of everything else.

Proof. Again we may restrict the processes $\widehat{\xi}^{(n)}$ to some time interval $[0, T]$. Let $\overline{\tau}_n$ be the discretization epoch right after $\tau_n + T/n$, and note that $\overline{\tau}_n$ is a stopping time independent of $\widehat{\xi}^{(n)}$. The idea is to replace $\Delta^{(n)}$ with the integrated difference $\tilde{\Delta}^{(n)}$, where the interval $[\tau_n, \overline{\tau}_n]$ and the respective space increment are ignored. More precisely, the new ξ is kept constant on

$[\tau_n, \overline{\tau}_n]$ and then it has the original increments. Observe that $\sup_{t\le T'}|\tilde{\Delta}_t^{(n)} - \Delta_t^{(n)}| = o_{\mathbb{P}}(1)$ using the strong Markov property at τ_n, see also the proof of Proposition I.2. But now the two parts are independent and the arguments from Lemma I.7 can be repeated, additionally using Theorem I.1 for the joint convergence of $\Delta^{(n)}$ and ξ_{t_i}.

The multivariate version follows the same reasoning. Note, that the intervals $[\tau_n^i, \tau_n^i + T/n]$ do not intersect with probability tending to 1 by assumption. Hence we may assume this property which then yields independent $\widehat{\xi}^i$. □

Proof of Theorem I.3. We use Lemma I.8 with $\tau_n^i = \tau(r_i^n - T/n)$ for a fixed $T > 0$. Note that

$$n(\tau(r_i^n) - \tau(r_i^n - T/n)) \to T\exp(-\alpha\xi_{\tau(r_i)}) =: s_i \quad \text{a.s.},$$

see also the proof of Proposition I.2. Thus we can add the required time shifts to the limit result, and these limiting shifts s_i are independent of the processes $(\widehat{\xi}_t^i)_{t\ge 0}$. But for any $T' > 0$ we can choose T large enough so that with arbitrarily large probability $s_i > T'$, and on this event $(\widehat{\xi}_{s_i+t}^i - \widehat{\xi}_{s_i}^i)_{t\ge -T'}$ has the law of $(\widehat{\xi}_t)_{t\ge -T'}$ and is independent of s_i. It is left to apply the continuous mapping theorem. □

In conclusion, the stopping time $\tau(r)$ has a particular structure allowing to extend zooming in at $\tau(r)$ also to the negative times.

I.4.2 Fractional parts and the standard uniform

Here we prove the joint convergence in Theorem I.5 for $d = 1$. For the purpose of extending it from $d = 1$ to $d \ge 1$ later we need to allow for perturbations in r. We state this result as a separate lemma.

Lemma I.9. *Assuming [I.A1] and [I.A2] we have for any $r_n \to r > 0$:*

$$\big(\Delta^{(n)}, a_n(\xi_{\tau(r_n)+t/n} - \xi_{\tau(r_n)})_{t\in\mathbb{R}}, \{\tau(r_n)n\}\big) \xrightarrow{st} \big(\Delta, (\widehat{\xi}_t)_{t\in\mathbb{R}}, U\big),$$

where U is a standard uniform independent of everything else.

The independent uniform will arise via the following lemma. Consider a random variable Z and a sequence of random variables (U_n, V_n, Y_n) defined on the same probability space. Let $\mathbb{P}_z$ be the regular conditional distribution $\mathbb{P}(\,\cdot \mid Z = z)$, which is unique almost everywhere with respect to the law $\mathbb{P}_Z$ of Z (Kallenberg, 2002, Thm. 6.3).

Lemma I.10. *Assume that $Y_n \xrightarrow{d} Y$, and that for $\mathbb{P}_Z$-almost all z we have under $\mathbb{P}_z$:*

- *U_n is independent of (V_n, Y_n) for each n.*
- *The distribution of U_n has no atoms and converges weakly to the standard uniform distribution.*

Then $(Y_n, \{U_n + V_n\}) \xrightarrow{d} (Y, U)$ with a standard uniform U independent of Y.

Proof. Below we work with $\mathbb{P}_Z$-almost all z. Let $F_{n,z}$ be the continuous distribution function of $U_n \mid Z = z$. Define

$$U_n' = F_{n,Z}(U_n)$$

and note that, given $Z = z$, U_n' is a standard uniform independent of (V_n, Y_n). Note that

$$\mathbb{P}(Y_n \in B, \{V_n + U_n'\} \le u) = u\mathbb{P}(Y_n \in B)$$

by conditioning on Z, Y_n, V_n and the fact that $\{v + U'_n\}$ is standard uniform. Hence

$$(Y_n, \{V_n + U'_n\}) \xrightarrow{d} (Y, U)$$

with Y and U independent.

It is left to show that

$$\{V_n + U'_n\} - \{V_n + U_n\} \xrightarrow{\mathbb{P}} 0. \tag{I.17}$$

Since $F_{n,z}(x) - x \to 0$ and the convergence is necessarily uniform in x, we see that $U'_n - U_n \xrightarrow{\mathbb{P}} 0$. Hence for any small $\delta > 0$ we have $|U'_n - U_n| < \delta$ with probability at least $1 - \delta$ for large n. Moreover, $\mathbb{P}(\{V_n + U'_n\} < \delta) = \mathbb{P}(\{V_n + U'_n\} > 1 - \delta) = \delta$ and thus the left-hand side of (I.17) does not exceed δ in absolute value with probability at least $1 - 3\delta$. ☐

Proof of Lemma I.9. Choose $0 < \delta < r$ and $\delta' > 0$, and consider the process $\xi'_t = \xi_{\tau(r-\delta)+t} - \xi_{\tau(r-\delta)}$ independent of $\mathcal{F}_{\tau(r-\delta)}$. Let τ be the time such that

$$\int_0^\tau \exp(\alpha \xi'_t)\, \mathrm{d}t = \delta'.$$

We consider the regular conditional distribution $\mathbb{P}_z$ corresponding to conditioning on $\xi'_\tau = z$. Note that for almost every z the variable τ has a density under $\mathbb{P}_z$ according to the assumption [I.A2], and so the distribution of $U_n = \{\tau n\}$ has no atoms and it converges weakly to a standard uniform law. Furthermore,

$$(I_{\tau(r-\delta)+\tau}, \xi_{\tau(r-\delta)+\tau}) = (r - \delta + \exp(\alpha \xi_{\tau(r-\delta)})\delta', \xi_{\tau(r-\delta)} + z),$$

and we assume that the first component is smaller than $r_n \wedge (r - \delta/2)$; we may do so since this is true for small enough δ' with arbitrarily high probability. Now letting $R_n = \tau(r_n) - (\tau(r - \delta) + \tau)$ be the remaining time, we note the decomposition of the fractional part of interest:

$$\{\tau(r_n)n\} = \{(\tau(r - \delta) + R_n)n + \{\tau n\}\} =: \{V_n + U_n\},$$

where U_n is independent of V_n under $\mathbb{P}_z$.

Next, we define the quantities of interest, which will be assembled into Y_n. The integrated difference process stopped at $\tau(r - \delta)$ is denoted by $\hat{\Delta}^{(n)}$. We consider this quantity jointly with $\xi_{t_i} \mathbf{1}_{\{t_i < \tau(r-\delta)\}}$ for some fixed times $t_i, i \le k$. Furthermore, consider the epoch τ_n following $\tau(r + \delta)$ with the corresponding incremental process $\tilde{\xi}_t = \xi_{\tau_n + t} - \xi_{\tau_n}$, which is independent of $\mathcal{F}_{\tau(r+\delta)}$. The integrated difference process for the times $\tau_n + t, t \ge 0$ is given by $\exp(\alpha \xi_{\tau(r+\delta)})(1 + o_\mathbb{P}(1))\tilde{\Delta}^{(n)}$, which is our second object of interest. It is considered jointly with $\xi_{\tau(r+\delta)+\tilde{t}_i} = \xi_{\tau(r+\delta)} + \tilde{\xi}_{\tilde{t}_i} + o_\mathbb{P}(1)$ for some fixed $\tilde{t}_i, i \le \tilde{k}$. Thirdly, we consider the zoomed-in process $\widehat{\xi}^{(n)}_t = a_n(\xi_{\tau(r_n)+t/n} - \xi_{\tau(r_n)})$ for $t \in [-T, T]$. The event where $\tau(r + \delta) > \tau(r_n) + T/n$ and $\tau(r - \delta/2) < \tau(r_n) - T/n$ occurs with arbitrarily high probability, and we assume these inequalities in the following. The above objects form the random quantity

$$Y_n = \big(\hat{\Delta}^{(n)}, (\xi_{t_i} \mathbf{1}_{\{t_i < \tau(r-\delta)\}})_{i=1,\ldots,k}, \xi_{\tau(r+\delta)}, \tilde{\Delta}^{(n)}, (\tilde{\xi}_{\tilde{t}_i})_{i=1,\ldots,\tilde{k}}, \widehat{\xi}^{(n)}\big),$$

and as required in Lemma I.10 the variable U_n is independent of (V_n, Y_n) and the above events under $\mathbb{P}_z$.

Observe that the quantities $\tilde{\Delta}^{(n)}, \tilde{\xi}_{\tilde{t}_i}$ are independent of the rest and have a joint weak limit as given by Theorem I.1. But the rest converges according to Theorem I.3, where stopping at $\tau(r - \delta)$ requires that ξ does not jump at this time, which is indeed true. Thus

Lemma I.10 yields $(Y_n, \{\tau(r)n\}) \xrightarrow{d} (Y, U)$ with an independent standard uniform U and obvious Y for any given $\delta > 0$.

Finally, we piece together different components to get the integrated difference processes with the time interval $(\tau(r-\delta), \tau(r+\delta))$ excluded, as well as the corresponding limiting expression, see also (I.6). Now we can take $\delta \downarrow 0$ using Jacod and Protter (2011, Prop. 2.2.4) to get

$$\begin{aligned}&\big(\Delta^{(n)}, (\xi_{t_i}\mathbf{1}_{\{t_i<\tau(r)\}})_{i=1,\dots,k}, (\xi_{\tau(r)+\tilde{t}_i})_{i=1,\dots,k}, \widehat{\xi}^{(n)}, \{\tau(r_n n)\}\big)\\ &\qquad\xrightarrow{d} \big(\Delta, (\xi_{t_i}\mathbf{1}_{\{t_i<\tau(r)\}})_{i=1,\dots,k}, (\xi_{\tau(r)+\tilde{t}_i})_{i=1,\dots,k}, \widehat{\xi}, U\big).\end{aligned}$$

It is only required to verify the assumptions of Jacod and Protter (2011, Prop. 2.2.4). Firstly, the limits converge a.s. as $\delta \downarrow 0$, because $\tau(r \pm \delta) \to \tau(r)$ and the process ξ is continuous at $\tau(r)$ and at $\tau(r) + \tilde{t}_i$. Secondly, we must show that the excluded integrated difference is uniformly negligible:

$$\lim_{\delta\downarrow 0}\limsup_{n\to\infty}\mathbb{P}\Big(\sup_{t\le\lceil\tau(r+\delta)n\rceil/n}\Big|n\int_{[\tau(r-\delta)n]/n}^{[tn]/n}(\exp(\alpha\xi_s)-\exp(\alpha\xi_s^{(n)}))\,\mathrm{d}s\Big|\ge\epsilon\Big)=0.$$

But the respective quantity converges weakly according to Theorem I.1, and the limit goes to 0 a.s. establishing this claim. The proof is now complete. □

I.4.3 Extension to multivariate case

Let us recall a basic result, which readily follows from Skorokhod's representation theorem. Assume that μ_n is a sequence of finite measures converging weakly to a finite measure μ and that f_n is a sequences of bounded functions that are continuously convergent, i.e. $f_n(z_n) \to f(z)$ whenever $z_n \to z$ for z in the support of μ. Then we also have

$$\int f_n\,\mathrm{d}\mu_n \to \int f\,\mathrm{d}\mu.$$

Proof of Theorem I.5. We prove the multivariate case inductively. Suppose the case $d \ge 1$ is proven. Consider $r = (r_d + r_{d+1})/2$, and let $\tau_n = \lceil\tau(r)n\rceil/n$ be the epoch following $\tau(r)$ which is a stopping time. Note also that $\tau_n \to \tau(r)$ and $r^{(n)} = I_{\tau_n} \to r$ a.s. We condition on $\xi_{\tau_n} = x$ and $r^{(n)} - r = \epsilon$ and use the strong Markov property to split the quantities of interest. The processes $\Delta^{(n)}$ are split into two parts: the one stopped at τ_n and the post-τ_n contribution. The latter corresponds to $\exp(\alpha x)\tilde{\Delta}^{(n)}$ for the process $\tilde{\xi}_t = \xi_{\tau_n+t} - \xi_{\tau_n}$ which is independent of $\mathcal{F}_{\tau_n}$. Moreover, note that

$$\tau(r_{d+1}) = \tau_n + \tilde{\tau}(\exp(-\alpha x)(r_{d+1} - r - \epsilon))$$

and zooming in at $\tau(r_{d+1})$ translates into zooming in on $\tilde{\xi}$ at the respective time, whereas

$$\{\tau(r_{d+1})n\} = \{\tilde{\tau}(\exp(-\alpha x)(r_{d+1} - r - \epsilon))n\}.$$

Finally, we may assume that none of the zoomed-in processes over $[-T, T]$ span both $[0, \tau_n]$ and (τ_n, ∞) since this is true with arbitrarily large probability for large enough n. This allows to split the variables of interest into two independent groups under the conditional law specified above.

Next, we construct the measures $\mu_n(\mathrm{d}x, \mathrm{d}\epsilon)$ and the functions f_n by simply applying bounded continuous functions to the two quantities of interest, where the latter also include

$\tilde{\xi}_{\tilde{t}_i}$ needed to guarantee the stable convergence. Weak convergence of measures follows from the inductive assumption and the facts that $\tau_n \to \tau(r), r^{(n)} \to r$ and ξ is continuous at $\tau(r)$. Convergence of $f_n(x_n, \epsilon_n)$ for $(x_n, \epsilon_n) \to (x, 0)$ follows from Lemma I.9 with $r_n \to r$ given by

$$\exp(-\alpha x_n)(r_{d+1} - r - \epsilon_n) \to \exp(-\alpha x)(r_{d+1} - r).$$

It is left to glue back the limits, where the only dependence comes from x needed to reconstruct the process $(\Delta_t)_{\geq 0}$ and the variables $\xi_{\tau(r)+\tilde{t}_i}$. Finally, note that convergence still holds when $\xi_{\tau_n+\tilde{t}_i}$ are replaced by $\xi_{\tau(r)+\tilde{t}_i}$ in the pre-limit. This yields the stated stable convergence for $d+1$, and the proof is complete. □

I.5 Zooming-in on pssMp

I.5.1 The result

Self-similarity of X implies that $n^{1/\alpha} X_{t/n}$ (with X starting at x) has the law of X_t (starting at $xn^{1/\alpha}$). There is, however, a different scaling resulting in a limit process as $n \to \infty$, which we now state. Importantly, it provides a zooming-in limit for the pssMp X and connects it to the zooming-in limit for ξ. It is noted that this result does not require the assumption (I.2). Furthermore, this result is somewhat related to the law of iterated logarithm for X_t at small times, see Lamperti (1972, Thm. 7.1) and Pardo and Rivero (2013, §2.3).

Theorem I.11. *Under the assumption [I.A1] there is the convergence for any* $x > 0$

$$a_n(X_{t/n} - x)_{t\geq 0} \xrightarrow{st} x^{1-\alpha/\beta}(\widehat{\xi}_t)_{t\geq 0} \quad \text{as } \mathbb{R} \ni n \to \infty, \tag{I.18}$$

where $\widehat{\xi}$ *is independent of* $\mathcal{F}$ *and* $1/\beta$ *is its Hurst index.*

Furthermore, [I.A1] is equivalent to the weak convergence of $a_n(X_{1/n} - 1)$ *to a non-zero limit for* $x = 1$.

Proof. For all $t \in [0, T]$ we have

$$a_n(X_{t/n} - x) = x a_n \xi_{\tau(x^{-\alpha}t/n)}(1 + R_{t,n}),$$

where $\sup_{t\leq T}|R_{t,n}| \xrightarrow{\mathbb{P}} 0$. It is left to show that

$$\sup_{t\leq T}|\tau(x^{-\alpha}t/n)n - x^{-\alpha}t| \xrightarrow{\mathbb{P}} 0, \tag{I.19}$$

since then by continuity of subordination (Whitt, 2002, Thm. 13.2.2) at the limiting time change $x^{-\alpha}t$, and $a_n\xi_{\cdot/n} \xrightarrow{st} \widehat{\xi}$ we find

$$(a_n\xi_{\tau(x^{-\alpha}t/n)})_{t\leq T} \xrightarrow{st} (\widehat{\xi}_{x^{-\alpha}t})_{t\leq T} \overset{d}{=} x^{-\alpha/\beta}(\widehat{\xi}_t)_{t\leq T}.$$

To this end, observe that

$$t/n - \tau(t/n) = \int_0^{\tau(t/n)} (e^{\alpha\xi_s} - 1)\,\mathrm{d}s = \tau(t/n)\mathrm{o}_{\mathbb{P}}(1),$$

which firstly shows that $n\tau(x^{-\alpha}T/n) \xrightarrow{\mathbb{P}} x^{-\alpha}T$ and then also yields (I.19).

Next, assume that $a_n(X_{1/n} - 1) \xrightarrow{d} Z \neq 0$ for $x = 1$. Then

$$a_n\xi_{\tau(1/n)} = a_n(e^{\xi_{\tau(1/n)}} - 1)\big(1 + \mathrm{o}_{\mathbb{P}}(1)\big) \xrightarrow{d} Z.$$

But $\tau(1/n)n \xrightarrow{\mathbb{P}} 1$ and according to Proposition I.12 below we must have

$$a_n \xi_{1/n} \xrightarrow{d} Z$$

The proof is now complete. □

Let us note that the non-zero weak limit of $a_n(X_{1/n} - 1)$, when it exists, is necessarily $\widehat{\xi}_1$. In fact, this assumption is equivalent to a seemingly weaker assumption, namely that $a_n(X_{t/n} - x) \xrightarrow{d} Z \neq 0$ for some $t, x > 0$. Importantly, Theorem I.11 allows to identify $\widehat{\xi}$ directly without determining the corresponding process ξ first. An application of this will be given in §I.6 below.

I.5.2 On convergence of Lévy processes at random times

The following basic result is essential for the second statement in Theorem I.11, and it may be useful in various other settings. Somewhat surprisingly, it is not contained in the standard monographs.

Proposition I.12. *Consider a sequence of Lévy processes ξ^n and assume that $\xi^n_{T_n} \xrightarrow{d} Z$ for some random times $0 \leq T_n \xrightarrow{\mathbb{P}} 1$. Then $\xi^n_1 \xrightarrow{d} Z$.*

Importantly, we do not assume that ξ^n and T_n are independent. The main difficulty is in proving that ξ^n_1 is tight, which is the content of the following two lemmas.

Lemma I.13. *Assume that $0 \leq T_n \xrightarrow{\mathbb{P}} 1$ and $\xi^n_{T_n} \xrightarrow{\mathbb{P}} 0$ for a sequence of Lévy processes ξ^n such that*

$$\mathbb{P}(\sup_{t \leq 1} |\xi^n_t| > 1) \leq 1/2. \tag{I.20}$$

Then $\xi^n_1 \xrightarrow{\mathbb{P}} 0$.

Proof. Suppose for contradiction that there exist $\epsilon, \delta > 0$ such that

$$\liminf_{n \to \infty} \mathbb{P}(\xi^n_{1-\delta} < -2\epsilon) > 0.$$

Let $\xi'^n_t = \xi^n_{1-\delta+t} - \xi^n_{1-\delta}$ be the incremental post-$(1-\delta)$ process. Using the typical notation $\overline{\xi}'^n_s$ for $\sup_{u \in [0,s]} \xi'^n_u$ we have

$$\mathbb{P}(\xi^n_{T_n} > -\epsilon, |T_n - 1| < \delta, \xi^n_{1-\delta} < -2\epsilon) \leq \mathbb{P}(\xi^n_{1-\delta} < -2\epsilon)\mathbb{P}(\overline{\xi}'^n_{2\delta} > \epsilon),$$

where on the right-hand side we used independence of ξ'^n and $\xi^n_{1-\delta}$. By the initial assumption we readily obtain

$$p_n := \mathbb{P}(\overline{\xi}^n_{2\delta} > \epsilon) \to 1 \quad \text{as } n \to \infty.$$

Applying the strong Markov property at first passage times we now find

$$1/2 \geq \mathbb{P}(\overline{\xi}^n_1 > 1) \geq p_n^{\lceil 1/\epsilon \rceil}$$

for all n, given that $2\delta\lceil 1/\epsilon \rceil < 1$. In this case the right-hand side tends to 1, which is a contradiction. Similar reasoning works when $\mathbb{P}(\xi^n_{1-\delta} > 2\epsilon)$ is assumed to be bounded away from 0. Thus we conclude that for any $\epsilon > 0$ and small enough $\delta > 0$ we have

$$\mathbb{P}(|\xi^n_{1-\delta}| > \epsilon) \to 0 \quad \text{as } n \to \infty.$$

Fix arbitrary $h, \epsilon > 0$ and choose δ small so that $\mathbb{P}(|\xi^n_{1-2\delta}| < \epsilon)$ and $\mathbb{P}(|\xi^n_{1-\delta}| < \epsilon)$ are larger than $1 - h$ for all large n. Thus $\mathbb{P}(|\xi^n_\delta| < 2\epsilon) > 1 - 2h$ implying that $\mathbb{P}(|\xi^n_1| < 3\epsilon) > 1 - 3h$, which completes the proof. □

Lemma I.14. *The conclusion of Lemma I.13 is true without the assumption* (I.20).

Proof. We choose the maximal b_n such that (I.20) is satisfied for $\xi_t'^n = b_n \xi_t^n$:

$$b_n = \sup\{b \in (0,1] : \mathbb{P}(\sup_{t\le 1} |b\xi_t^n| > 1) \le 1/2\}.$$

Since b_n is upper bounded by construction, we still have $\xi_{T_n}'^n \xrightarrow{\mathbb{P}} 0$. Now the previous lemma implies that $b_n\xi_1^n \xrightarrow{\mathbb{P}} 0$, and then according to the standard theory (Kallenberg, 2002, Thm. 15.17) we also have convergence on the process level. By the continuous mapping theorem we find

$$b_n \sup_{t\le 1} |\xi_t^n| \xrightarrow{\mathbb{P}} 0,$$

whereas by maximality of b_n it must be that $\mathbb{P}(\sup_{t\le 1}|2b_n\xi_t^n| > 1) > 1/2$ for any $b_n < 1$. Hence $b_n = 1$ for all large n and the proof is complete. □

Proof of Proposition I.12. Take any sequence $0 \le h_n \to 0$ and note that $h_n\xi_{T_n}^n \xrightarrow{\mathbb{P}} 0$. By Lemma I.14 we also have $h_n\xi_1^n \xrightarrow{\mathbb{P}} 0$. According to Kallenberg (2002, Lem. 4.9) the sequence ξ_1^n is tight. Thus every subsequence has a weakly convergent further subsequence $\xi_1^{n_k}$ (Kallenberg, 2002, Prop. 5.21). It must be (Kallenberg, 2002, Thm. 15.12) that the limit is Z_1' for some Lévy process Z', and $\xi^{n_k} \xrightarrow{d} Z'$, see Kallenberg (2002, Thm. 15.17). But Z' is necessarily continuous at time 1 a.s., and thus $\xi_{T_{n_k}}^{n_k} \xrightarrow{d} Z_1'$ showing that Z_1' and Z have the same distribution. Thus $\xi_1^{n_k} \xrightarrow{d} Z$ and the proof is now complete. □

I.6 Application to self-similar Lévy processes conditioned to stay positive

I.6.1 Definition and properties

Let $(X_t^0)_{t\ge 0}$ be a non-monotone $1/\alpha$-self-similar Lévy process. In particular, X^0 is either (I) a drift-less Brownian motion ($\alpha = 2$) or (II) a strictly α-stable Lévy process with $\alpha \in (0,2)$. Without real loss of generality we may fix the scale, and so in case (I) we assume that X_0 is a standard Brownian motion. We also define the negativity parameter

$$\rho = \mathbb{P}(X_1^0 < 0) \in (0,1),$$

which additionally must satisfy $\alpha - 1 \le \alpha\rho \le 1$ and, in particular, $\rho = 1/2$ in case (I).

Let X be the process X^0 conditioned to stay positive when started from $x > 0$. Formally it is defined via Doob's h-transform (Caballero and Chaumont, 2006):

$$\mathbb{P}_x^{\uparrow}(A) := h^{-1}(x)\mathbb{E}[h(x+X_t^0)\,\mathbf{1}_A\,\mathbf{1}_{\{x+\underline{X}_t^0>0\}}], \qquad t\ge 0, A\in\mathcal{F}_t, \tag{I.21}$$

where $h(x) = x^{\alpha\rho}$ and $\underline{X}_t^0 = \inf_{s\le t} X_s^0$, see also Bertoin (1993) for the case when X^0 is a general Lévy process. We write $(X,\mathbb{P})$ for the pair $(x+X^0, \mathbb{P}_x^{\uparrow})$ and specify $x > 0$ separately when needed. Let us also mention that the new law in (I.21) coincides with the limit of $\mathbb{P}(A \,|\, x+\underline{X}_s^0 > 0)$ as $s\to\infty$, explaining the name 'conditioned to stay positive'. Importantly, X is a strictly positive pssMp with Hurst parameter $1/\alpha$. Such processes naturally arise in limit theory concerned with extremes, first passage times and Skorokhod reflection, see Ivanovs (2018) and Ivanovs and Podolskij (2020) and references therein. Importantly, in case (I) the process X is Bessel-3.

As before, let ξ be the Lévy process in the Lamperti representation (I.1) of X. In case (I) the process ξ is a Brownian motion with unit variance and drift 1/2, see Carmona et al.

(2001). In case (II) the Lévy triplet of ξ has been identified in Caballero and Chaumont (2006)[1], excluding the non-symmetric Cauchy case. It is worth mentioning that ξ has no Brownian component, and its Lévy density behaves as the Lévy density of the original stable process X^0 both at $0+$ and at $0-$. Furthermore, ξ is a pure jump process when $\alpha \in (0,1)$.

Importantly, Theorem I.11 allows to identify $\widehat{\xi}$ and to verify assumption [I.A1] without the knowledge of ξ. It turns out that $\widehat{\xi}$ has the law of the original process X^0 and, in particular, $\beta = \alpha$.

Proposition I.15. *Let X be X^0 conditioned to stay positive. Then the assumption [I.A1] is satisfied with*

$$a_n = n^{1/\alpha} \quad \textit{and} \quad \widehat{\xi} \overset{d}{=} X^0.$$

Proof. According to Theorem I.11 we only need to verify that

$$n^{1/\alpha}(X_{1/n} - 1) \xrightarrow{d} X_1^0$$

for $x = 1$ as $\mathbb{R} \ni n \to \infty$. Using (I.21) and self-similarity of X^0 one easily verifies that

$$\mathbb{P}(n^{1/\alpha}(X_{1/n} - 1) \leq z) = \mathbb{E}[(n^{-1/\alpha} X_1^0 + 1)^{\alpha\rho} \mathbf{1}_{X_1^0 \leq z} \mathbf{1}_{n^{-1/\alpha} \underline{X}_1^0 > -1}],$$

for any $z \in \mathbb{R}$. But the right-hand side converges to $\mathbb{P}(X_1^0 \leq z)$, and we are done. □

Alternatively, one may prove Proposition I.15 using the knowledge of the Lévy triplet of ξ by checking the conditions of Ivanovs (2018, Thm. 2). The latter approach requires verification that the drift parameter of ξ is zero in case $\alpha < 1$. Furthermore, calculations are somewhat tedious in the symmetric Cauchy case, whereas the triplet of ξ in the non-symmetric case is not yet available.

In various applications the law of interest corresponds to the weak limit of $\mathbb{P}_x^{\uparrow}$ as $x \downarrow 0$, which corresponds to the conditioned process started at 0. This can be approximated by taking small $x > 0$, which then results in large $r = x^{-\alpha} t$. Thus it would be interesting to understand the behavior of $L(r)$ as $r \to \infty$.

I.6.2 Simulations

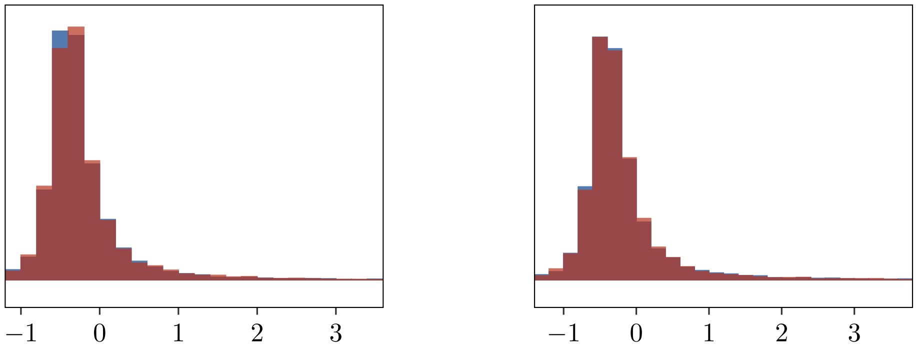
Here we present a small simulation study in order to illustrate our results. For simplicity, we take a standard Brownian motion conditioned to stay positive (Bessel-3 process) as the pssMp X of interest. Let us stress that simple and exact simulation methods exist for Bessel-3 process, and our only purpose is to illustrate the results of §I.3. In this case $\alpha = 2, a_n = \sqrt{n}$, ξ is a Brownian motion with unit variance and drift $1/2$, whereas $\widehat{\xi}$ is a standard Brownian motion, see Proposition I.15. We also note that assumption [I.A2] is satisfied since the density $g_t(x, y)$ in Lemma I.17 is indeed jointly continuous in $t, x, y > 0$, see Borodin and Salminen (2002, 1.8.8, p. 613).

We start X at $x = 1$ and simulate at time $t = 1$. Hence $X_1 = \exp(\xi_{\tau(1)})$. We use two rather coarse discretization grids corresponding to $n = 10$ and $n = 100$. The true quantities are computed using $N = 10^6$, so that $\xi^{(N)}$ and $X^{(N)}$ are used in place of ξ and X, respectively. The process Δ in (I.6) is approximated by taking $\xi^{(N)}$ in the Brownian integral and removing the sum over jump times, which must be 0 in the case of continuous ξ. Finally, $\tau(1)$ is replaced by $\tau_N(1)$. Importantly, the increments of $\xi^{(N)}$ are assembled into

[1] Caballero and Chaumont (2006, Eq. (17)) has a typo: the second term should come with a minus sign, which only affects the drift parameter.

the increments of $\xi^{(n)}$, so that the two processes correspond to the same sample path. These sample paths are then reused in construction of the limit variables.

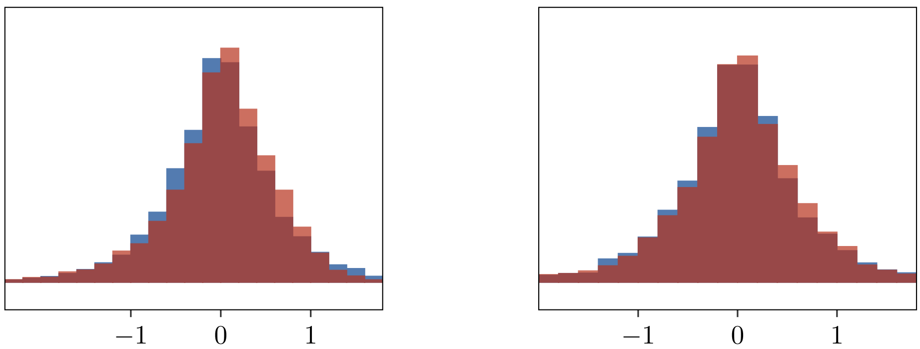
In Figure I.2 below we compare the distributions of $n(\tau(1) - \tau_n(1))$ and the limit $L(1)$, see Proposition I.2. All histograms are based on simulation of 10 000 independent copies of the relevant random variable. In red we have $n(\tau(1) - \tau_n(1))$ and in blue we have $L(1)$. Since some values are quite large the histogram has been trimmed to contain at least 98% of the realizations. More precisely the lower limit is the minimum of the 1%-quantiles for $n(\tau(1) - \tau_n(1))$ and $L(1)$, and the upper limit is the maximum of the 99%-quantiles. We discuss these large values in detail later. Let us remark that already at $n = 10$ we see very similar histograms and at $n = 100$ the fit is even better.



(a) $n = 10$ **(b)** $n = 100$

Figure I.2: Histograms for $n(\tau(1) - \tau_n(1))$ in red and $L(1)$ in blue trimmed to contain at least 98% of the realizations.

In Figure I.3 we depict the discretization errors for the pssMp itself. That is, we compare the distributions of $\sqrt{n}(X_1 - X_1^{(n)})/X_1$ in red and the limit $\widehat{\xi}_{L(1)+U}$ in blue, see Corollary I.6. The fit is worse than in Figure I.2, which is to be expected since now we combine the error in time and the zooming-in approximation. Again we have trimmed the histograms to contain at least 98% of the realizations.

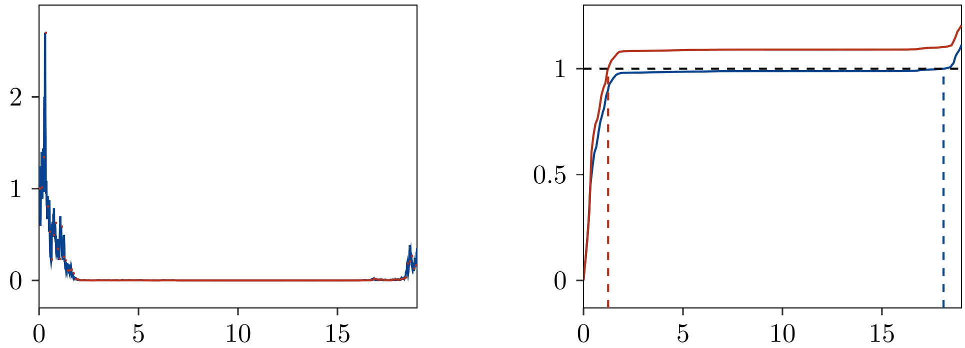


(a) $n = 10$ **(b)** $n = 100$

Figure I.3: Histograms for $\sqrt{n}(X_1 - X_1^{(n)})/X_1$ in red and $\widehat{\xi}_{L(1)+U}$ in blue trimmed to contain at least 98% of the realizations.

In order to understand the extreme values of $L(1)$ and $n(\tau(1) - \tau_n(1))$ we depict a sample path in Figure I.4a which results in large values of both variables. In this case $n = 10$. Notice that $I^{(n)}$ hits 1 right before $\exp(\alpha\xi_t)$ vanishes ($-\xi_t$ becomes large), whereas I hits 1 upon much later. This illustrates how $n(\tau(1) - \tau_n(1))$ can become very large. Furthermore,

$\exp(\alpha\xi_{\tau(1)}) = X_1^\alpha$ is close to zero and so $L(1) = -\Delta_{\tau(1)}X_1^{-\alpha}$ can be large as well. It seems that heaviness of the tails of $L(1)$ is determined by $X_1^{-\alpha}$; for the Bessel-3 process this quantity has a power tail with exponent $-3/2$, see Borodin and Salminen (2002, 1.0.6, p. 373).



(a) The processes $\exp(\alpha\xi_t)$ and $\exp(\alpha\xi_t^{(n)})$ in respectively blue and red with $n = 10$.

(b) Integrals of the processes in Figure I.4a. The colored dashed lines mark $\tau(1)$ and $\tau_n(1)$.

Figure I.4: A sample path and corresponding integrals producing extreme values of $L(1)$ and $n(\tau(1) - \tau_n(1))$.

In conclusion, discretization provides the standard rate of convergence $n^{-1/\alpha}$, but the limit variables normally exhibit heavy-tails.

I.7 Extensions and comments

I.7.1 Trapezoidal approximation

An interesting modification of our approximation scheme is obtained by considering the trapezoidal rule (instead of the left Riemann sum) in computation of the integral I_t, so that the points $(i/n, \exp(\alpha\xi_{i/n}))$ are connected by straight lines. Importantly, all the results and proofs of this paper continue to hold true given that Theorem I.1 and the definition of Δ in (I.6) are adjusted accordingly, which we now discuss.

Observe that the trapezoidal approximation $\widetilde{I}^{(n)}$ satisfies

$$\widetilde{I}_{i/n}^{(n)} = I_{i/n}^{(n)} + (f(\xi_{i/n}) - f(0))/(2n)$$

and hence the form of the new limiting process is intuitively clear:

$$\widetilde{\Delta}_t = \frac{\sigma^2}{\sqrt{12}} \int_0^t f'(\xi_s)\,\mathrm{d}W_s' + \sum_{m:T_m \le t} (f(\xi_{T_m}) - f(\xi_{T_m-}))(\kappa_m - \frac{1}{2}),$$

that is, the bias $(f(\xi_t) - f(0))/2$ is removed from (I.6).

Theorem I.16. *The trapezoidal approximation $\widetilde{I}^{(n)}$ satisfies*

$$n(I_{[tn]/n} - \widetilde{I}_{[tn]/n}^{(n)})_{t\ge 0} \xrightarrow{st} (\widetilde{\Delta}_t)_{t\ge 0}.$$

Proof. The proof requires only some simple adaptations of the proof in Jacod and Protter (2011, Ch. 6). Firstly, we note that the reduction of the problem in §6.2.2 is still true, because of the u.c.p. convergence of processes in (6.2.13). Secondly, (6.3.6) now contains our new term, which is rewritten using Itô's formula, and the limiting expression in (6.3.7) is modified accordingly. The expressions, in fact, become even shorter, and the rest of the proof applies. □

It must be noted that this result cannot be directly retrieved from Jacod and Protter (2011, Thm. 6.1.2) and the basic relation between left Riemann sum and trapezoidal rule. The problem is that the continuous mapping theorem does not apply for the sum of the two processes of interest since both components may jump at the same time. This issue does not arise in the setting of continuous Itô semimartingales considered in Altmeyer (2019).

I.7.2 Absolute continuity of the inverse

Here we establish a sufficient condition for [I.A2] in terms of the integral I_t and the end-value $\exp(\alpha\xi_t)$. We assume that the pair

$$(I_t, Y_t) = \Big(\int_0^t \exp(\alpha\xi_s)\,\mathrm{d}s, \exp(\alpha\xi_t)\Big)$$

has a density $g_t(x, y)$ for all $t > 0$. Recall that $\tau(r)$ is defined by the relation $I_{\tau(r)} = r$, and that we need simple conditions implying that the pair $(\tau(r), Y_{\tau(r)})$ has a density for all $r > 0$.

Lemma I.17. *Assume that $g_t(x, y)$ is jointly continuous in $x, y, t > 0$. Then for any $r > 0$*

$$\mathbb{P}(\tau(r) \in \mathrm{d}t, Y_{\tau(r)} \in \mathrm{d}y) = y g_t(r, y)\,\mathrm{d}t\,\mathrm{d}y, \qquad t, y > 0.$$

Proof. For fixed $r > 0$ and $0 < a < b < \infty$ consider

$$F(t) = \mathbb{P}(\tau(r) \le t, Y_{\tau(r)} \in [a, b]), \qquad t \ge 0.$$

We note that it is sufficient to show that $F(t)$ is a (left-) continuous function with the right-derivative

$$\partial_+ F(t) = \int_a^b y g_t(r, y)\,\mathrm{d}y =: f(t) \tag{I.22}$$

for all $t > 0$. This is so, because f is continuous and so with $G(t) = \int_0^t f(u)\,\mathrm{d}u$ we have $\partial_+(F(t) - G(t)) = 0$ for all $t > 0$, implying that $F(t)$ coincides with $G(t)$ on $t > 0$ up to a constant. By taking $t \downarrow 0$ we see that this constant is 0 and hence

$$F(t) = \int_0^t \int_a^b y g_u(r, y)\,\mathrm{d}y\,\mathrm{d}u$$

establishing the claim.

For $h > 0$ we note the identity

$$F(t + h) - F(t) = \mathbb{P}(I_t < r \le I_{t+h}, Y_{\tau(r)} \in [a, b]). \tag{I.23}$$

Moreover, $I_{t+h} = I_t + Y_t I'_h$ with I'_h corresponding to $\xi'_u = \xi_{t+u} - \xi_t$, and so the latter is independent of $\mathcal{F}_t$. Next, we note for any $z > 0$ that

$$\begin{aligned}\frac{1}{h}\mathbb{P}(I_t < r \le I_t + Y_t z h, Y_t \in [a, b]) &= \frac{1}{h}\int_a^b \int_{r-yzh}^r g_t(x, y)\,\mathrm{d}x\,\mathrm{d}y \\ &\to z\int_a^b y g_t(r, y)\,\mathrm{d}y.\end{aligned} \tag{I.24}$$

This follows from the mean value theorem and the fact that $g_t(x, y)$ is bounded for x of interest.

Define $\Delta'_h = \exp(\alpha \sup_{u\le h} \xi'_u)$ and note that $I'_h \le h\Delta'_h$. Moreover, for any $\epsilon > 0$ we may choose $c > 1$ large enough so that $\mathbb{P}(\Delta'_h > c) < \epsilon h$ for h small enough. This can be seen from the inequality (Gikhman and Skorokhod, 2004, Lem. 2, p. 420)

$$\mathbb{P}(\sup_{u\le h}|\xi'_u| > \log c/\alpha) \le (1+\mathrm{o}(1))\mathbb{P}(|\xi'_h| \ge \log c/(2\alpha)),$$

and the standard bounds on the right-hand side, see the argument in Sato (2013, Lem. 30.3). Thus in the following we may always assume that $\Delta'_h \le c$. Similarly, we may also assume that $\underline{\Delta}'_h = \exp(\alpha \inf_{u\le h} \xi'_u) \ge \underline{c} \in (0,1)$.

Now, we readily find that

$$\mathbb{P}(I_t < r \le I_t + Y_t ch, Y_t \in [a/c, b/\underline{c}], \Delta'_h > 1+\epsilon) = \mathrm{o}(h)$$

and the analogous statement with $\underline{\Delta}'_h < 1-\epsilon$. Hence we have the following upper bound on (I.23)

$$\mathbb{P}(I_t < r \le I_t + Y_t I'_h, I'_h < ch, Y_t \in [a/(1+\epsilon), b/(1-\epsilon)])$$

up to some negligible terms, and a similar lower bound. It is left to condition on I'_h/h, to apply the arguments from (I.24) and to notice that

$$\lim_{h\downarrow 0} \mathbb{E}(I_h/h\, \mathbf{1}_{I_h/h<c}) = 1,$$

where the latter is a consequence of the mean value theorem and the dominated convergence theorem. Hence (I.22) is now proven. Left-continuity of $F(t)$ follows from $\mathbb{P}(I_{t-h} < r \le I_{t-h} + (b/\underline{c})ch) \to 0$. □

Acknowledgements

The authors gratefully acknowledge financial support of Sapere Aude Starting Grant 8049-00021B "Distributional Robustness in Assessment of Extreme Risk" from Independent Research Fund Denmark.

References

D. J. Aldous and G. K. Eagleson. On mixing and stability of limit theorems. *Ann. Probability*, 6(2):325–331, 1978. ISSN 0091-1798.

R. Altmeyer. Central limit theorems for discretized occupation time functionals. *arXiv preprint arXiv:1909.00474*, 2019.

S. Asmussen and J. Ivanovs. Discretization error for a two-sided reflected Lévy process. *Queueing Syst.*, 89(1-2):199–212, 2018. ISSN 0257-0130.

E. J. Baurdoux, A. E. Kyprianou, and C. Ott. Optimal prediction for positive self-similar Markov processes. *Electron. J. Probab.*, 21:Paper No. 48, 24, 2016.

J. Bertoin. Splitting at the infimum and excursions in half-lines for random walks and Lévy processes. *Stochastic Process. Appl.*, 47(1):17–35, 1993. ISSN 0304-4149.

J. Bertoin. *Lévy Processes*. Number vb. 121 in Cambridge Tracts in Mathematics. Cambridge University Press, 1998. ISBN 9780521646321.

K. Bisewski and J. Ivanovs. Zooming-in on a Lévy process: Failure to observe threshold exceedance over a dense grid. *arXiv preprint arXiv:1904.06162*, 2020.

A. N. Borodin and P. Salminen. *Handbook of Brownian motion—facts and formulae*. Probability and its Applications. Birkhäuser Verlag, Basel, second edition, 2002. ISBN 3-7643-6705-9.

M. E. Caballero and L. Chaumont. Conditioned stable Lévy processes and the Lamperti representation. *Journal of Applied Probability*, 43(4):967–983, 2006. ISSN 00219002.

P. Carmona, F. Petit, and M. Yor. Exponential functionals of Lévy processes. In *Lévy processes*, pages 41–55. Springer, 2001.

L. Chaumont, A. Kyprianou, J. C. Pardo, and V. Rivero. Fluctuation theory and exit systems for positive self-similar Markov processes. *Ann. Probab.*, 40(1):245–279, 2012. ISSN 0091-1798.

S. Engelke and J. Ivanovs. A Lévy-derived process seen from its supremum and max-stable processes. *Electron. J. Probab.*, 21:Paper No. 14, 19, 2016.

I. I. Gikhman and A. V. Skorokhod. *The Theory of Stochastic Processes. I.* Classics in Mathematics. Springer-Verlag, Berlin, 2004. ISBN 3-540-20284-6. Translated from the Russian by S. Kotz, Reprint of the 1974 edition.

J. González Cázares, A. Mijatović, and G. Uribe Bravo. ε-strong simulation of the convex minorants of stable processes and meanders. *arXiv preprint arXiv:1910.13273*, 2019.

J. Ivanovs. Zooming in on a Lévy process at its supremum. *Ann. Appl. Probab.*, 28(2): 912–940, 04 2018.

J. Ivanovs and M. Podolskij. Optimal estimation of some random quantities of a Lévy process. *arXiv preprint arXiv:2001.02517*, 2020.

J. Jacod and P. Protter. *Discretization of Processes.* Stochastic Modelling and Applied Probability. Springer Berlin Heidelberg, 2011. ISBN 9783642241277.

J. Jacod, A. Jakubowski, and J. Mémin. On asymptotic errors in discretization of processes. *The Annals of Probability*, 31(2):592–608, 2003.

O. Kallenberg. *Foundations of Modern Probability.* Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. ISBN 0-387-95313-2.

P. Kosulajeff. Sur la répartition de la partie fractionnaire d'une variable. *Math. Sbornik*, 2 (5):1017–1019, 1937.

J. Lamperti. Semi-stable Markov processes. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 22:205–225, 1972.

J. C. Pardo. The upper envelope of positive self-similar Markov processes. *J. Theoret. Probab.*, 22(2):514–542, 2009. ISSN 0894-9840.

J. C. Pardo and V. Rivero. Self-similar Markov processes. *Bol. Soc. Mat. Mexicana (3)*, 19 (2):201–235, 2013. ISSN 1405-213X.

J. C. Pardo, V. Rivero, and K. van Schaik. On the density of exponential functionals of Lévy processes. *Bernoulli*, 19(5A):1938–1964, 2013.

M. Podolskij and M. Vetter. Understanding limit theorems for semimartingales: A short survey. *Statistica Neerlandica*, 64(3):329–351, 8 2010. ISSN 0039-0402.

A. Rényi. On stable sequences of events. *Sankhyā Ser. A*, 25:293 302, 1963. ISSN 0581-572X.

P. Salminen and L. Vostrikova. On exponential functionals of processes with independent increments. *Theory of Probability & Its Applications*, 63(2):267–291, 2018.

K.-I. Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2013. ISBN 978-1-107-65649-9. Translated from the 1990 Japanese original, Revised edition of the 1999 English translation.

J. W. Tukey. On the distribution of the fractional part of a statistical variable. *Math. Sbornik*, 4(3):561–562, 1938.

W. Whitt. *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer Series in Operations Research and Financial Engineering. Springer New York, 2002. ISBN 9780387953588.

Paper II

Lévy processes conditioned to stay in a half-space with applications to directional extremes

Jevgenijs Ivanovs
Department of Mathematics, Aarhus University

Jakob D. Thøstesen
Department of Mathematics, Aarhus University

Abstract. This paper provides a multivariate extension of Bertoin's pathwise construction of a Lévy process conditioned to stay positive/negative. Thus obtained processes conditioned to stay in half-spaces are closely related to the original process on a compact time interval seen from its directional extremal points. In the case of a correlated Brownian motion the law of the conditioned process is obtained by a linear transformation of a standard Brownian motion and an independent Bessel-3 process. Further motivation is provided by a limit theorem corresponding to zooming in on a Lévy process with a Brownian part at the point of its directional infimum. Applications to zooming in at the point farthest from the origin are envisaged.

Keywords: Conditioning to stay positive, directional extremes, exchangeability, local behavior, Sparre-Andersen identity

2010 MSC: 60G51; 60G17; 60F17

II.1 Introduction

There are multiple examples of conditioning a univariate Lévy process in some limiting sense, which alternatively can be described by Doob h-transforms, see Bertoin (1993), Döring et al. (2019) and Döring and Weissmann (2020) and references therein. Most often the focus is on establishing properties directly related to these conditional processes. The case of conditioning to stay positive/negative is special in the sense that it is intimately related to the post- and pre-infimum processes (Bertoin, 1993), leading to various important applications. Further links to path decomposition results can be found in Duquesne (2003).

Local behavior of a univariate Lévy process at its extremal points is studied in Ivanovs (2018), see also Bertoin (1993) for a self-similar case and Asmussen et al. (1995) for a linear Brownian motion. It is shown that zooming in at the point of infimum results in a pair of processes obtained from the underlying self-similar Lévy process conditioned to stay positive and negative. Further applications of this theory in the setting of high-frequency statistics include estimation of threshold exceedance in Bisewski and Ivanovs (2020) and optimal estimation of extremes in Ivanovs and Podolskij (2020). Bertoin's pathwise construction of conditioned processes in Bertoin (1993) plays a fundamental role in these works. For yet another application see Asmussen and Ivanovs (2018) studying the discretization error in the two-sided Skorokhod reflection map.

In this work we extend Bertoin's construction to the multivariate setting to define a Lévy process conditioned to stay in a half-space specified by some normal vector $\eta \neq 0$, see §II.3. Importantly, the link to post- and pre-extremum processes is preserved, where extrema are understood with respect to the direction η. Furthermore, in §II.4 we establish an associated invariance principle which, in particular, yields a limit result when zooming in on a Lévy process at the point of directional extremum. This is achieved via a short and direct argument relying on the path-wise construction. Applications of this result to high frequency statistics and the study of discretization errors in problems related to directional extrema and exceedance are anticipated.

In the multivariate case we have a continuum of possible directions, and the effect of linear transformations is studied in §II.5. It is shown that conditioning with respect to any direction η can be reduced to, say, conditioning an appropriately rotated process so that its first component stays positive. Furthermore, we provide a simple expression for the conditioned correlated Brownian motion in terms of a certain linear transformation of independent standard Brownian motions and a Bessel-3 process. In §II.6 we present the semigroup of the conditioned process in the general case, which turns out to have an intuitive structure. In §II.7 we utilize the arguments and insights from Chaumont and Doney (2005) to establish some important properties of the conditioned process. This leads to a natural definition of the respective Feller process started from an arbitrary point in the closed half-space.

We have attempted to present the multivariate theory in a streamlined and concise form, while emphasizing the main novelties stemming from the multivariate setting. Finally, in §II.8 we state a conjecture related to the local behavior at the point farthest from the origin, which hints at even greater application potential of the multivariate theory.

II.2 Preliminaries

Fix an integer $d \geq 1$ and let $\mathcal{D}$ denote the space of càdlàg functions $\omega : \mathbb{R} \to \mathbb{R}^d \cup \{\dagger\}$, where $\dagger$ is an isolated absorbing state. As usual we equip the path space $\mathcal{D}$ with the Skorokhod topology and let $\mathcal{F}$ denote the Borel σ-field. Furthermore we denote the coordinate process by $X = (X_t)$ and its natural completed filtration by $(\mathcal{F}_t)$. Unless stated otherwise we work with a subclass of processes satisfying $X_t = 0$ for $t < 0$, and let $\zeta := \inf\{t \geq 0 \mid X_t = \dagger\} \in [0, \infty]$ be the lifetime.

II.2.1 Directional infimum

We shall consider a fixed vector $\eta \in \mathbb{R}^d \setminus \{0\}$ and the respective open and closed half-spaces

$$S := \{x \in \mathbb{R}^d \mid \langle x, \eta \rangle > 0\} \quad \text{and} \quad \overline{S} := \{x \in \mathbb{R}^d \mid \langle x, \eta \rangle \geq 0\};$$

for ease of notation we omit η here and in the following. The *projected process* is defined by

$$Z_t := \langle X_t, \eta \rangle \in \mathbb{R} \cup \{\dagger\},$$

where $\langle \dagger, \eta \rangle = \dagger$ by convention.

Assume for a moment that the lifetime is finite and strictly positive, $\zeta \in (0, \infty)$. Consider the directional infimum $\underline{Z} := \inf\{Z_t \mid t \geq 0\}$ and the respective (last) time

$$\tau := \sup\{t \geq 0 \mid Z_t \wedge Z_{t-} = \underline{Z}\} \in [0, \zeta],$$

where $z \wedge \dagger = z$. Letting $\underline{X} := X_\tau \mathbf{1}_{\{Z_\tau \le Z_{\tau-}\}} + X_{\tau-} \mathbf{1}_{\{Z_\tau > Z_{\tau-}\}}$ be the position of X at the time of directional infimum, we define the (directional) post-infimum and reversed pre-infimum processes by

$$\underset{\rightarrow}{X}_t := \begin{cases} X_{\tau+t} - \underline{X} & \text{if } 0 \le t < \zeta - \tau, \\ \dagger & \text{if } t \ge \zeta - \tau, \end{cases}$$

$$\underset{\leftarrow}{X}_t := \begin{cases} X_{(\tau-t)-} - \underline{X} & \text{if } 0 \le t < \tau, \\ \dagger & \text{if } t \ge \tau, \end{cases}$$

see also Figure II.1 for a schematic illustration. According to the above convention we set $\underset{\rightarrow}{X}_t = \underset{\leftarrow}{X}_t = 0$ for $t < 0$. Note that $\underset{\rightarrow}{X}_t = \dagger$ for $t \ge 0$ if $\tau = \zeta$, and similarly $\underset{\leftarrow}{X}_t = \dagger$ for $t \ge 0$ if $\tau = 0$. The pair of processes $(\underset{\leftarrow}{X}, \underset{\rightarrow}{X})$ is a representation of the process X seen from the time-space point $(\tau, \underline{X})$. Alternatively, we could have defined a proper two-sided process.

II.2.2 Lévy processes

Throughout this paper $\mathbb{P}$ will be a probability measure on $(\mathcal{D}, \mathcal{F})$ such that X is a d-dimensional Lévy process with infinite lifetime. We write $X : \mathbb{P}$ when there is a need to specify the law of X explicitly. For a deterministic $T \in (0, \infty)$ the process $X : \mathbb{P}$ sent to $\dagger$ at T is denoted by $X : \mathbb{P}^T$, and in particular $\mathbb{P}^T(\zeta = T) = 1$. By default we work with $\mathbb{P}$ if no law is mentioned explicitly. The Lévy measure of X is denoted by $\Pi(\mathrm{d}x)$. Additional notation will be introduced in the following when required.

Throughout this paper we assume (the excluded case is simple but somewhat cumbersome):

Assumption II.A. *For the chosen direction η the projected process Z is not a compound Poisson process.*

Under Assumption II.A it is well known that the process $Z : \mathbb{P}^T$ achieves its infimum once only (at the time τ) a.s. This means that $\underset{\rightarrow}{X}$ and $\underset{\leftarrow}{X}$ are inside the open half-space S for strictly positive times preceding ζ. Our next result shows that X cannot jump perpendicularly to η at τ, see Figure II.1, and so $\underset{\rightarrow}{X}_0$ and $\underset{\leftarrow}{X}_0$ are either at the origin or inside $S \cup \{\dagger\}$. For the definition of regular/irregular points we refer to Bertoin (1996, p. 104).

Lemma II.1. *The following trichotomy holds with respect to the projected process $Z : \mathbb{P}$.*

($\updownarrow$) *If 0 is regular for $(-\infty, 0)$ and for $(0, \infty)$ then $\underset{\leftarrow}{X}_0 = \underset{\rightarrow}{X}_0 = 0$ $\quad \mathbb{P}^T$-a.s.*

($\uparrow$) *If 0 is irregular for $(-\infty, 0)$ then $\underset{\leftarrow}{X}_0 \in S \cup \{\dagger\}$ and $\underset{\rightarrow}{X}_0 = 0$ $\quad \mathbb{P}^T$-a.s.*

($\downarrow$) *If 0 is irregular for $(0, \infty)$ then $\underset{\leftarrow}{X}_0 = 0$ and $\underset{\rightarrow}{X}_0 \in S \cup \{\dagger\}$ $\quad \mathbb{P}^T$-a.s.*

Proof. The latter two statements are easy and follow from the univariate case. Suppose instead that 0 is regular for both half-lines, in which case $\mathbb{P}^T(\tau \in \{0, T\}) = 0$. We may choose a sequence (T_n) of stopping times, ranging over all jump epochs of X. Applying the strong Markov property yields $\mathbb{P}^T(Z_{T_n} = \underline{Z}) = 0$ since Z is regular for $(-\infty, 0)$. Thus, if X jumps at τ then $Z_\tau > \underline{Z}$ $\mathbb{P}^T$-a.s. The same argument applied to the time reversed process $(X_T - X_{(T-t)-})$ having the law of $X : \mathbb{P}^T$ shows that $Z_{\tau-} > \underline{Z}$ if X jumps at τ $\mathbb{P}^T$-a.s.; here we employ regularity for $(0, \infty)$. We conclude that X is $\mathbb{P}^T$-a.s. continuous at τ and this proves the statement. $\square$

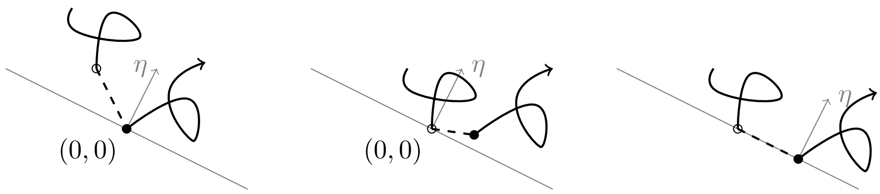


Figure II.1: Schematic illustration of the process in $\mathbb{R}^2$ seen from its directional infimum: (↑) jump into η-minimum (left), (↓) jump out of η-infimum (center) and an impossible case (right).

II.3 The fundamental representation and the limit object

We start with a fundamental representation of the law of the pair $(\underleftarrow{X}, \underrightarrow{X}) : \mathbb{P}^T$, which extends a univariate construction by Bertoin (1993) based, in turn, on an implicit identity for random walks appearing in Feller (1971, Lem. XII.8.3). Our representation is in terms of time-changed stochastic integrals, since the construction in Bertoin (1993) in terms of the local time at 0 does not have a simple analogue in the multivariate setting.

Consider the non-killed process X and let $\tilde{X}_t := X_{(-t)-}$ be its time-reversal, which is a process with stationary and independent increments for negative times. Define two $(\mathcal{F}_t)$-adapted càdlàg processes $Y^\pm$ by

$$Y_t^+ := -\int_{[-t,0]} \mathbf{1}_{\{\langle \tilde{X}_{s-}, \eta\rangle > 0\}} \,\mathrm{d}\tilde{X}_s, \qquad Y_t^- := -\int_{[-t,0]} \mathbf{1}_{\{\langle \tilde{X}_{s-}, \eta\rangle \le 0\}} \,\mathrm{d}\tilde{X}_s \quad \text{for } t \ge 0,$$

and $Y_\infty^\pm := \dagger$. These stochastic integrals can be understood intuitively as $\int_0^t \mathbf{1}_{\{\langle X_s, \eta\rangle > 0\}} \,\mathrm{d}X_s$ and $\int_0^t \mathbf{1}_{\{\langle X_s, \eta\rangle \le 0\}} \,\mathrm{d}X_s$, where the integrands are not predictable.

The cumulative times when X is and is not in S are denoted by A^+ and A^- respectively. That is,

$$A_t^+ := \int_0^t \mathbf{1}_{\{\langle X_s, \eta\rangle > 0\}} \,\mathrm{d}s, \qquad A_t^- := \int_0^t \mathbf{1}_{\{\langle X_s, \eta\rangle \le 0\}} \,\mathrm{d}s \quad \text{for } t \ge 0.$$

Consider now the right-continuous inverses $\alpha_t^\pm := \inf\{s \ge 0 \mid A_s^\pm > t\}$ of $A^\pm$, and define

$$X_t^\uparrow := Y^+_{\alpha_t^+}, \qquad X_t^\downarrow := Y^-_{\alpha_t^-} \quad \text{for } t \ge 0.$$

The processes $X^\uparrow$ and $X^\downarrow$ under $\mathbb{P}^T$ are obtained by killing $X^\uparrow$ and $X^\downarrow$ at the times A_T^+ and A_T^- under $\mathbb{P}$, respectively. The times $A_T^\pm$ are non-decreasing in T, which results in longer lifetimes $\zeta^\uparrow$ and $\zeta^\downarrow$ for larger time horizons T.

Theorem II.2. *Under $\mathbb{P}^T$ for $T \in (0, \infty)$ there is the following identity in law:*

$$(\underleftarrow{X}, \underrightarrow{X}) \stackrel{d}{=} (-X^\downarrow, X^\uparrow),$$

where $-\dagger = \dagger$ by convention.

Proof. The proof is based on a random walk approximation and exchangeability of increments as in the one-dimensional cases of Bertoin (1993); it is deferred to §II.A. □

Importantly, the above construction of the pair $(X^\downarrow, X^\uparrow) : \mathbb{P}^T$ depends on T via the killing times $A_T^\pm$ alone. In particular, for $0 < T_1 < T_2$ the paths of $X^\uparrow : \mathbb{P}^{T_1}$ and $X^\uparrow : \mathbb{P}^{T_2}$ coincide up to the time $A_{T_1}^+$ when the former is sent to $\dagger$, whereas the latter is killed at $A_{T_2}^+ \ge A_{T_1}^+$. It is convenient to think of growing the paths as T increases. As $T \to \infty$ we obtain $(X^\downarrow, X^\uparrow)$.

Corollary II.3. *It holds that*

$$(\underleftarrow{X}, \underrightarrow{X}) : \mathbb{P}^T \xrightarrow{d} (-X^{\downarrow}, X^{\uparrow}) \quad \textit{as } T \to \infty.$$

It is noted that the above weak convergence statement can be strengthened, see Bertoin (1993, Cor. 3.2), but we prefer using Theorem II.2 directly when needed. The pair $(X^{\downarrow}, X^{\uparrow})$ is our main object of interest. According to Corollary II.3, the process $X^{\uparrow} : \mathbb{P}$ can be called a limiting post-infimum process. In analogy to the univariate case we instead call it X *conditioned to stay in the half-plane* S, and provide a justification below.

Observe that $-X^{\downarrow}_t, X^{\uparrow}_t \in S \cup \{\dagger\}$ for $t > 0$ a.s., whereas the initial values are classified according to the trichotomy in Lemma II.1. In particular, $X^{\uparrow}_0 = 0$ in cases $(\updownarrow), (\uparrow)$, and $X^{\downarrow}_0 = 0$ in cases $(\updownarrow), (\downarrow)$. Importantly, the projected conditioned processes $\langle X^{\uparrow}, \eta\rangle$ and $\langle X^{\downarrow}, \eta\rangle$ coincide with the univariate Lévy process Z conditioned to stay positive and negative, respectively. In particular, the lifetimes $\zeta^{\uparrow}$ and $\zeta^{\downarrow}$ can be studied using the univariate theory, and so

$$\zeta^{\uparrow} = \infty \quad \text{iff} \quad \limsup_{t\to\infty} Z_t = \infty, \qquad \zeta^{\downarrow} = \infty \quad \text{iff} \quad \liminf_{t\to\infty} Z_t = -\infty$$

with probability 1. Furthermore, $\zeta^{\uparrow} > 0$ unless Z is a non-increasing process and then $\zeta^{\uparrow} = 0$ a.s. Yet another useful observation is given by the following result.

Lemma II.4. *The processes $X^{\uparrow}$ and $X^{\downarrow}$ do not jump at a fixed $t > 0$ a.s.*

Proof. Assume that $\underleftarrow{X} : \mathbb{P}^T$ jumps at $t > 0$ with positive probability. Then by an argument as in the proof of Lemma II.1 we find that we must be in the case $(\uparrow)$. Hence X has two jumps separated by time t with positive probability, which is impossible. According to Theorem II.2 we find that $X^{\downarrow}$ has no jump at t a.s. when excluding the jump into $\dagger$. The latter would imply $\mathbb{P}(\zeta^{\downarrow} = t) > 0$, which is again impossible by a similar argument. By time-reversal the same property is true with respect to $X^{\uparrow}$. □

Importantly, (under Assumption II.A) the process $X^{\downarrow}$ is a.s. the same if the non-strict inequalities in its definition are replaced by strict inequalities, which follows from basic properties of Lévy processes. In particular, we find that $X^{\downarrow} = -(-X)^{\uparrow}$ a.s. The respective equality in distribution can also be seen using the representation in Theorem II.2 and the standard time-reversal argument. Finally, observe a close link to the classical Sparre-Andersen identity (Bertoin, 1996, Lem. VI.15): A^+_T has the same law as the time of the supremum of Z on $[0, T]$, which by time-reversal coincides with the law of the lifetime of the respective post-infimum process.

II.4 Motivating limit theorem

Bertoin's representation and its above stated generalization are indispensable in the study of Lévy processes around their extremes. In the one-dimensional setting it has been fundamental for the results in Bisewski and Ivanovs (2020) and Ivanovs and Podolskij (2020). We further demonstrate its usefulness by establishing an invariance principle, see Chaumont and Doney (2010) and Ivanovs (2018) for alternative approaches in the univariate case (the latter needs a better justification of convergence of Markov processes). The following short proof requires certain assumptions, and for simplicity we consider only the case of an oscillating $Z_t = \langle X_t, \eta\rangle$:

$$\limsup_{t\to\infty} Z_t = \infty \quad \text{and} \quad \liminf_{t\to\infty} Z_t = -\infty \quad \text{a.s.} \tag{II.1}$$

Recall that this assumption implies that both $X^{\uparrow}$ and $X^{\downarrow}$ have infinite lifetimes.

Theorem II.5. *Let $X^{(n)}$ be a sequence of Lévy processes weakly convergent to a Lévy process X satisfying* (II.1) *and Assumption II.A. Then for any sequence of finite deterministic times $T_n \to \infty$ there is the weak convergence*

$$(\underleftarrow{\underline{X}}^{(n)}, \underrightarrow{\underline{X}}^{(n)}) : \mathbb{P}^{T_n} \xrightarrow{d} (-X^{\downarrow}, X^{\uparrow}).$$

Proof. Fix an arbitrary finite $T > 0$. By the continuous mapping theorem we have under $\mathbb{P}^T$:

$$(\underleftarrow{\underline{X}}^{(n)}, \underrightarrow{\underline{X}}^{(n)}) \xrightarrow{d} (\underleftarrow{\underline{X}}, \underrightarrow{\underline{X}}).$$

Indeed, for converging paths the directional infima and their (right) times must converge assuming the limiting path has no jump at T and it achieves the directional infimum only once (this is a.s. true). Furthermore, X has no jump perpendicular to η at τ, see Lemma II.1 and Figure II.1 (right). Note that making all processes stay at 0 for negative times is essential in the case when the limit process jumps at τ.

According to Theorem II.2 we have

$$(-X^{(n)\downarrow}, X^{(n)\uparrow}) : \mathbb{P}^T \xrightarrow{d} (-X^{\downarrow}, X^{\uparrow}) : \mathbb{P}^T$$

for every $T > 0$, and the latter weakly converges to $(-X^{\downarrow}, X^{\uparrow})$ as $T \to \infty$. Thus it is left to apply a standard approximation result, see Billingsley (1999, Thm. 3.2) or Kallenberg (2002, Thm. 4.28), to obtain

$$(-X^{(n)\downarrow}, X^{(n)\uparrow}) : \mathbb{P}^{T_n} \xrightarrow{d} (-X^{\downarrow}, X^{\uparrow}), \tag{II.2}$$

and hence also the stated result (apply Theorem II.2 to the left hand side). The crux of the approximation result consists in showing that the Skorokhod distance (on each compact time interval $[0, t]$) between the left hand side in (II.2) and the same object for the time horizon T converges to 0 in probability as $T \to \infty$ uniformly for large n. In our case it is sufficient to check that

$$\lim_{T\to\infty} \limsup_{n} \mathbb{P}(A^{(n)\pm}_{T_n} \wedge A^{(n)\pm}_{T} > t) = 1, \qquad t > 0,$$

where the event corresponds to two identical paths on the time interval $[0, t]$. We may assume that $T_n \geq T$ implying $A^{(n)\pm}_{T_n} \geq A^{(n)\pm}_{T}$, but the latter weakly converges to $A^{\pm}_{T}$. Finally, note that (II.1) implies $A^{\pm}_{\infty} = \infty$ a.s. □

The above argument can be adapted to include the case where $\lim_{t\to\infty} Z_t = \infty$ and $\lim_{t\to\infty} Z^{(n)}_t = \infty$ for all large enough n, as well as the case with $-\infty$ limits. That is, the infinite-time behavior of Z and the approximating sequence $Z^{(n)}$ is the same. Otherwise, the proof becomes substantially more difficult and it is then required to work with a compactified space where † is a point at infinity.

Finally, we show that zooming in on X at the time-space location of the directional infimum results in the pair of conditioned processes corresponding to the underlying Brownian part. This limit law is studied in Proposition II.9 below.

Corollary II.6. *Let B be the Brownian part of the d-dimensional X, and assume that $\langle B_1, \eta\rangle$ is not a.s. zero. Then*

$$\sqrt{n}(\underleftarrow{\underline{X}}_{\cdot/n}, \underrightarrow{\underline{X}}_{\cdot/n}) : \mathbb{P}^1 \xrightarrow{d} (-B^{\downarrow}, B^{\uparrow}).$$

Proof. Define a scaled time-changed process $X^{(n)}_t = \sqrt{n}X_{t/n}$ and note that $X^{(n)} \xrightarrow{d} B$, see Bertoin (1996, Prop. 2) and Kallenberg (2002, Thm. 15.17). It is left to apply Theorem II.5 with $T_n = n$. □

II.5 Linear transformations and the Brownian example

Linear transformations play an important role in the multivariate theory as demonstrated by the following result.

Lemma II.7. *Consider a $d' \times d$ matrix M and d'-dimensional vector $\eta' \neq 0$ such that $M^\top \eta' \neq 0$. Then $(MX)^\uparrow$ defined using $\eta = \eta'$ coincides with $M(X^\uparrow)$ defined using $\eta = M^\top \eta'$.*

Proof. Note that $\langle MX, \eta' \rangle = \langle X, M^\top \eta' \rangle$ and use linearity of the stochastic integral in the definition of $Y^\pm$. □

Consequently, it suffices to study conditioning for just one direction, say

$$\eta_1 = (1, 0, \dots, 0)^\top \in \mathbb{R}^d.$$

For any unit vector $\eta \in \mathbb{R}^d$ we may choose an orthogonal matrix R ($RR^\top = I$) such that $R\eta = \eta_1$. Then $X^\uparrow$ coincides with $R^\top (RX)^\uparrow$ where the latter is defined for the direction η_1. Our next result allows us reduce certain multivariate cases to the univariate theory.

Lemma II.8. *Consider $X = X'v + X''$, where X' and X'' are independent Lévy processes with dimensions 1 and d respectively, and additionally $\langle v, \eta \rangle > 0$, $\langle X''_t, \eta \rangle = 0, t \geq 0$. Then $X^\uparrow \overset{d}{=} X'^\uparrow v + X''$, where $X'^\uparrow$ is the univariate X' conditioned to stay positive and by convention $\dagger \cdot v + x'' = \dagger$.*

Proof. Note that the process $\underset{\rightarrow}{X} : \mathbb{P}^T$ has the same law as $\underset{\rightarrow}{X'} v + X'' : \mathbb{P}^T$, where $\underset{\rightarrow}{X'}$ is the post-infimum process of univariate X'. This is so, because $Z_t = \langle v, \eta \rangle X'_t$ and the process X'' is independent of τ, whereas $\underset{\rightarrow}{X' v} = \underset{\rightarrow}{X'} v$ under $\mathbb{P}^T$. It is left to apply Corollary II.3 and the continuous mapping theorem. □

We are now ready to treat the basic example of a conditioned Brownian motion. In this regard note that a univariate standard Brownian $B^{(1)}$ conditioned to stay positive is a Bessel-3 process which we denote by $B^{(1)\uparrow}$.

Proposition II.9. *Let X be a (driftless) Brownian motion with covariance matrix Σ such that $\Sigma \eta \neq 0$. Then*

$$X^\uparrow \overset{d}{=} -X^\downarrow \overset{d}{=} MR(B^{(1)\uparrow}, B^{(2)}, \dots, B^{(d)})^\top,$$

where $B = (B^{(1)}, \dots, B^{(d)})^\top$ is a standard Brownian motion in $\mathbb{R}^d$, and the square matrices M and R satisfy

$$MM^\top = \Sigma, \qquad RR^\top = I, \qquad R^\top M^\top \eta = \sqrt{\eta^\top \Sigma \eta}\, \eta_1.$$

Proof. The first distributional equality is a consequence of $-X \overset{d}{=} X$. Next, using $X \overset{d}{=} MRB$ and Lemma II.7 we find that $X^\uparrow$ has the law of $MR(B^\uparrow)$ for the direction $R^\top M^\top \eta$, where the latter is proportional to η_1. It is left to apply Lemma II.8 to find that $B^\uparrow$ for the direction η_1 has the law of $(B^{(1)\uparrow}, B^{(2)}, \dots, B^{(d)})^\top$. □

Example II.10. *Take $d = 2, \eta = (a, b)^\top$ and a Brownian motion X with standard deviations $\sigma_1, \sigma_2 > 0$ and correlation $\rho \in (-1, 1)$. Then Proposition II.9 yields*

$$X^\uparrow \overset{d}{=} \frac{1}{\sqrt{a^2\sigma_1^2 + 2ab\sigma_1\sigma_2\rho + b^2\sigma_2^2}} \begin{pmatrix} a\sigma_1^2 + b\sigma_1\sigma_2\rho & -b\sigma_1\sigma_2\sqrt{1-\rho^2} \\ a\sigma_1\sigma_2\rho + b\sigma_2^2 & a\sigma_1\sigma_2\sqrt{1-\rho^2} \end{pmatrix} \begin{pmatrix} B^{(1)\uparrow} \\ B^{(2)} \end{pmatrix},$$

where we used a Cholesky square-root M.

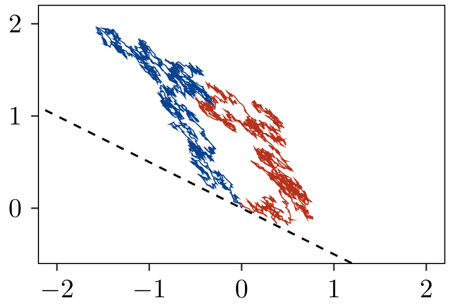
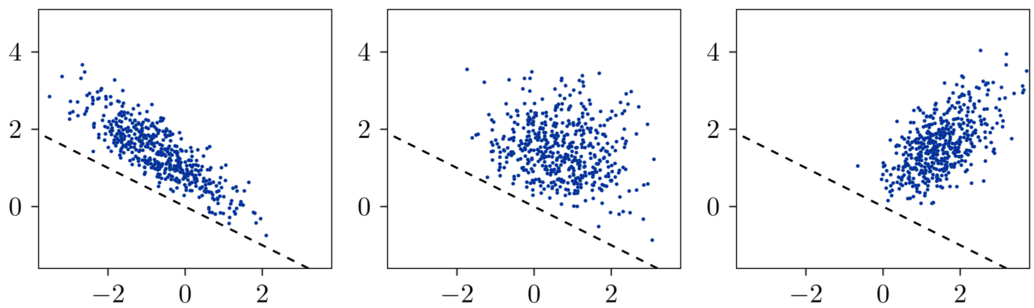


Figure II.2: Two independent paths simulated from the common law of $-X^{\downarrow}$ and $X^{\uparrow}$ for the direction $\eta = (1,2)^{\top}$, where $\sigma_1 = \sigma_2 = 1$ and $\rho = -0.8$. The dashed line is the boundary of the corresponding half-space.

In particular, simulation of $X^{\uparrow}$ over a grid is a trivial task when X is a driftless Brownian motion. We depict two independent sample paths in Figure II.2.

Further insight can be obtained from Figure II.3 consisting of three plots, each containing 500 simulations of $X_1^{\uparrow}$ for different values of ρ.



(a) $\rho = -0.8$. **(b)** $\rho = 0$. **(c)** $\rho = 0.8$.

Figure II.3: Simulated values of $X_1^{\uparrow}$ for the direction $\eta = (1,2)^{\top}$ and for different values of ρ. The standard deviations are $\sigma_1 = \sigma_2 = 1$.

II.6 The law of the limit pair

We need some additional notation. Consider an extension of the probability space $(D, \mathcal{F}, \mathbb{P})$ supporting a standard exponential random variable e_1 independent of everything else. Define $e_q = e_1/q$, an exponential random variable of rate $q > 0$, and let $X : \mathbb{P}^{e_q}$ be the process $X : \mathbb{P}$ killed at e_q. Finally, the process $X : \mathbb{P}_x^{e_q}$ corresponds to the shifted process $X_t + x\,\mathbf{1}_{\{t \geq 0\}}$ killed at e_q, and in the case of no killing we write $\mathbb{P}_x$.

II.6.1 Exponential time horizon

As in the univariate case the characterization of the law of the limit object $(-X^{\downarrow}, X^{\uparrow})$ proceeds by first studying the respective pair of processes under $\mathbb{P}^{e_q}$, that is, when the killing time of the original process is an independent exponential random variable of rate $q > 0$. We start with a simple observation that under $\mathbb{P}^{e_q}$ (and Assumption II.A) we have

$$(\underleftarrow{X}, \underrightarrow{X}) \overset{d}{=} (-\underrightarrow{X}, -\underleftarrow{X}),$$

which readily follows by time-reversal; alternatively one may use Theorem II.2. The following splitting result is based on some classical arguments, and we only provide appropriate references.

Proposition II.11. *Under $\mathbb{P}^{e_q}$ the processes $\underleftarrow{X}$ and $\underrightarrow{X}$ are independent Markov processes. The semigroup of $\underrightarrow{X}$ is given by*

$$\frac{\mathbb{P}_x^{e_q}(X_t \in \mathrm{d}y, \underline{Z}_t > 0)\mathbb{P}_y^{e_q}(\underline{Z} > 0)}{\mathbb{P}_x^{e_q}(\underline{Z} > 0)} = \mathbb{P}_x^{e_q}(X_t \in \mathrm{d}y \,|\, \underline{Z} > 0), \qquad t > 0, x, y \in S.$$

Moreover, in case (↓) the initial distribution is given by

$$\mathbb{P}^{e_q}(\underrightarrow{X}_0 \in \mathrm{d}y) = \mathbb{P}_y^{e_q}(\underline{Z} > 0)\Pi(\mathrm{d}y)/\big(q + \int_{\{\langle z,\eta\rangle > 0\}} \mathbb{P}_z^{e_q}(\underline{Z} > 0)\Pi(\mathrm{d}z)\big), \qquad y \in S.$$

Proof. The fact that $\underrightarrow{X}$ is Markov with the stated semigroup is proven in Millar (1978). Independence of the processes follows by discretizing the local time of Z at its infimum as in the proof by Bertoin (1996, Lem. VI.6). The initial distribution in case (↓) can be obtained analogously to Ivanovs (2017, Prop. 3.3) using an enumeration of jumps of X. □

II.6.2 Infinite time horizon

We are now ready to characterize the law of the limit object $(-X^{\downarrow}, X^{\uparrow})$. Consider a so-called renewal function associated to the ladder height process $\underline{H}$ corresponding to $-Z$:

$$h(x) := \int_0^\infty \mathbb{P}(\underline{H}_t \le x)\,\mathrm{d}t,$$

where the scaling of local time is arbitrary, see also Bertoin (1996, p. 157, 171). This is exactly the h-function appearing in the Doob h-transform corresponding to the univariate Z conditioned to stay positive, see Bertoin (1993) and Chaumont and Doney (2005) for alternative representations. The function h is finite, continuous and increasing.

Theorem II.12. *The processes $-X^{\downarrow}$ and $X^{\uparrow}$ are independent Markov processes, and the former has the law of $(-X)^{\uparrow}$. The semigroup of $X^{\uparrow}$ is given by*

$$p_t^{\uparrow}(x, \mathrm{d}y) := \frac{h(\langle y, \eta\rangle)}{h(\langle x, \eta\rangle)}\mathbb{P}_x(X_t \in \mathrm{d}y, \underline{Z}_t > 0), \qquad t > 0,\ x, y \in S.$$

Furthermore, in case (↓) and if Z is non-monotone we have

$$\mathbb{P}(X_0^{\uparrow} \in \mathrm{d}y) = h(\langle y, \eta\rangle)\Pi(\mathrm{d}y)/\int_{\{\langle z,\eta\rangle > 0\}} h(\langle z, \eta\rangle)\Pi(\mathrm{d}z), \qquad y \in S. \tag{II.3}$$

Proof. We apply Theorem II.2 with $T = e_q$ and Proposition II.11, and then let $q \downarrow 0$. Since $e_q \to \infty$ we indeed retrieve $-X^{\downarrow}$ and $X^{\uparrow}$. The Markov property follows from the strong convergence result implied by Theorem II.2 upon recalling that the distribution of $\zeta^{\uparrow}$ has no atoms, see Lemma II.4. Let us check that the semigroup in Proposition II.11 has the stated weak limit. From the univariate theory (Chaumont and Doney, 2005, Eq. (2.5)) we know that for a certain $c_q > 0$

$$\mathbb{P}_x^{e_q}(\underline{Z} > 0)/c_q \to h(\langle x, \eta\rangle), \qquad x \in S$$

as $q \downarrow 0$, and it is left to apply the dominated convergence theorem as in Chaumont and Doney (2005, Prop. 1).

With respect to the initial distribution (for the assumed case) we observe that

$$\mathbb{P}^{e_q}(\underset{\rightarrow}{X}_0 \in A) \to \mathbb{P}(X_0^\uparrow \in A), \qquad \int_A \mathbb{P}_y^{e_q}(\underline{Z} > 0)/c_q \Pi(\mathrm{d}y) \to \int_A h(\langle y, \eta\rangle)\Pi(\mathrm{d}y) < \infty$$

for any bounded Borel set A, also bounded away from 0 (by the dominated convergence theorem). It is left to recall that $\zeta^\uparrow > 0$ and $X_0^\uparrow \in S$ a.s., and A can be chosen so that $\int_A h(\langle y, \eta\rangle)\Pi(\mathrm{d}y) > 0$. The latter is true since $h(z) > 0$ for $z > 0$ and necessarily $\Pi(S) > 0$. $\square$

In the univariate case the initial distribution formula (II.3) is known in the case of no negative jumps, where $\underline{H}(t) = t$ implying $h(x) = x$, see Chaumont (1994) and also Chaumont and Doney (2005, Eq. (2.12)). Let us also stress the following relation to the univariate conditioned processes.

Remark II.13. *Choosing the direction* $\eta_1 = (1, 0, \dots, 0)^\top$ *we observe that*

$$p_t^\uparrow(x, \mathrm{d}y) = p_t^{(1)\uparrow}(x^{(1)}, \mathrm{d}y^{(1)})\mathbb{P}_{x^{(2:d)}}(X_t^{(2:d)} \in \mathrm{d}y^{(2:d)} \mid X_t^{(1)} = y^{(1)}, \underline{X}_t^{(1)} > 0),$$

where $X = (X^{(1)}, X^{(2:d)})$ *and* $p_t^{(1)\uparrow}$ *corresponds to* $X^{(1)}$ *conditioned to stay positive.*

II.7 Starting away from the origin

Theorem II.12 characterizes the law of $X^\uparrow$ in case ($\downarrow$), but otherwise it lacks convergence of the semigroup as $x \to 0$. In this section we address this issue and also state a number of further useful properties. The proofs follow closely the univariate analogues in Chaumont and Doney (2005) and thus we only state the main steps and observations.

It is easy to see that the semigroup $p_t^\uparrow(x, \mathrm{d}y)$ of $X^\uparrow$, see Theorem II.12, is conservative and satisfies the Feller properties on $\hat{S} := S \cup \{\dagger\}$. Note that the hyperplane defining this half-space has been excluded. We write $X : P_x^\uparrow$ for the respective Feller process indexed by $[0, \infty)$ and started at $x \in \hat{S}$, and note that it satisfies the strong Markov property (Kallenberg, 2002, Thm. 19.17).

Observe that the law of the Markov process with the semigroup in Proposition II.11 when started in $x \in \hat{S}$ can be conveniently written as

$$(X \mid \underline{Z} > 0) = (X \mid X_t \in \hat{S}\ \forall t \geq 0) \quad \text{under } \mathbb{P}_x^{e_q}. \tag{II.4}$$

Furthermore, Chaumont and Doney (2005, Prop. 1) readily generalizes to

$$P_x^\uparrow(\Lambda, t < \zeta) = \lim_{q \downarrow 0} \mathbb{P}_x^{e_q}(\Lambda, t < \zeta \mid \underline{Z} > 0), \qquad \Lambda \in \mathcal{F}_t, t > 0, x \in S, \tag{II.5}$$

which explains the name 'conditioned to stay in a half-space'. Note that $\langle X, \eta\rangle : P_x^\uparrow$ is the univariate process $\langle X, \eta\rangle$ conditioned to stay positive and started from $\langle x, \eta\rangle$.

Proposition II.14. *For any* $x \in S$ *the process* $Z_t = \langle X_t, \eta\rangle$ *under* $P_x^\uparrow$ *has a unique and finite time of infimum,* $\underset{\rightarrow}{X}$ *and* $\underset{\leftarrow}{X}$ *are independent under* $P_x^\uparrow$*, and*

$$\underset{\rightarrow}{X} : P_x^\uparrow \overset{d}{=} X^\uparrow.$$

Furthermore,

$$X : P_x^\uparrow \xrightarrow{d} X^\uparrow \quad \text{as } S \ni x \to 0,$$

where by convention the sample paths satisfy $\omega_t = 0, t < 0$.

Proof. It follows from the calculations in Chaumont and Doney (2005, p. 956) that the time of infimum is finite. Consider the process in (II.4) and establish a splitting result analogous to Proposition II.11. The post-infimum process has the law of $\underset{\rightarrow}{X} : \mathbb{P}^{e_q}$, and so we can apply (II.5) and Theorem II.2 to get the first statement.

In view of the first part, it is only required to show that the pre-infimum process $\underset{\leftarrow}{X} : P_x^{\uparrow}$ becomes negligible in probability as $x \to 0$. The arguments of Chaumont and Doney (2005, Thm. 2) still apply, and we additionally show that the maximal fluctuation of the pre-limit process perpendicular to η is negligible. This can be done by considering the stopping time $\nu = \inf\{t \geq 0 \mid \|X_t - x\| > \epsilon\}$ and employing similar analysis based on the strong Markov property. In case ($\updownarrow$) and ($\downarrow$) we then need to show that $\mathbb{P}_x(\underline{Z}_\nu > 0) \to 0$, which is indeed true. ☐

The above proof, in fact, shows that

$$X : P_x^{\uparrow} \xrightarrow{d} x_0 + X^{\uparrow} \quad \text{as } S \ni x \to x_0, \langle x_0, \eta\rangle = 0.$$

In cases ($\updownarrow$), ($\uparrow$) the process $x_0 + X^{\uparrow}$ starts at x_0 and according to Lemma II.4 it does not jump at fixed times. Hence in these cases we may extend our Feller process $X : P_x^{\uparrow}$ to the state space $\overline{S} \cup \{\dagger\}$, the closed half-space with an absorbing state, by setting

$$X : P_{x_0}^{\uparrow} := x_0 + X^{\uparrow} \quad \text{for any } x_0 \text{ with } \langle x_0, \eta\rangle = 0.$$

Note that this definition coincides with the result of the construction presented in Section II.3 if we take X started at x_0 and let $Y_t^+ = -\int_{[-t,0]} \mathbf{1}_{\{\langle \tilde{X}_{s-}, \eta\rangle \geq 0\}} \, \mathrm{d}\tilde{X}_s$ which yields an a.s. identical process in the original case.

II.8 Conjecture: zooming in at the maximal distance from the origin

For a possible further application we turn our attention to the local behavior of a Lévy process at the time when it reaches the maximal distance from the origin. Assuming finite life time, $\zeta \in (0, \infty)$, we let

$$\tau := \sup\{t \geq 0 \mid \|X_t\| \vee \|X_{t-}\| = \sup_{s\geq 0}\|X_s\|\} \in [0, \zeta]$$

be the (last) time when the Euclidean norm is maximal. Consider the respective position $M := X_\tau$ if $\|X_\tau\| \geq \|X_{\tau-}\|$ and $M := X_{\tau-}$ otherwise, and define the processes

$$\overrightarrow{X}_t := \begin{cases} X_{\tau+t} - M & \text{if } 0 \leq t < \zeta - \tau, \\ \dagger & \text{if } t \geq \zeta - \tau, \end{cases}$$

$$\overleftarrow{X}_t := \begin{cases} X_{(\tau-t)-} - M & \text{if } 0 \leq t < \tau, \\ \dagger & \text{if } t \geq \tau. \end{cases}$$

Observe that the pair $(\overleftarrow{X}, \overrightarrow{X})$ coincides with $(\underset{\leftarrow}{X}, \underset{\rightarrow}{X})$ studied above for the (path-dependent) direction $\eta = -M$, see also Figure II.4 for a schematic illustration. Inspired by Corollary II.6 and using the intuition from the one-dimensional stable convergence in Ivanovs (2018), we conjecture the following result; proving it seems to be exceedingly challenging at the moment. We anticipate that the convergence is again stable (Aldous and Eagleson, 1978) but avoid complicating the statement.

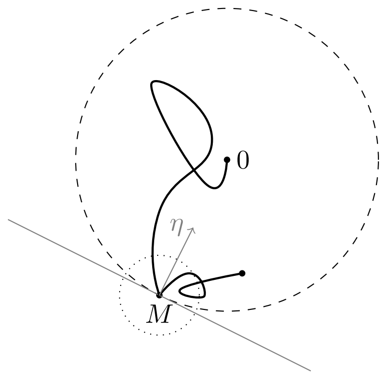


Figure II.4: Schematic illustration of zooming in at the maximal distance.

Conjecture II.15. *Let B be the Brownian part of X with a non-singular covariance matrix. Then*

$$\sqrt{n}(\overleftarrow{X}_{\cdot/n}, \overrightarrow{X}_{\cdot/n}) : \mathbb{P}^1 \xrightarrow{d} (-B^{\Downarrow}, B^{\Uparrow}),$$

where the limit pair is a mixture of $(-B^{\downarrow}, B^{\uparrow})$ for the independent direction $\eta = -M : \mathbb{P}^1$.

We illustrate this conjecture by a simulation study where $X = B$ is a 2-dimensional Brownian motion with correlation $\rho = -0.8$ as in Section II.5. We simulate K (approximate) copies of the random vector $\sqrt{n}\overrightarrow{X}_{1/n}$ under $\mathbb{P}^1$ for $n = 1000$ using discretization with step size 10^{-5}. The K samples of the limit quantity $B_1^{\Uparrow}$ are constructed by reusing the directions $\eta = -M : \mathbb{P}^1$ and then independently sampling $B_1^{\uparrow}$ according to Example II.10.

Note that $\overrightarrow{X}$ may have a lifetime strictly smaller than $1/n$, making $\sqrt{n}\overrightarrow{X}_{1/n}$ undefined. In our simulation we exclude these cases, effectively conditioning $\overrightarrow{X}$ to have a lifetime larger than $1/n$. We simulated 5000 times, resulting in $K = 4833$ (conditional) samples of $\sqrt{n}\overrightarrow{X}_{1/n}$. The respective bivariate densities are presented in Figure II.5, and we observe that they are indeed rather close.

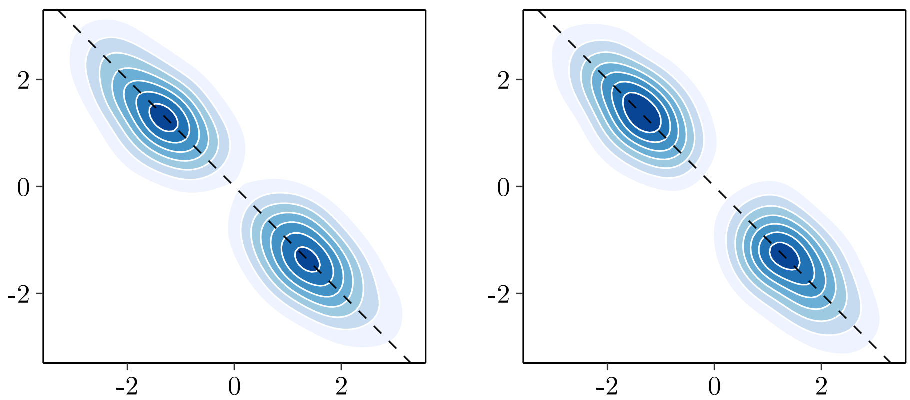


Figure II.5: Estimated bivariate densities for $\sqrt{n}\overrightarrow{X}_{1/n}$ and $B_1^{\Uparrow}$ on the left and right respectively. A darker shade of blue indicates a higher density. The dashed line is the line through the origin with slope -1.

It is noted that the Brownian motion X with correlation ρ tends to achieve its maximal distance from the origin in the NW or SE direction, which leads to the two clusters of points in Figure II.5.

II.A Proof of Theorem II.2

II.A.1 Discrete time

We begin by stating a discrete-time version of Theorem II.2. Fix $\zeta \in \mathbb{N}$ and consider a process $X \in \mathbb{R}^d$ over the index set $\{0, \ldots, \zeta\}$ together with the projected process $Z_i := \langle X_i, \eta \rangle$. Let $\tau := \sup\{i \le \zeta \mid Z_i = \underline{Z}_i\}$ be the index of the (last) minimum of Z, and $\underline{X} := X_\tau$ be the value of the directional minimum. The directional post-minimum and reversed pre-minimum chains $\underset{\rightarrow}{X}$ and $\underset{\leftarrow}{X}$ are given by

$$\underset{\rightarrow}{X}_i := \begin{cases} X_{\tau+i} - \underline{X} & \text{if } i \le \zeta - \tau, \\ \dagger & \text{if } i > \zeta - \tau, \end{cases}$$

$$\underset{\leftarrow}{X}_i := \begin{cases} X_{\tau-i} - \underline{X} & \text{if } i \le \tau, \\ \dagger & \text{if } i > \tau. \end{cases}$$

Next, define

$$A_i^+ := \sum_{j=1}^{i} \mathbf{1}_{\{Z_j > 0\}}, \qquad A_i^- := \sum_{j=1}^{i} \mathbf{1}_{\{Z_j \le 0\}} \tag{II.6}$$

when $i \le \zeta$, and let $\alpha_i^\pm := \inf\{j \in \mathbb{N} \mid A_j^\pm = i\}$ denote the inverses of $A^\pm$. With $\Delta X_j := X_j - X_{j-1}$ we define the chains $X^\uparrow$ and $X^\downarrow$ by

$$X_i^\uparrow := \begin{cases} \sum_{j=1}^{\alpha_i^+} \mathbf{1}_{\{Z_j > 0\}} \Delta X_j & \text{if } i \le A_\zeta^+ \\ \dagger & \text{if } i > A_\zeta^+, \end{cases}$$

$$X_i^\downarrow := \begin{cases} \sum_{j=1}^{\alpha_i^-} \mathbf{1}_{\{Z_j \le 0\}} \Delta X_j & \text{if } i \le A_\zeta^- \\ \dagger & \text{if } i > A_\zeta^-. \end{cases}$$

We are now ready to state the discrete analogue of Theorem II.2.

Theorem II.16. *Assume that $\zeta \in \mathbb{N}$ and X has exchangeable increments. Then the pairs of processes $(X^\downarrow, X^\uparrow)$ and $(-\underset{\leftarrow}{X}, \underset{\rightarrow}{X})$ have the same law.*

Proof. The proof of Bertoin (1993, Thm. 2.1) is easily adapted to this setting. □

II.A.2 Continuous time

The proof of Theorem II.2 proceeds much like the proof of Bertoin (1993, Thm. 3.1). We discretize, apply Theorem II.16 and take the limit.

Recall that we are considering a Lévy process $X : \mathbb{P}^T$ up to a finite time horizon $T > 0$. For each $n \in \mathbb{N}$ let X^n be the chain given by $X_i^n := X_{i/n}$, and let $X^{n\uparrow}$ and $X^{n\downarrow}$ be the chains obtained from X^n by the procedure in §II.A.1. Define

$$Y_i^{n+} := \sum_{j=1}^{i} \mathbf{1}_{\{\langle X_j^n, \eta \rangle > 0\}} \Delta X_j^n,$$

and note that almost surely

$$Y_{[tn]}^{n+} = -\sum_{i=-[tn]}^{-1} \mathbf{1}_{\{\langle \tilde{X}_{i/n}, \eta \rangle > 0\}} (\tilde{X}_{(i+1)/n} - \tilde{X}_{i/n}).$$

By Kallenberg (2002, Cor. 17.13) we have

$$\sup_{0\le t\le T} \|Y^{n+}_{[tn]} - Y^+_t\| \xrightarrow{\mathbb{P}} 0.$$

Consider further the increasing chains $A^{n\pm}$ obtained from X^n through the construction in (II.6), and let $\alpha^{n\pm}$ be the inverses. Note that $\frac{1}{n}A^{n+}_{[tn]} \to A^+_t$ for all $t\ge 0$ a.s. since the zero set of $\langle X\,\cdot\,, \eta\rangle$ is a Lebesgue null-set a.s. It follows that almost surely $\frac{1}{n}\alpha^{n+}_{[tn]} \to \alpha^+_t$ for all $t\in\mathcal{C}(\alpha^+)$, where $\mathcal{C}(\alpha^+)$ is the set of continuity points for α^+. To see this, observe first that

$$\inf\{s\ge 0 \mid \tfrac{1}{n}A^{n+}_{[sn]} > t\} = \inf\{s\ge 0 \mid A^{n+}_{[sn]} \ge [tn]+1\} = \frac{1}{n}\alpha^{n+}_{[tn]+1}.$$

Almost surely the expression on the left converges to α^+_t for all $t\in\mathcal{C}(\alpha^+)$. This basic convergence of right-continuous inverses is easy to prove (e.g. using the arguments in the proof by Resnick (1987, Prop. 0.1)). Lastly one verifies that $\frac{1}{n}\alpha^{n+}_{[tn]+1}$ can indeed by replaced by $\frac{1}{n}\alpha^{n+}_{[tn]}$.

Let $f\colon [0,\infty)\to\mathbb{R}^d$ be a continuous function. Then it follows from the observations above that

$$\frac{1}{n}\sum_{i=0}^{[Tn]} \langle f(i/n), X^{n\uparrow}_i\rangle \xrightarrow{\mathbb{P}} \int_0^T \langle f(s), X^\uparrow_s\rangle\,\mathrm{d}s,$$

where we make the convention that $\langle a, \dagger\rangle = \infty$ for $a\neq 0$ and $\langle 0,\dagger\rangle = 0$. To prove this we use the fact that α^+ is strictly increasing and has at most countably many discontinuities, with the former implying that Y^+ jumps at α^+_t for at most countably many t.

Similarly, if $g\colon [0,\infty)\to\mathbb{R}^d$ is a continuous function we obtain the convergence

$$\frac{1}{n}\sum_{i=0}^{[Tn]} \langle g(i/n), X^{n\downarrow}_i\rangle \xrightarrow{\mathbb{P}} \int_0^T \langle g(s), X^\downarrow_s\rangle\,\mathrm{d}s.$$

Using the fact that almost surely $(X_t)_{t\in[0,T]}$ reaches its infimum in the direction given by η exactly once, it follows that almost surely

$$\frac{1}{n}\sum_{i=0}^{[Tn]} \langle f(i/n), \underrightarrow{X}^n_i\rangle \to \int_0^T \langle f(s), \underrightarrow{X}_s\rangle\,\mathrm{d}s,$$

and

$$\frac{1}{n}\sum_{i=0}^{[Tn]} \langle g(i/n), \underleftarrow{X}^n_i\rangle \to \int_0^T \langle g(s), \underleftarrow{X}_s\rangle\,\mathrm{d}s,$$

where f and g are as above. By Theorem II.16 we obtain the distributional identity

$$\Big(\int_0^T \langle g(s), X^\downarrow_s\rangle\,\mathrm{d}s, \int_0^T \langle f(s), X^\uparrow_s\rangle\,\mathrm{d}s\Big) \overset{d}{=} \Big(-\int_0^T \langle g(s), \underleftarrow{X}_s\rangle\,\mathrm{d}s, \int_0^T \langle f(s), \underrightarrow{X}_s\rangle\,\mathrm{d}s\Big)$$

under $\mathbb{P}^T$, thus proving Theorem II.2. □

Acknowledgements

The authors gratefully acknowledge financial support of Sapere Aude Starting Grant 8049-00021B "Distributional Robustness in Assessment of Extreme Risk" from Independent Research Fund Denmark.

References

D. J. Aldous and G. K. Eagleson. On mixing and stability of limit theorems. *Ann. Probability*, 6(2):325–331, 1978. ISSN 0091-1798.

S. Asmussen and J. Ivanovs. Discretization error for a two-sided reflected Lévy process. *Queueing Syst.*, 89(1-2):199–212, 2018. ISSN 0257-0130.

S. Asmussen, P. Glynn, and J. Pitman. Discretization error in simulation of one-dimensional reflecting Brownian motion. *Ann. Appl. Probab.*, 5(4):875–896, 1995. ISSN 1050-5164.

J. Bertoin. Splitting at the infimum and excursions in half-lines for random walks and Lévy processes. *Stochastic Process. Appl.*, 47(1):17–35, 1993. ISSN 0304-4149.

J. Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996. ISBN 0-521-56243-0.

P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. ISBN 0-471-19745-9. A Wiley-Interscience Publication.

K. Bisewski and J. Ivanovs. Zooming-in on a Lévy process: failure to observe threshold exceedance over a dense grid. *Electron. J. Probab.*, 25:Paper No. 113, 33, 2020.

L. Chaumont. Sur certains processus de Lévy conditionnés à rester positifs. *Stochastics Stochastics Rep.*, 47(1-2):1–20, 1994. ISSN 1045-1129.

L. Chaumont and R. A. Doney. On Lévy processes conditioned to stay positive. *Electron. J. Probab.*, 10:no. 28, 948–961, 2005. ISSN 1083-6489.

L. Chaumont and R. A. Doney. Invariance principles for local times at the maximum of random walks and Lévy processes. *Ann. Probab.*, 38(4):1368–1389, 2010. ISSN 0091-1798.

L. Döring and P. Weissmann. Stable processes conditioned to hit an interval continuously from the outside. *Bernoulli*, 26(2):980–1015, 2020. ISSN 1350-7265.

L. Döring, A. R. Watson, and P. Weissmann. Lévy processes with finite variance conditioned to avoid an interval. *Electron. J. Probab.*, 24:Paper No. 55, 32, 2019.

T. Duquesne. Path decompositions for real Levy processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 39(2):339–370, 2003. ISSN 0246-0203.

W. Feller. *An introduction to probability theory and its applications. Vol. II.* John Wiley & Sons, Inc., New York-London-Sydney, second edition, 1971.

J. Ivanovs. Splitting and time reversal for Markov additive processes. *Stochastic Process. Appl.*, 127(8):2699–2724, 2017. ISSN 0304-4149.

J. Ivanovs. Zooming in on a Lévy process at its supremum. *Ann. Appl. Probab.*, 28(2): 912–940, 2018. ISSN 1050-5164.

J. Ivanovs and M. Podolskij. Optimal estimation of some random quantities of a Lévy process. *arXiv preprint arXiv:2001.02517*, 2020.

O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. ISBN 0-387-95313-2.

P. W. Millar. A path decomposition for Markov processes. *Ann. Probability*, 6(2):345–348, 1978. ISSN 0091-1798.

S. I. Resnick. *Extreme values, regular variation, and point processes*, volume 4 of *Applied Probability. A Series of the Applied Probability Trust.* Springer-Verlag, New York, 1987. ISBN 0-387-96481-9.

Paper III

Local behavior of diffusions at the supremum

Jakob D. Thøstesen
Department of Mathematics, Aarhus University

Abstract. This paper studies small-time behavior at the supremum of a diffusion process. For a solution to the SDE $\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t$ (where W is a standard Brownian motion) we consider $(\epsilon^{-1/2}(X_{m^X+\epsilon t} - \overline{X}))_{t\in\mathbb{R}}$ as $\epsilon \downarrow 0$, where $\overline{X}$ is the supremum of X on the time interval $[0,1]$ and m^X is the time of the supremum. It is shown that this process converges in law to a process $\hat{\xi}$, where $(\hat{\xi}_t)_{t\geq 0}$ and $(\hat{\xi}_{-t})_{t\geq 0}$ arise as independent Bessel-3 processes multiplied by $-\sigma(\overline{X})$. The proof is based on the fact that a continuous local martingale can be represented as a time-changed Brownian motion. This representation is also used to prove a limit theorem for zooming in on X at a fixed time. As an application of the zooming-in result at the supremum we consider estimation of the supremum $\overline{X}$ based on observations at equidistant times.

Keywords: Diffusion process, functional limit theorem, small-time behavior, stable convergence, discretization error, Bessel process

2020 MSC: 60J60; 60F17

III.1 Introduction

Differentiation is a central concept in classical analysis and it is useful in many areas with one example being approximation. When dealing with stochastic processes, however, we rarely care about differentiation as the paths of many typical processes are differentiable at few (if any) points. This means that there is a need for a similar tool to handle the local behavior of such processes.

A differentiation-type concept for stochastic processes was introduced in Asmussen et al. (1995) with the purpose of describing local behavior at the supremum of the Brownian motion. This concept was revisited in Ivanovs (2018) where it was called zooming in. A stochastic process X starting at zero is said to satisfy the *zooming-in condition* if

$$(a_\epsilon X_{\epsilon t})_{t\geq 0} \xrightarrow{fdd} (\tilde{X}_t)_{t\geq 0} \quad \text{as } \epsilon \downarrow 0, \tag{III.1}$$

where a_ϵ is a scaling function and $\tilde{X}$ is a non-trivial stochastic process. It is clear that this is connected to differentiation (from the right) at time 0. Indeed, if $t \mapsto X_t$ is differentiable from the right at 0 then the convergence holds with $a_\epsilon = \epsilon^{-1}$ and $\tilde{X}$ being a line.

The related concept of zooming out was studied in Lamperti (1962). While this sounds like quite a different framework it is in fact possible to transfer many of ideas to the zooming-in setting. This includes the study of the scaling function and the limit process. For more details see Ivanovs (2018).

The zooming-in condition has proven to be a very useful regularity assumption in e.g. Bisewski and Ivanovs (2020), Ivanovs and Podolskij (2020) and Ivanovs and Thøstesen (2021). In those papers the zooming-in theory plays a large role in various discretization problems.

Naturally there is a big difference between zooming in at a fixed time and at a random time. With X being a Lévy process satisfying the zooming-in assumption it was shown in Ivanovs (2018) that one may also zoom in at the supremum of X over the interval $[0, 1]$. The scaling is again a_ϵ and the law of the limit process is related to $\tilde{X}$. This theory was used in Ivanovs and Podolskij (2020) to derive limit theorems related to estimation of the supremum of X in a high-frequency setting, and it was used in Bisewski and Ivanovs (2020) to study threshold exceedance for Lévy processes.

This paper presents limit results for zooming in at a fixed time and at the supremum of a diffusion process. Estimation of the supremum is studied as an application of the limit theory. The approach is based on the fact that a continuous local martingale can be represented as a time-changed Brownian motion. For zooming in at the supremum this lets us build on an existing zooming-in result for the Brownian motion.

All relevant definitions and prerequisites are contained in §III.2. In §III.3 the main results are presented. Generality of the results and possible extensions are covered in §III.4, and finally the most technical proofs are found in §III.5.

III.2 Definitions and prerequisites

III.2.1 The setup

Consider the SDE

$$\mathrm{d}X_t = \mu(X_t)\,\mathrm{d}t + \sigma(X_t)\,\mathrm{d}W_t \quad \text{and} \quad X_0 = x_0, \tag{III.2}$$

where W is a standard Brownian motion. We assume that there exists a weak solution (X, W) to (III.2), defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ such that X is $(\mathcal{F}_t)$-adapted and W is an $(\mathcal{F}_t)$-Brownian motion. We assume that $(\mathcal{F}_t)$ satisfies the usual conditions. In this paper we will encounter several $(\mathcal{F}_t)$-adapted processes which are almost surely continuous, X and W being the first examples. Since $(\mathcal{F}_t)$ is complete we may and will assume that these processes are continuous for all $\omega \in \Omega$.

We need some regularity assumptions on μ and σ which are stated in Assumption III.A below. Here, the range of X is the set of points $x \in \mathbb{R}$ for which $\mathbb{P}(X_t = x \text{ for some } t \in [0, \infty)) > 0$. Note that the positivity in assumption (ii) is quite standard and guarantees the presence of some amount of noise at any time. This is important for zooming in since the presence of a Brownian motion affects the scaling function. For example, if X is a Brownian motion plus a linear drift then $a_\epsilon \sim c_1\epsilon^{-1/2}$ (for some $c_1 > 0$), and if X is just a linear drift then $a_\epsilon \sim c_2\epsilon^{-1}$ (for some $c_2 > 0$), see Ivanovs (2018, Thm. 2).

Assumption III.A.

(i) *The function* $\mu\colon \mathbb{R} \to \mathbb{R}$ *is locally bounded.*

(ii) *The function* $\sigma\colon \mathbb{R} \to [0, \infty)$ *is continuous and strictly positive on the range of* X*.*

We let $\overline{X} := \sup_{t\in[0,1]} X_t$ denote the supremum of X over the unit interval, and we denote the time of the ultimate supremum by $m^X := \sup\{t \in [0, 1] \mid X_t = \overline{X}\}$. We then define the pre- and post-supremum processes, $\underleftarrow{X}$ and $\underrightarrow{X}$, by

$$\underleftarrow{X}_t := \begin{cases} X_{m^X - t} - \overline{X} & \text{if } 0 \le t < m^X, \\ \dagger & \text{if } t \ge m^X, \end{cases}$$
$$\underrightarrow{X}_t := \begin{cases} X_{m^X + t} - \overline{X} & \text{if } 0 \le t < 1 - m^X, \\ \dagger & \text{if } t \ge 1 - m^X. \end{cases}$$

III.2.2 Path space and topology

The processes appearing in this paper are viewed as random variables taking values in the measurable space $(D[0,\infty), \mathcal{D})$, where $D[0,\infty)$ is the space of real-valued càdlàg functions defined on $[0,\infty)$ and $\mathcal{D}$ is the Borel σ-algebra induced by the Skorokhod topology. A standard reference treating this space is Billingsley (1999, §16).

For convergence in distribution it is often sufficient to consider the restrictions of processes to intervals of the form $[0,T]$ for $T > 0$. Consider $D[0,\infty)$-valued random variables (i.e. stochastic processes) $X, X^1, X^2, \dots$. Then $X^n \xrightarrow{d} X$ if and only if $(X^n_t)_{t\in[0,T]} \xrightarrow{d} (X_t)_{t\in[0,T]}$ for all $T > 0$ where X is almost surely continuous at T, see e.g. Billingsley (1999, Thm. 16.7). Here the restrictions are seen as random variables in $D[0,T]$ (the space of càdlàg functions on $[0,T]$).

III.2.3 The central representation

Suppose for a moment that X solves the SDE (III.2) with $x_0 = 0$ and $\mu \equiv 0$. Then X is a continuous local $(\mathcal{F}_t)$-martingale starting at zero. We denote the quadratic variation of X by $[X]$ and recall that it is almost surely given by

$$[X]_t = \int_0^t \sigma^2(X_s)\,\mathrm{d}s, \qquad t \ge 0.$$

Note that $[X]$ is continuous and strictly increasing and let $\tau_t := \inf\{s \ge 0 \mid [X]_s > t\}$ denote its (right-continuous) inverse. We define a new filtration $(\mathcal{G}_t)$ by $\mathcal{G}_t := \mathcal{F}_{\tau_t}$. A standard result (Kallenberg, 2021, Thm. 19.4) gives the existence of a Brownian motion $\tilde{W}$ with respect to a standard extension $(\hat{\mathcal{G}}_t)$ of $(\mathcal{G}_t)$ (Kallenberg, 2021, p. 420) such that $X = (\tilde{W}_{[X]_t})_{t\ge 0}$ a.s. Furthermore, for any $s \ge 0$ the random variable $[X]_s$ is a $(\mathcal{G}_t)_{t\ge 0}$-stopping time.

III.2.4 Stable convergence

A central concept in this paper is the notion of *stable convergence* which was originally introduced by Rényi (1963). Later papers which are also of interest include Aldous and Eagleson (1978) and Podolskij and Vetter (2010). In this subsection we present only the results which are relevant for this paper.

We consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a sequence of random variables (X_n) taking values in some Polish space. We say that X_n converges stably to X (written $X_n \xrightarrow{st} X$) defined on an extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ of the space if

$$\mathbb{E}[f(X_n)Z] \to \tilde{\mathbb{E}}[f(X)Z] \tag{III.3}$$

for all bounded continuous functions f and all bounded $\mathcal{F}$-measurable Z.

The extension of $(\Omega, \mathcal{F}, \mathbb{P})$ is a product space $(\tilde{\Omega}, \tilde{\mathcal{F}}) = (\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}')$ equipped with a probability measure $\tilde{\mathbb{P}}$ which satisfies $\tilde{\mathbb{P}}(A \times \Omega') = \mathbb{P}(A)$ for any $A \in \mathcal{F}$. A random variable Z defined on $(\Omega, \mathcal{F}, \mathbb{P})$ becomes a random variable on the extension by defining $Z(\omega, \omega') := Z(\omega)$. We often need the extension to support a random variable X which is independent of $\mathcal{F}$. In that case we let $(\Omega', \mathcal{F}', \mathbb{P}')$ be a probability space on which X can be defined. As before X can be viewed as a random variable on $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}')$, and taking $\tilde{\mathbb{P}} = \mathbb{P} \otimes \mathbb{P}'$ gives the desired independence. In this case, and when $X_n \xrightarrow{st} X$, we sometimes say that the convergence is *mixing*. This concept was first introduced in Rényi (1958).

Proving the main results of this paper we will require a few lemmas about stable convergence. These can be found in §III.5.1.

III.3 Main results

III.3.1 Zooming in at a fixed time

We begin with a limit theorem that formalizes the intuitive understanding of a diffusion process. Namely that the local behavior of X at a fixed time $T > 0$ is that of a scaled Brownian motion. To simplify we consider the time point $T = 1$.

For $\epsilon > 0$ and $t \in \mathbb{R}$ we let $X_t^{(\epsilon)} := \epsilon^{-1/2}(X_{1+\epsilon t} - X_1)$. Consider further two standard Brownian motions $U^{(1)}$ and $U^{(2)}$ defined on an extension of $(\Omega, \mathcal{F}, \mathbb{P})$ which are independent of each other and of $\mathcal{F}$.

Theorem III.1. *It holds that*

$$\big(l(X_{-t}^{(\epsilon)})_{t\geq 0}, (X_t^{(\epsilon)})_{t\geq 0}\big) \xrightarrow{st} \big(\sigma(X_1)U^{(1)}, \sigma(X_1)U^{(2)}\big) \quad \text{as } \epsilon \downarrow 0.$$

Dealing with $(X_t^{(\epsilon)})_{t\geq 0}$ is fairly simple as we look forward in time. Looking backwards in time is generally harder and proving the convergence of $(X_{-t}^{(\epsilon)})_{t\geq 0}$ is indeed somewhat technical. The proof of Theorem III.1 is deferred to §III.5.2.

Looking backwards in time may be difficult but it is quite useful. The following result is very intuitive in addition to being necessary for proving Theorem III.4 below, and proving it is now trivial.

Corollary III.2. *Almost surely $m^X \neq 1$.*

Proof. Let $A \subseteq D[0, \infty)$ be the set of functions f in $D[0, \infty)$ with $f(t) \leq 0$ for all $t \in [0, 1)$. Using Billingsley (1999, Thm. 16.1) it is easy to verify that A is closed in the Skorokhod topology. It follows from Theorem III.1 and the Portmanteau theorem that

$$\mathbb{P}(m^X = 1) \leq \limsup_{\epsilon\downarrow 0} \mathbb{P}\big((X_{-t}^{(\epsilon)})_{t\geq 0} \in A\big) \leq \tilde{\mathbb{P}}\big((\sigma(X_1)U_t^{(1)})_{t\geq 0} \in A\big) = 0.$$

□

Remark III.3. With a few straight-forward modifications to the proof we may extend Theorem III.1 to cover the case where X is given by the more general SDE $\mathrm{d}X_t = \mu_t \,\mathrm{d}t + \sigma_t \,\mathrm{d}W_t$. In this case we assume that the process μ is locally bounded and that σ is continuous.

III.3.2 Zooming in at the supremum

The local behavior of X at time 1 is described by the zooming-in result in Theorem III.1. In a similar fashion we want to describe the local behavior at the supremum through a zooming-in result. It is well-known (Bertoin, 1993) that the negated pre- and post-supremum processes for a Brownian motion are two independent Bessel-3 processes (killed at certain random times). With this in mind the following result is somewhat intuitive.

Theorem III.4. *Let $B^{(1)}$ and $B^{(2)}$ be two independent Bessel-3 processes defined on an extension of $(\Omega, \mathcal{F}, \mathbb{P})$ such that both processes are independent of $\mathcal{F}$. Then it holds that*

$$\big((\epsilon^{-1/2}\underleftarrow{X}_{\epsilon t})_{t\geq 0}, (\epsilon^{-1/2}\underrightarrow{X}_{\epsilon t})_{t\geq 0}\big) \xrightarrow{st} \big(-\sigma(\overline{X})B^{(1)}, -\sigma(\overline{X})B^{(2)}\big) \quad \textit{as } \epsilon \downarrow 0. \tag{III.4}$$

The proof of Theorem III.4 is deferred to §III.5.3.

III.3.3 Estimation of the supremum

As an application of Theorem III.4 we consider a high-frequency setting in which the process X is observed on the set of times $\epsilon(\mathbb{N}_0 + U) \cap [0, 1]$ for some small $\epsilon > 0$, where U is a standard uniform defined on an extension of the space such that it is independent of $\mathcal{F}$ and $B^{(1)}, B^{(2)}$. The objective is to estimate the supremum $\overline{X}$ over $[0, 1]$. To avoid constantly having to intersect with the unit interval we consider X as being restricted to this interval.

We take the basic estimator $M^{(\epsilon)} := \sup_{t\in\epsilon(\mathbb{N}_0+U)} X_t$. The following result establishes the convergence rate $\epsilon^{-1/2}$.

Proposition III.5. *For all $\epsilon > 0$ it holds that*

$$0 \geq \epsilon^{-1/2}(M^{(\epsilon)} - \overline{X}) \geq \epsilon^{-1/2}\underrightarrow{X}_{\epsilon\{U-m^X/\epsilon\}},$$

where $\{U-m^X/\epsilon\}$ is the fractional part of $U-m^X/\epsilon$. Furthermore, there is stable convergence of the lower bound:

$$\epsilon^{-1/2}\underrightarrow{X}_{\epsilon\{U-m^X/\epsilon\}} \xrightarrow{st} -\sigma(\overline{X})B^{(2)}_U.$$

Proof. Observe that

$$\epsilon^{-1/2}(M^{(\epsilon)} - \overline{X}) = \sup_{i\in\mathbb{N}_0} \epsilon^{-1/2}(X_{\epsilon(i+U)} - \overline{X}) = \sup_{i\in\mathbb{Z}} \epsilon^{-1/2}(X_{\epsilon(i+\{U-m^X/\epsilon\})+m^X} - \overline{X})$$

for all $\epsilon > 0$. We can get a lower bound by taking a specific i instead of taking the supremum over $\mathbb{Z}$. With $i = 0$ we get the claimed lower bound.

By conditioning one sees that for all $\epsilon > 0$ the fractional part $U_\epsilon := \{U - m^X/\epsilon\}$ is a standard uniform independent of $\mathcal{F}$ and $B^{(1)}, B^{(2)}$. In combination with Theorem III.4 and Whitt (2002, Prop. 13.2.1) we obtain the convergence of the lower bound. □

Remark III.6. The lower bound in Proposition III.5 is somewhat conservative. Indeed, in the proof we see that the discretization error can be written as $\sup_{i\in\mathbb{Z}} \epsilon^{-1/2}(X_{\epsilon(i+\{U-m^X/\epsilon\})+m^X} - \overline{X})$. Looking to Theorem III.4 it is expected that this quantity will converge to $\sup_{i\in\mathbb{Z}} \hat{\xi}_{i+U}$, where $\hat{\xi}_t = -\sigma(\overline{X})B^{(1)}_{-t}$ for $t < 0$ and $\hat{\xi}_t = -\sigma(\overline{X})B^{(2)}_t$ for $t \geq 0$. However, this is not straight-forward to prove. The issue is that taking the supremum over an unbounded set of times is not continuous. This was solved by Bisewski and Ivanovs (2020, App. B) where the authors corrected the proof by Ivanovs (2018, Thm. 5). In those papers X is a Lévy process satisfying the zooming-in condition. The approach is not directly applicable here because it is based on results known only for Lévy processes.

It is perfectly valid to ask why we choose to sample at times $\epsilon(i + U)$ rather than ϵi for $i \in \mathbb{N}_0$. In the latter case one would consider the estimator $\tilde{M}^{(\epsilon)} := \sup_{t\in\epsilon\mathbb{N}_0} X_t$. For this estimator it holds that

$$\epsilon^{-1/2}(\tilde{M}^{(\epsilon)} - \overline{X}) = \sup_{i\in\mathbb{Z}} \epsilon^{-1/2}(X_{\epsilon(i+\{-m^X/\epsilon\})+m^X} - \overline{X})$$

for any $\epsilon > 0$. This gives the lower bound $\epsilon^{-1/2}\underrightarrow{X}_{\epsilon\{-m^X/\epsilon\}}$. In order to obtain a limit theorem for this quantity we need to know what happens to $\{-m^X/\epsilon\}$ as $\epsilon \downarrow 0$. By the classical result of Kosulajeff (1937) it is known that $\{-m^X/\epsilon\}$ converges to the standard uniform distribution if m^X has a density wrt. the Lebesgue measure. As seen in Proposition III.5 we are able to avoid such considerations by translating the sampling times by ϵU.

III.4 Further comments

III.4.1 Generality of the results

Theorem III.1 describes zooming in at time 1. Naturally there is nothing special about the time 1 so the result also holds if we zoom in at some other fixed time $T > 0$. In that case one simply replaces $\sigma(X_1)$ by $\sigma(X_T)$ in the limit. The time point 1 is chosen only to simplify notation.

In the same way there is nothing special about the time interval $[0, 1]$ in the formulation of Theorem III.4. This interval can be replaced by $[T_1, T_2]$ where $0 \le T_1 < T_2 < \infty$ are fixed. In the formulation of the result one will then have to define $\overline{X} := \sup_{t\in[T_1,T_2]} X_t$.

III.4.2 Extending to other processes

The approach used to prove Theorem III.1 and Theorem III.4 is based on representing the local martingale part of X as a time-changed Brownian motion. The time-change is differentiable and this lets us apply zooming-in results for the Brownian motion to obtain corresponding results for X.

It is possible to extend the result about zooming in at the supremum to other classes of stochastic processes. In Ivanovs (2018) this was done for any Lévy process satisfying the zooming-in condition (III.1). With the approach used to prove Theorem III.4 it is likely that this result can be used to prove limit results for zooming in at the supremum of time-changed Lévy processes. Below are two examples where this appears to be do-able.

Example A: Let X be a positive $1/\alpha$-self-similar Markov process (pssMp) starting at some value $x > 0$. The classical result of Lamperti (1972) tells us that there exists a Lévy process ξ such that

$$X_t = x \exp(\xi_{\tau(tx^{-\alpha})}), \qquad t \ge 0,$$

where $\tau(tx^{-\alpha}) = \inf\{s > 0 \,|\, \int_0^s \exp(\alpha\xi_u)\,\mathrm{d}u \ge tx^{-\alpha}\}$. The key point is that X_t is obtained by time-changing a Lévy process and applying a strictly increasing and differentiable function. Note also that the time-change is differentiable. The last ingredient is that ξ must satisfy the zooming-in condition. This is completely characterized in Ivanovs (2018, Thm. 2) in terms of the characteristics of ξ. Note also that one must pay special attention to a possible jump at the time of supremum.

Example B: Let X be a continuous-state branching process. Then there exists (Kyprianou, 2006, Thm. 10.2) a Lévy process ζ such that

$$X_t = \zeta_{\theta(t)\wedge\tau_0^-}, \qquad t \ge 0,$$

where $\tau_0^- = \inf\{s > 0 \mid \zeta_s < 0\}$ and $\theta(t) = \inf\{s > 0 \mid \int_0^s \zeta_u^{-1}\,\mathrm{d}u > t\}$. We note that the time-change is not as well-behaved as for the class of pssMps. For example, $t \mapsto \theta(t)$ is not differentiable everywhere. As a consequence one will again have to be particularly aware of any jump at the supremum.

III.5 Proofs

III.5.1 Useful lemmas

This subsection contains a few results which are useful when working with stable convergence.

Lemma III.7. *Assume that* $X_n \xrightarrow{st} X$. *Then we have the following:*

(i) *If* $Y, Y_1, Y_2, \ldots$ *are random variables (taking values in some Polish space) on* $(\Omega, \mathcal{F}, \mathbb{P})$ *and* $Y_n \xrightarrow{\mathbb{P}} Y$, *then* $(X_n, Y_n) \xrightarrow{st} (X, Y)$.

(ii) *If* g *is a Borel-measurable function taking values in a Polish space and* g *is almost surely continuous at* X *then* $g(X_n) \xrightarrow{st} g(X)$.

Proof. See e.g. Häusler and Luschgy (2015, Thm. 3.18). □

If $\mathcal{H} \subseteq \mathcal{F}$ is a sub-σ-algebra and (III.3) is only known to hold for $\mathcal{H}$-measurable Z we say that X_n converges $\mathcal{H}$-stably to X (written $X_n \xrightarrow{\mathcal{H}-st} X$). The following basic lemma shows that sometimes stable convergence can be obtained just by proving $\mathcal{H}$-stable convergence for a suitable sub-σ-algebra $\mathcal{H}$. This trick is used in e.g. the proof of Jacod and Protter (2012, Thm. 4.3.1).

Lemma III.8. *Let* $\mathcal{H} \subseteq \mathcal{F}$ *be a sub-*σ*-algebra. Assume that each* X_n *is* $\mathcal{H}$*-measurable,* X *is independent of* $\mathcal{F}$ *and* $X_n \xrightarrow{\mathcal{H}-st} X$. *Then* $X_n \xrightarrow{st} X$.

Proof. We must verify (III.3) for all bounded continuous functions f and all bounded $\mathcal{F}$-measurable Z. Since X_n is $\mathcal{H}$-measurable and $X_n \xrightarrow{\mathcal{H}-st} X$ it holds that

$$\mathbb{E}[f(X_n)Z] = \mathbb{E}[f(X_n)\mathbb{E}[Z \mid \mathcal{H}]] \to \tilde{\mathbb{E}}[f(X)\mathbb{E}[Z \mid \mathcal{H}]].$$

Finally the assumed independence yields

$$\tilde{\mathbb{E}}[f(X)\mathbb{E}[Z \mid \mathcal{H}]] = \tilde{\mathbb{E}}[f(X)]\tilde{\mathbb{E}}[Z] = \tilde{\mathbb{E}}[f(X)Z].$$

□

It is often useful to work with equivalent definitions of stable convergence.

Lemma III.9. *For a sub-*σ*-algebra* $\mathcal{H} \subseteq \mathcal{F}$ *the following statements are equivalent:*

(i) $X_n \xrightarrow{\mathcal{H}-st} X$.

(ii) $(X_n, Y) \xrightarrow{\mathcal{H}-st} (X, Y)$ *for any* $\mathcal{H}$*-measurable* Y *taking values in some Polish space.*

(iii) $(X_n, Y) \xrightarrow{d} (X, Y)$ *for any* $\mathcal{H}$*-measurable* Y *taking values in some Polish space.*

(iv) $(X_n, \mathbf{1}_F) \xrightarrow{d} (X, \mathbf{1}_F)$ *for any* $F \in \mathcal{E}$, *where* $\mathcal{E} \subseteq \mathcal{H}$ *is closed under finite intersections and further satisfies* $\Omega \in \mathcal{E}$ *and* $\sigma(\mathcal{E}) = \mathcal{H}$.

Proof. For equivalence of (i)–(iii) see Podolskij and Vetter (2010, Prop. 1), and for equivalence of (i) and (iv) see Häusler and Luschgy (2015, Thm. 3.17). □

Independence plays a large role for convergence of joint distributions. The following lemma shows that joint stable convergence can also be obtained under certain independence assumptions.

Lemma III.10. *Let $(X_n), (Y_n)$ be independent sequences of random variables, and let X, Y be independent random variables such that X and Y are independent of $\mathcal{F}$, $X_n \xrightarrow{st} X$ and $Y_n \xrightarrow{st} Y$. Then $(X_n, Y_n) \xrightarrow{st} (X, Y)$.*

Proof. Let $\mathcal{A} = \sigma(\{X_n \mid n \in \mathbb{N}\})$, $\mathcal{B} = \sigma(\{Y_n \mid n \in \mathbb{N}\})$ and $\mathcal{H} = \sigma(\mathcal{A} \cup \mathcal{B})$. According to Lemma III.8 it is sufficient to prove $\mathcal{H}$-stable convergence. For $A \in \mathcal{A}$ and $B \in \mathcal{B}$ we see that

$$(X_n, \mathbf{1}_A, Y_n, \mathbf{1}_B) \xrightarrow{d} (X, \mathbf{1}_A, Y, \mathbf{1}_B)$$

due to the assumed independence. Hence, $(X_n, Y_n, \mathbf{1}_{A\cap B}) \xrightarrow{d} (X, Y, \mathbf{1}_{A\cap B})$. The $\mathcal{H}$-stable convergence follows since condition (iv) in Lemma III.9 is satisfied with $\mathcal{E}$ being the collection of sets on the form $A \cap B$ where $A \in \mathcal{A}$ and $B \in \mathcal{B}$. □

III.5.2 Proof of Theorem III.1

As in the formulation of Theorem III.1 we let $(U^{(1)}, U^{(2)})$ denote a pair of independent standard Brownian motions, defined on an extension of $(\Omega, \mathcal{F}, \mathbb{P})$ such that they are also independent of $\mathcal{F}$.

We may write X_t as

$$X_t = x_0 + A_t + M_t, \qquad t \geq 0,$$

where A is a continuous and $(\mathcal{F}_t)$-adapted process with bounded variation, M is a continuous $(\mathcal{F}_t)$-local martingale and $A_0 = M_0 = 0$ a.s. We see that

$$X_t^{(\epsilon)} = \epsilon^{-1/2}(X_{1+\epsilon t} - X_1) = \epsilon^{-1/2}(A_{1+\epsilon t} - A_1) + \epsilon^{-1/2}(M_{1+\epsilon t} - M_1)$$

for all $t \geq -1/\epsilon$. We treat each term from the right-hand side separately.

Note that $A_t = \int_0^t \mu(X_s)\,\mathrm{d}s$ for all $t \geq 0$ a.s. Since μ and X are both locally bounded we immediately find that

$$\sup_{t\in[-T,T]} \epsilon^{-1/2}|A_{1+\epsilon t} - A_1| \leq 2T\epsilon^{1/2} \sup_{t\in[1-\epsilon T, 1+\epsilon T]} |\mu(X_t)| \to 0$$

a.s. as $\epsilon \downarrow 0$ for any $T > 0$.

Below in the proof of Theorem III.4 it is necessary to deal with the drift differently. The same approach could be used here, however it is the author's belief that the calculation above is more illustrative since it clearly shows that the drift vanishes due to the $\epsilon^{-1/2}$ scaling.

It remains to show that

$$\begin{aligned}\big((\epsilon^{-1/2}(M_{1-\epsilon t} - M_1))_{t\geq 0}, (\epsilon^{-1/2}(M_{1+\epsilon t} - M_1))_{t\geq 0}\big) \\ \xrightarrow{st} \big(\sigma(X_1)U^{(1)}, \sigma(X_1)U^{(2)}\big).\end{aligned} \tag{III.5}$$

To do so we will represent M as a time-changed Brownian motion. Let $(\mathcal{F}_t^M)$ denote the completed natural filtration generated by M, let τ denote the inverse of $[M]$, and define $\mathcal{G}_t^M = \mathcal{F}_{\tau_t}^M$. Now, as in §III.2.3 a standard result gives the existence of a Brownian motion $\tilde{W}$ with respect to a standard extension $(\hat{\mathcal{G}}_t^M)$ of $(\mathcal{G}_t^M)$ such that $M = (\tilde{W}_{[M]_t})_{t\geq 0}$ a.s. Recall that $[M]_s$ is a $(\mathcal{G}_t^M)$-stopping time for any $s \geq 0$. Finally we note that the quadratic variation of M is given by

$$[M]_t = [X]_t = \int_0^t \sigma^2(X_s)\,\mathrm{d}s, \qquad t \geq 0,$$

almost surely.

The next step in the proof of Theorem III.1 is Lemma III.11 below which allows for zooming in on $\tilde{W}$ from the right. Instead of simply zooming in at time 1 we generalize to zooming in at $1-\epsilon R$ with $R \geq 0$ since we will need this in the proof of Lemma III.12 below. This slight generalization requires very little extra effort.

Lemma III.11. *For any $R \geq 0$ it holds that*

$$\big(\epsilon^{-1/2}(\tilde{W}_{[M]_{1-\epsilon R}+\epsilon t} - \tilde{W}_{[M]_{1-\epsilon R}})\big)_{t\geq 0} \xrightarrow{st} U, \tag{III.6}$$

where U is a standard Brownian motion defined on an extension of $(\Omega, \mathcal{F}, \mathbb{P})$ such that U is independent of $\mathcal{F}$.

Proof. We fix $R \geq 0$ and recall that $[M]_{1-\epsilon R}$ is a $(\mathcal{G}_t^M)$-stopping time. It follows that the left-hand side of (III.6) is a standard Brownian motion for any $\epsilon > 0$ so the convergence in distribution is trivial.

Now, let $\mathcal{A}$ denote the σ-algebra generated by the process $\tilde{W}' := (\tilde{W}_{[M]_1+t} - \tilde{W}_{[M]_1})_{t\geq 0}$. The first step is proving $\mathcal{A}$-stable convergence. It is sufficient to show that

$$\begin{aligned} &\big((\epsilon^{-1/2}(\tilde{W}_{[M]_{1-\epsilon R}+\epsilon t} - \tilde{W}_{[M]_{1-\epsilon R}}))_{t\in[0,T]}, (\tilde{W}'_{t_i})_{i=1,\ldots,k}\big) \\ &\quad \xrightarrow{d} \big((U_t)_{t\in[0,T]}, (\tilde{W}'_{t_i})_{i=1,\ldots,k}\big) \end{aligned} \tag{III.7}$$

for any $T > 0$, $k \in \mathbb{N}$ and $0 < t_1 < \ldots < t_k$. To this end define $a_i^{(\epsilon)} := \tilde{W}_{[M]_1+t_i} - \tilde{W}_{[M]_1+\epsilon T}$ and $b_i^{(\epsilon)} := \tilde{W}_{[M]_1+\epsilon T} - \tilde{W}_{[M]_1}$. Then $\tilde{W}'_{t_i} = a_i^{(\epsilon)} + b_i^{(\epsilon)}$, $b_i^{(\epsilon)} \to 0$ a.s. as $\epsilon \downarrow 0$, and for $\epsilon \in (0, t_1/T)$ we see that $a_i^{(\epsilon)}$ is independent of $(\tilde{W}_t)_{t\in[0,[M]_{1-\epsilon R}+\epsilon T]}$. Hence,

$$\begin{aligned} &\big((\epsilon^{-1/2}(\tilde{W}_{[M]_{1-\epsilon R}+\epsilon t} - \tilde{W}_{[M]_{1-\epsilon R}}))_{t\in[0,T]}, (a_i^{(\epsilon)})_{i=1,\ldots,k}\big) \\ &\quad \xrightarrow{d} \big((U_t)_{t\in[0,T]}, (\tilde{W}'_{t_i})_{i=1,\ldots,k}\big). \end{aligned}$$

The convergence in (III.7) follows immediately. This establishes (III.6) with $\xrightarrow{st}$ replaced by $\xrightarrow{\mathcal{A}-st}$.

We let $\mathcal{H} := \sigma(\mathcal{G}^M_{[M]_1} \cup \mathcal{A}) = \sigma(\mathcal{F}^M_1 \cup \mathcal{A})$ and note that the left-hand side in (III.6) is $\mathcal{H}$-measurable. Thus, proving $\mathcal{H}$-stable convergence automatically yields $\mathcal{F}$-stable convergence by Lemma III.8. We note that $\mathcal{F}^M_1 = \sigma(\bigcup_{\delta>0} \mathcal{F}^M_{1-\delta})$ since $M(\omega)$ is continuous for all $\omega \in \Omega$ (recall the considerations in the beginning of §III.2.1). According to Lemma III.9 it is sufficient to show that

$$\big((\epsilon^{-1/2}(\tilde{W}_{[M]_{1-\epsilon R}+\epsilon t} - \tilde{W}_{[M]_{1-\epsilon R}}))_{t\geq 0}, \mathbf{1}_A, \mathbf{1}_F\big) \xrightarrow{d} (U, \mathbf{1}_A, \mathbf{1}_F)$$

for any $\delta > 0$, $F \in \mathcal{F}^M_{1-\delta}$ and $A \in \mathcal{A}$. Since the first two components on the left-hand side are independent of $\mathbf{1}_F$ for small enough ϵ this is a trivial consequence of the $\mathcal{A}$-stable convergence. This concludes the proof. □

We proceed by proving the following lemma, stating that we can zoom in on $\tilde{W}$ at time $[M]_1$. The proof follows the same strategy as the proof of Ivanovs and Thøstesen (2021, Thm. 3).

Lemma III.12. *As $\epsilon \downarrow 0$ it holds that*

$$\big((\tilde{W}^{(\epsilon)}_{-t})_{t\geq 0}, (\tilde{W}^{(\epsilon)}_{t})_{t\geq 0}\big) \xrightarrow{st} (U^{(1)}, U^{(2)}), \tag{III.8}$$

where $\tilde{W}^{(\epsilon)}_t := \epsilon^{-1/2}(\tilde{W}_{[M]_1+\epsilon t} - \tilde{W}_{[M]_1})$.

Proof. There are two immediate things to note. Firstly, the convergence $(\tilde{W}_t^{(\epsilon)})_{t\geq 0} \xrightarrow{st} U^{(2)}$ is nothing more than the case $R = 0$ in Lemma III.11. Secondly, since $(\tilde{W}_{-t}^{(\epsilon)})_{t\geq 0}$ and $(\tilde{W}_t^{(\epsilon)})_{t\geq 0}$ are independent for all $\epsilon > 0$ it is sufficient, according to Lemma III.10, to show that the former converges stably to $U^{(1)}$. Again it is sufficient to show stable convergence of the process restricted to the time interval $[0, T]$ for all $T > 0$.

For any $R \geq 0$ we have the almost sure convergence $\epsilon^{-1}([M]_1 - [M]_{1-\epsilon R}) \to R\sigma^2(X_1) =: s$. Given $T > 0$ we can pick R such that $s > T$ with probability arbitrarily close to 1. With $Y_t^{(\epsilon)} = \epsilon^{-1/2}(\tilde{W}_{[M]_{1-\epsilon R}+\epsilon t} - \tilde{W}_{[M]_{1-\epsilon R}})$ we then write

$$\epsilon^{-1/2}(\tilde{W}_{[M]_1-\epsilon t} - \tilde{W}_{[M]_1}) = -(Y^{(\epsilon)}_{\epsilon^{-1}([M]_1-[M]_{1-\epsilon R})} - Y^{(\epsilon)}_{\epsilon^{-1}([M]_1-[M]_{1-\epsilon R}-\epsilon t)}). \quad \text{(III.9)}$$

That is, on $\{s > T\}$ the increment of $\epsilon^{-1/2}\tilde{W}$ over $[[M]_1 - \epsilon t, [M]_1]$ can be viewed as the increment of $Y^{(\epsilon)}$ over $[\epsilon^{-1}([M]_1 - [M]_{1-\epsilon R} - \epsilon t), \epsilon^{-1}([M]_1 - [M]_{1-\epsilon R})]$ (for small enough $\epsilon > 0$).

Almost surely $\epsilon^{-1}([M]_1 - [M]_{1-\epsilon R} - \epsilon t) \to s - t$ uniformly for $t \in [0, T]$. By combining this with (III.9), Lemma III.11, continuity of subordination (Whitt, 2002, Thm. 13.2.2) and Lemma III.7 we find that

$$\mathbb{E}[\mathbf{1}_{\{s>T\}} f((\tilde{W}_{-t}^{(\epsilon)})_{t\in[0,T]})Z] \to \tilde{\mathbb{E}}[\mathbf{1}_{\{s>T\}} f(-(U_s - U_{s-t})_{t\in[0,T]})Z] \quad \text{(III.10)}$$

for all bounded continuous f and all bounded $\mathcal{F}$-measurable Z, where U is a standard Brownian motion defined on an extension of $(\Omega, \mathcal{F}, \mathbb{P})$ such that U is independent of $\mathcal{F}$ and independent of $U^{(2)}$. We conclude by noting that the limit in (III.10) is equal to $\tilde{\mathbb{E}}[\mathbf{1}_{\{s>T\}} f((U_t^{(1)})_{t\in[0,T]})Z]$, where $U^{(1)}$ is a standard Brownian motion defined on an extension of $(\Omega, \mathcal{F}, \mathbb{P})$, independent of $\mathcal{F}$ and independent of $U^{(2)}$. □

We are now ready to finish the proof of Theorem III.1 which we have reduced to proving the convergence

$$\big((M_{-t}^{(\epsilon)})_{t\geq 0}, (M_t^{(\epsilon)})_{t\geq 0}\big) \xrightarrow{st} \big(\sigma(X_1)U^{(1)}, \sigma(X_1)U^{(2)}\big),$$

where $M_t^{(\epsilon)} := \epsilon^{-1/2}(M_{1+\epsilon t} - M_1)$.

Firstly, we have the almost sure convergence

$$\sigma_\epsilon^2(t) := \epsilon^{-1}([M]_{1+\epsilon t} - [M]_1) \to t\sigma^2(X_1).$$

This convergence is uniform in t over compact intervals so we get the a.s. functional convergence

$$\big((\sigma_\epsilon^2(-t))_{t\geq 0}, (\sigma_\epsilon^2(t))_{t\geq 0}\big) \to \big((-t\sigma^2(X_1))_{t\geq 0}, (t\sigma^2(X_1))_{t\geq 0}\big),$$

which we may add to the stable convergence in (III.8).

Now, for $t \in \mathbb{R}$ we can write

$$M_t^{(\epsilon)} = \epsilon^{-1/2}(M_{1+\epsilon t} - M_1) = \epsilon^{-1/2}(\tilde{W}_{[M]_{1+\epsilon t}} - \tilde{W}_{[M]_1}) = \tilde{W}^{(\epsilon)}_{\sigma_\epsilon^2(t)},$$

where $\tilde{W}^{(\epsilon)}$ is defined in Lemma III.12. By piecing the above together we obtain the convergence

$$\begin{aligned}\big((M_{-t}^{(\epsilon)})_{t\geq 0}, (M_t^{(\epsilon)})_{t\geq 0}\big) &\xrightarrow{st} \big((U^{(1)}_{t\sigma^2(X_1)})_{t\geq 0}, (U^{(2)}_{t\sigma^2(X_1)})_{t\geq 0}\big) \\ &= \big(\sigma(X_1)\tilde{U}^{(1)}, \sigma(X_1)\tilde{U}^{(2)}\big),\end{aligned}$$

where $\tilde{U}_t^{(i)} := \sigma^{-1}(X_1)U^{(i)}_{t\sigma^2(X_1)}$. Again we use continuity of subordination (Whitt, 2002, Thm. 13.2.2). We conclude by remarking that $(\tilde{U}^{(1)}, \tilde{U}^{(2)})$ is again a pair of independent standard Brownian motions, also independent of $\mathcal{F}$.

III.5.3 Proof of Theorem III.4

We begin by establishing that we may assume that X starts at zero and has no drift. As in Theorem III.4 $(B^{(1)}, B^{(2)})$ denotes a pair of independent Bessel-3 processes, defined on an extension of $(\Omega, \mathcal{F}, \mathbb{P})$ such that they are also independent of $\mathcal{F}$.

Following Kallenberg (2021, Ch. 33) we let p be the function given by

$$p'(x) = \exp\Big\{-2\int_{x_0}^{x} (\mu/\sigma^2)(u)\,\mathrm{d}u\Big\} \quad \text{and} \quad p(x_0) = 0.$$

Note that this definition of p has a problem at a value x if the function μ/σ^2 is not integrable over the interval $[x_0, x]$ (or $[x, x_0]$ depending on which is larger). However, if x is in the range of X then μ/σ^2 is bounded on $[x_0, x]$ (or $[x, x_0]$) due to Assumption III.A. As we will only need to evaluate p at such points we need not worry.

Now, let $Y_t := p(X_t)$ for $t \geq 0$. The choice of p has two particularly useful implications. Firstly, p is strictly increasing so $\overline{Y} = p(\overline{X})$ and $m^X = m^Y$. Secondly, Y is a diffusion process solving the SDE

$$\mathrm{d}Y_t = \tilde{\sigma}(Y_t)\,\mathrm{d}W_t \quad \text{and} \quad Y_0 = 0, \tag{III.11}$$

where $\tilde{\sigma} = (\sigma p') \circ p^{-1}$.

Now we are able to prove the following lemma which is an essential step in proving Theorem III.4.

Lemma III.13. *It is sufficient to prove Theorem III.4 under the assumption that $x_0 = 0$ and $\mu \equiv 0$.*

Proof. Assume that Theorem III.4 holds for any diffusion process which starts at zero, has no drift and satisfies Assumption III.A.

We consider the transformation $Y := p(X)$ introduced above. In addition to solving the SDE (III.11) we further note that Y satisfies Assumption III.A. So by our initial assumption there is the convergence

$$\big((\epsilon^{-1/2}\underset{\leftarrow}{Y}_{\epsilon t})_{t\geq 0}, (\epsilon^{-1/2}\underset{\rightarrow}{Y}_{\epsilon t})_{t\geq 0}\big) \xrightarrow{st} (-\tilde{\sigma}(\overline{Y})B^{(1)}, -\tilde{\sigma}(\overline{Y})B^{(2)}),$$

where $\underset{\leftarrow}{Y}$ and $\underset{\rightarrow}{Y}$ are pre- and post-supremum processes defined for the interval $[0, 1]$. Using the mean value theorem we find that

$$\epsilon^{-1/2}\underset{\rightarrow}{X}_{\epsilon t} = \epsilon^{-1/2}(p^{-1})'(c_\epsilon(t))\underset{\rightarrow}{Y}_{\epsilon t},$$

where $c_\epsilon(t)$ is between $Y_{m^X+\epsilon t}$ and $\overline{Y}$. One easily verifies that $(p^{-1})'(c_\epsilon(\,\cdot\,))$ converges (in the Skorokhod topology) to the constant function $(p^{-1})'(\overline{Y})$. Hence,

$$\big(\epsilon^{-1/2}(p^{-1})'(c_\epsilon(t))\underset{\rightarrow}{Y}_{\epsilon t}\big)_{t\geq 0} \xrightarrow{st} -(p^{-1})'(\overline{Y})\tilde{\sigma}(\overline{Y})B^{(2)} = -\sigma(\overline{X})B^{(2)},$$

where the final identity comes from the definition of $\tilde{\sigma}$. Obviously we can do similar calculations for the pre-supremum process. Hence,

$$\big((\epsilon^{-1/2}\underset{\leftarrow}{X}_{\epsilon t})_{t\geq 0}, (\epsilon^{-1/2}\underset{\rightarrow}{X}_{\epsilon t})_{t\geq 0}\big) \xrightarrow{st} (-\sigma(\overline{X})B^{(1)}, -\sigma(\overline{X})B^{(2)}).$$

□

For the rest of this subsection we assume that $x_0 = 0$ and $\mu \equiv 0$. Then, as in §III.2.3, we can write $X_t = \tilde{W}_{[X]_t}$ where $\tilde{W}$ is a standard Brownian motion and $[X]$ is the quadratic variation of X. To proceed we need the following result about zooming in at the supremum of

$\tilde{W}$, defined for the stochastic interval $[0,[X]_1]$. This result is essentially a direct consequence of Ivanovs (2018, Cor. 2) except for one technical complication. That paper works only on the canonical path space and since stable convergence is not only concerned with laws but also very much with the probability space the result does not apply directly. Instead we provide a short proof which fixes this problem.

Lemma III.14. *It holds that*

$$\big((\epsilon^{-1/2}\underleftarrow{\tilde{W}}_{\epsilon t})_{t\geq 0},(\epsilon^{-1/2}\underrightarrow{\tilde{W}}_{\epsilon t})_{t\geq 0}\big)\xrightarrow{st}(-B^{(1)},-B^{(2)}),\tag{III.12}$$

where $\underleftarrow{\tilde{W}}$ *and* $\underrightarrow{\tilde{W}}$ *are the pre- and post-supremum processes defined for the interval* $[0,[X]_1]$.

Proof. For each $T>0$ we let $\underleftarrow{\tilde{W}}^{(T)}$ and $\underrightarrow{\tilde{W}}^{(T)}$ denote the pre- and post-supremum processes for $\tilde{W}$, defined for the interval $[0,T]$. According to Ivanovs (2018, Thm. 4) there is the stable convergence

$$\big((\epsilon^{-1/2}\underleftarrow{\tilde{W}}^{(T)}_{\epsilon t})_{t\geq 0},(\epsilon^{-1/2}\underrightarrow{\tilde{W}}^{(T)}_{\epsilon t})_{t\geq 0}\big)\xrightarrow{\mathcal{H}-st}(-B^{(1)},-B^{(2)}),$$

where $\mathcal{H}$ is the σ-algebra generated by $\tilde{W}$. Since the left-hand side is obviously $\mathcal{H}$-measurable the $\mathcal{H}$-stable convergence extends to $\mathcal{F}$-stable convergence by Lemma III.8.

At this point it remains to extend to the case $T=[X]_1$. Corollary III.2 tells us that the supremum of $\tilde{W}$ over the interval $[0,[X]_1]$ is almost surely attained strictly before time $[X]_1$. Using this the convergence in (III.12) follows via the same arguments as in the proof of Ivanovs (2018, Cor. 2). □

Finally we are ready to prove Theorem III.4 in the case with $x_0=0$ and $\mu\equiv 0$. As in Lemma III.14 we let $\underleftarrow{\tilde{W}}$ and $\underrightarrow{\tilde{W}}$ denote the pre- and post-supremum processes for $\tilde{W}$ defined for the interval $[0,[X]_1]$.

Since $[X]_t=\int_0^t\sigma^2(X_s)\,\mathrm{d}s$ it follows immediately that

$$\sigma^2_\epsilon(t):=\epsilon^{-1}([X]_{m^X+\epsilon t}-[X]_{m^X})\to t\sigma^2(\overline{X})$$

a.s. for any $t\in\mathbb{R}$ since σ is continuous on the range of X. We note that this convergence is uniform on compact sets. Hence we have the almost sure functional convergence

$$\big((\sigma^2_\epsilon(-t))_{t\geq 0},(\sigma^2_\epsilon(t))_{t\geq 0}\big)\to\big((-t\sigma^2(\overline{X}))_{t\geq 0},(t\sigma^2(\overline{X}))_{t\geq 0}\big),\tag{III.13}$$

which we may add to the stable convergence in (III.12). We further note that

$$\begin{aligned}\epsilon^{-1/2}\underrightarrow{X}_{\epsilon t}&=\epsilon^{-1/2}(X_{m^X+\epsilon t}-\overline{X})\\&=\epsilon^{-1/2}(\tilde{W}_{[X]_{m^X+\epsilon t}}-\tilde{W}_{[X]_{m^X}})\\&=\epsilon^{-1/2}\underrightarrow{\tilde{W}}_{\epsilon\sigma^2_\epsilon(t)}\end{aligned}$$

for each $t\geq 0$. Similarly, it holds that $\epsilon^{-1/2}\underleftarrow{X}_{\epsilon t}=\epsilon^{-1/2}\underleftarrow{\tilde{W}}_{-\epsilon\sigma^2_\epsilon(-t)}$ for all $t\geq 0$. By continuity of subordination (Whitt, 2002, Thm. 13.2.2) we have the convergence

$$\begin{aligned}\big((\epsilon^{-1/2}\underleftarrow{X}_{\epsilon t})_{t\geq 0},(\epsilon^{-1/2}\underrightarrow{X}_{\epsilon t})_{t\geq 0}\big)&\xrightarrow{st}\big((-B^{(1)}_{t\sigma^2(\overline{X})})_{t\geq 0},(-B^{(2)}_{t\sigma^2(\overline{X})})_{t\geq 0}\big)\\&=\big(-\sigma(\overline{X})\tilde{B}^{(1)},-\sigma(\overline{X})\tilde{B}^{(2)}\big),\end{aligned}$$

where $\tilde{B}^{(i)}_t:=\sigma^{-1}(\overline{X})B^{(i)}_{t\sigma^2(\overline{X})}$. We note that $(\tilde{B}^{(1)},\tilde{B}^{(2)})$ is again a pair of Bessel-3 processes, independent of $\mathcal{F}$ and of each other. This concludes the proof of Theorem III.4.

Acknowledgements

I am thankful to Jan Pedersen and to my supervisor Jevgenijs Ivanovs for providing valuable feedback and for guiding me towards relevant literature.

Furthermore I gratefully acknowledge financial support of Sapere Aude Starting Grant 8049-00021B “Distributional Robustness in Assessment of Extreme Risk” from Independent Research Fund Denmark.

References

D. J. Aldous and G. K. Eagleson. On mixing and stability of limit theorems. *Ann. Probability*, 6(2):325–331, 1978. ISSN 0091-1798.

S. Asmussen, P. Glynn, and J. Pitman. Discretization error in simulation of one-dimensional reflecting Brownian motion. *Ann. Appl. Probab.*, 5(4):875–896, 1995. ISSN 1050-5164.

J. Bertoin. Splitting at the infimum and excursions in half-lines for random walks and Lévy processes. *Stochastic Process. Appl.*, 47(1):17–35, 1993. ISSN 0304-4149.

P. Billingsley. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. ISBN 0-471-19745-9. A Wiley-Interscience Publication.

K. Bisewski and J. Ivanovs. Zooming-in on a Lévy process: failure to observe threshold exceedance over a dense grid. *Electron. J. Probab.*, 25:Paper No. 113, 33, 2020.

E. Häusler and H. Luschgy. *Stable convergence and stable limit theorems*, volume 74 of *Probability Theory and Stochastic Modelling.* Springer, Cham, 2015. ISBN 978-3-319-18328-2; 978-3-319-18329-9.

J. Ivanovs. Zooming in on a Lévy process at its supremum. *Ann. Appl. Probab.*, 28(2): 912–940, 2018. ISSN 1050-5164.

J. Ivanovs and M. Podolskij. Optimal estimation of some random quantities of a Lévy process. *arXiv preprint arXiv:2001.02517*, 2020.

J. Ivanovs and J. D. Thøstesen. Discretization of the Lamperti representation of a positive self-similar Markov process. *Stochastic Process. Appl.*, 137:200–221, 2021. ISSN 0304-4149.

J. Jacod and P. Protter. *Discretization of processes*, volume 67 of *Stochastic Modelling and Applied Probability.* Springer, Heidelberg, 2012. ISBN 978-3-642-24126-0.

O. Kallenberg. *Foundations of modern probability*, volume 99 of *Probability Theory and Stochastic Modelling.* Springer, Cham, third edition, 2021. ISBN 978-3-030-61871-1; 978-3-030-61870-4.

P. Kosulajeff. Sur la répartition de la partie fractionnaire d’une variable. *Mat. Sbornik*, 2(5): 1017–1019, 1937.

A. E. Kyprianou. *Introductory lectures on fluctuations of Lévy processes with applications.* Universitext. Springer-Verlag, Berlin, 2006. ISBN 978-3-540-31342-7; 3-540-31342-7.

J. Lamperti. Semi-stable stochastic processes. *Trans. Amer. Math. Soc.*, 104:62–78, 1962. ISSN 0002-9947.

J. Lamperti. Semi-stable Markov processes. I. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 22:205–225, 1972.

M. Podolskij and M. Vetter. Understanding limit theorems for semimartingales: a short survey. *Stat. Neerl.*, 64(3):329–351, 2010. ISSN 0039-0402.

A. Rényi. On mixing sequences of sets. *Acta Math. Acad. Sci. Hungar.*, 9:215–228, 1958. ISSN 0001-5954.

A. Rényi. On stable sequences of events. *Sankhyā Ser. A*, 25:293 302, 1963. ISSN 0581-572X.

W. Whitt. *Stochastic-process limits*. Springer Series in Operations Research. Springer-Verlag, New York, 2002. ISBN 0-387-95358-2. An introduction to stochastic-process limits and their application to queues.

Paper IV

Graphical models for Lévy processes

Sebastian Engelke
Research Center for Statistics, University of Geneva

Jevgenijs Ivanovs
Department of Mathematics, Aarhus University

Jakob D. Thøstesen
Department of Mathematics, Aarhus University

Abstract. Conditional independence and graphical models for Lévy processes are seemingly unexplored topics. In this paper we consider a multivariate Lévy process and study conditional independence of its components, in particular leading to a notion of graphical models. Any Lévy process decomposes into the sum of a Brownian part and an independent jump part, and we show that these two parts may be considered separately when studying conditional independence. We further characterize conditional independence for the jump part in terms of the Lévy measure. Here we rely on a generalized notion of conditional independence which originates in extreme value theory. The class of stable Lévy processes is particularly interesting due to homogeneity of the Lévy measure. This is illustrated by considering models where the graphical structure is a tree. In this setting the jumps of the process have a particular product structure. We further present a method for consistent estimation of the tree given high frequency observations in a compact time interval. Lastly we discuss a common method for simulating a Lévy process, involving approximation of the small jumps using a Brownian motion. We show that under certain assumptions the Brownian approximation inherits the conditional independence structure of the Lévy process.

IV.1 Introduction

A Lévy process is an $\mathbb{R}^d$-valued stochastic process $\mathbf{X} = (X(t))_{t\geq 0}$ with independent and stationary increments that satisfies $X(0) = 0$ almost surely. For any fixed $t \geq 0$, the univariate marginal distribution of $X(t)$ is infinitely divisible and, thus, it arises as the limit of row sums of a triangular array. For this reason, Lévy processes are fundamental objects in limit theory that arise naturally in numerous settings. The Lévy–Itô decomposition states that any Lévy process can be decomposed into two independent processes,

$$\mathbf{X} = \mathbf{W} + \mathbf{J},$$

where the continuous part $\mathbf{W} = (W(t))_{t\geq 0}$ is a Brownian motion with covariance Σ and drift $\gamma \in \mathbb{R}$, and the jump process $\mathbf{J} = (J(t))_{t\geq 0}$ is described by the so-called Lévy measure Λ.

In the last decades, the probabilistic properties of Lévy processes have been extensively studied, ranging from detailed analysis of the sample path behavior to applications in physics, finance and other areas. Surprisingly, there is little known about conditional independence in this class of stochastic processes. In probability theory and statistics, conditional independence is a crucial notion of irrelevance that is at the heart of the many fields such as graphical models and causality. It further allows the definition of sparsity and is therefore the backbone for modern theory and statistical methods in larger dimensions.

One reason for the lacking connection is that the seemingly most natural definition of conditional independence for a Lévy process does not lead to useful characterizations. Indeed, for disjoint subsets $A, B, C \subseteq \{1, \dots, d\}$ and a fixed time point $t > 0$ one might be tempted to consider the conditional independence $X_A(t) \perp\!\!\!\perp X_B(t) \mid X_C(t)$. While such a statement is of course well-defined, we will argue that it is neither natural nor useful for theory or practice. Intuitively, the issue is that that conditioning only on the vector $X_C(t)$ collapses important information on the stochastic process up to time t into a single vector.

Instead, we propose to study conditional independence on the level of sample paths. As a first fundamental result, we show the (non-trivial) equivalence

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C \quad \Leftrightarrow \quad \mathbf{J}_A \perp\!\!\!\perp \mathbf{J}_B \mid \mathbf{J}_C \ \text{ and } \ \mathbf{W}_A \perp\!\!\!\perp \mathbf{W}_B \mid \mathbf{W}_C.$$

That is, the characterization of conditional independence of the process $\mathbf{X}$ can be separated into the corresponding statements for the Brownian and jump parts. Conditional independence for the Brownian motion part is well understood; if the corresponding covariance matrix Σ is invertible, then the any any conditional independence statement can be read off from the precision matrix Σ^{-1}.

Our main result characterizes conditional independence of the jump part $\mathbf{J}$. Since this part of the Lévy process is described by the Lévy measure Λ, we would like to describe stochastic properties in terms of this object. In fact, it turns out that exactly this is possible for conditional independence statement. Under a mild condition on Λ we show that for the jump process we have

$$\mathbf{J}_A \perp\!\!\!\perp \mathbf{J}_B \mid \mathbf{J}_C \quad \Leftrightarrow \quad A \perp B \mid C\,[\Lambda],$$

where the conditional independence notion on the right-hand side is a natural generalization of classical conditional independence to infinite measures exploding at the origin, see Engelke et al. (2022). This provides an effective tool to study conditional independence at the stochastic process level, to construct concrete examples and develop statistical methodology for Lévy processes by considering the much simpler object Λ.

Moreover, our theory allows us to define graphical models for Lévy processes, a mostly unexplored field with numerous open questions and potential for efficient statistical inference methods.

Figure IV.1 below shows a realization of a 3-dimensional Lévy process $\mathbf{X}$. It is clear that neither of the three components are independent since the paths have rather similar behavior. However, the processes $\mathbf{X}_1$ and $\mathbf{X}_3$ are conditionally independent given $\mathbf{X}_2$ but this is not obvious from looking at the plot. In §IV.6 we present a method that lets us learn this dependence structure from data.

In §IV.8 we consider daily stock prices for a selection of US companies. It is common to model the log prices using a Lévy process. In Figure IV.2 below we see the log prices for Target (TGT), United Parcel Service (UPS) and The Coca-Cola Company (KO) for the period January 2nd 2013 until December 31st 2015. Note that each time series has been shifted by its initial value such that they all start at zero. The analysis in §IV.8 suggests that TGT and KO are conditionally independent given UPS.

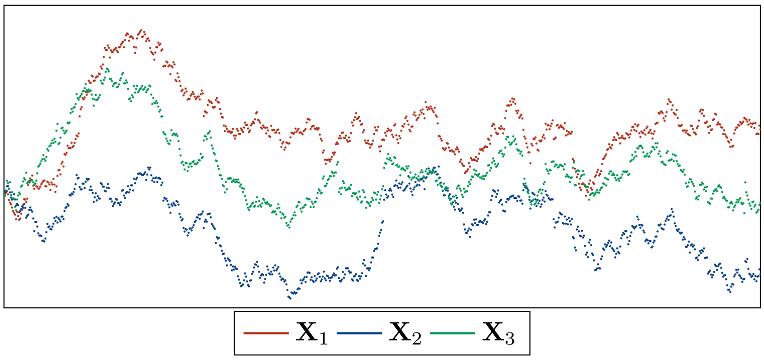


Figure IV.1: A realization of a 3-dimensional Lévy process $\mathbf{X}$ where $\mathbf{X}_1 \perp\!\!\!\perp \mathbf{X}_3 \mid \mathbf{X}_2$.

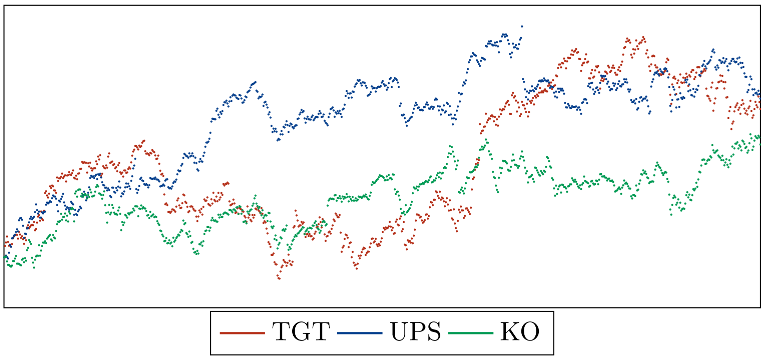


Figure IV.2: Shifted log prices for the tickers TGT, UPS and KO.

IV.1.1 Dependence models for stochastic processes

In the study of multivariate stochastic processes it is natural to want to model the dependence between different components. Numerous papers have aimed to do this and before we proceed we will briefly mention some of the proposed approaches. As is often the case, none of these are suited for all classes of processes and all applications. In practice one must therefore browse the literature for the most appropriate dependence model.

A popular concept is *Granger causality*, originally introduced by Granger (1969). The general idea is to study how information about one component of a time series up to time $n-1$ influences prediction of the value of another component at time n. Granger causality is the basis of many other papers such as Eichler (2012) which introduces graphical models in the context of multivariate time series.

The ideas from Granger causality can be extended to continuous time to obtain what is commonly known as *local independence*. The idea is to study how information about one coordinate influences the infinitesimal increment of another coordinate. Local independence is considered for various classes of stochastic processes. For instance, Didelez (2008) uses it to define graphical models for marked point processes. Another example is Mogensen and Hansen (2022) who study diffusion models with correlated driving noise.

The concepts of Granger causality and local independence are both based on dependence in time. This is entirely trivial if one considers a Lévy process since it has independent and stationary increments. For example, the evolution of the process after a fixed time point $T > 0$ is independent of the process up to time T. In this paper we concentrate instead on dependence in space. We look at individual jumps and study the dependence structure of such random vectors.

The idea of *α-stable graphical models* was introduced by Misra and Kuruoglu (2016). These models describe α-stable random vectors which are recursively constructed according to a directed acyclic graph. It is shown by Engelke et al. (2022, §7.3) that these models can be seen as a special case of directed graphical models based on the notion of conditional independence given in Definition IV.5 below. We note that the Lévy measures associated to these models are quite basic (for example, they do not admit a Lebesgue density) whereas our approach includes more complicated measures.

To model dependence of the jumps of a Lévy process Kallsen and Tankov (2006) use *Lévy copulas*, which are an analogue to copulas for random vectors. A somewhat similar idea based on so-called *Pareto Lévy measures* is suggested by Eder and Klüppelberg (2012). Both works choose to model the dependence via the Lévy measure. A feature they share with the approach presented in §IV.4 below.

IV.2 Preliminaries

In this section we will provide a brief introduction to conditional independence and Lévy processes along with a bit of notation.

Throughout we denote the index set $\{1, \ldots, d\}$ by V. This will eventually be the set of vertices in our graphical models. We will often consider sets of points $x \in \mathbb{R}^d$ which satisfy some conditions. To keep things simple we use the notation $\{x \in A\} = \{x \in \mathbb{R}^d \mid x \in A\}$. For example, we will write $\{x_1 \geq 1\}$ for the set $\{x \in \mathbb{R}^d \mid x_1 \geq 1\}$.

IV.2.1 Conditional independence

The following is a quick recap of conditional independence. A far more detailed account can be found in Kallenberg (2021, Ch. 8).

For random variables X, Y, Z, defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in spaces S_X, S_Y, S_Z, we say that X is conditionally independent of Y given Z if

$$\mathbb{P}(X \in A, Y \in B \mid Z) = \mathbb{P}(X \in A \mid Z)\mathbb{P}(Y \in B \mid Z)$$

almost surely for all measurable sets A and B. If this holds we write $X \perp\!\!\!\perp Y \mid Z$.

It is often convenient to write conditional probabilities using a probability kernel μ. That is,

$$\mathbb{P}((X, Y) \in F \mid Z = z) = \mu(z, F)$$

for all measurable F and $\mathbb{P}(Z \in \ \cdot\ )$-almost all z. If such a kernel exists the conditional independence $X \perp\!\!\!\perp Y \mid Z$ holds if and only if $\mu(z, A \times B) = \mu(z, A \times S_Y) \cdot \mu(z, S_X \times B)$ for for all measurable A, B and $\mathbb{P}(Z \in \ \cdot\ )$-almost all z.

A sufficient condition for existence of μ is that S_X, S_Y are Polish spaces. In this paper we will typically consider either $\mathbb{R}^n$ for some natural number n or the space of càdlàg functions from $[0, \infty)$ into $\mathbb{R}^n$. These are both examples of Polish spaces (where the latter is endowed with the Skorokhod topology, see Jacod and Shiryaev (2003, Thm. 1.14)).

IV.2.2 Lévy processes

Consider a stochastic process $\mathbf{X} = (X(t))_{t \geq 0}$ taking values in $\mathbb{R}^d$. Generally we will use bold letters to denote stochastic processes. We say that $\mathbf{X}$ is a Lévy process if it satisfies the following:

1. $\mathbb{P}(X(0) = 0) = 1$.

2. For $n \in \mathbb{N}$ and $0 \leq t_0 < t_1 < \cdots < t_n$ the increments $X(t_1) - X(t_0), \ldots, X(t_n) - X(t_{n-1})$ are independent.

3. The distribution of $X(t+s) - X(s)$ does not depend on s.

4. The sample path $t \mapsto X(t)$ is càdlàg $\mathbb{P}$-almost surely.

Typical references for Lévy processes include Bertoin (1996), Sato (1999) and Applebaum (2009).

A basic consequence of the definition above is that the distribution of $X(1)$ is infinitely divisible. Therefore, see e.g. Sato (1999, Thm. 8.1), its characteristic function is given by the Lévy–Khintchine formula

$$\mathbb{E}[e^{i\langle u, X(1)\rangle}] = \exp\Big(i\langle u, \gamma\rangle - \frac{1}{2}\langle u, \Sigma u\rangle + \int_{\mathbb{R}^d} e^{i\langle u,x\rangle} - 1 - i\langle u, x\rangle\, \mathbf{1}_{\{\|x\|\leq 1\}}\, \Lambda(\mathrm{d}x)\Big), \qquad u \in \mathbb{R}^d,$$

where $\gamma \in \mathbb{R}^d$, Σ is a positive semidefinite $d \times d$ matrix, $\|\cdot\|$ is a norm on $\mathbb{R}^d$ and Λ is a measure on $\mathbb{R}^d$. We say that $\mathbf{X}$ has characteristic triplet $(\gamma, \Sigma, \Lambda)$. The so-called Lévy measure Λ satisfies

$$\int_{\mathbb{R}^d} (1 \wedge \|x\|^2)\, \Lambda(\mathrm{d}x) < \infty. \tag{IV.1}$$

In particular, $\Lambda(F) < \infty$ for any $F \in \mathcal{B}(\mathbb{R}^d)$ bounded away from 0.

Just like any Lévy process gives rise to an infinitely divisible distribution the converse is also true. That is, for any infinitely divisible distribution ν on $\mathbb{R}^d$ there exists a Lévy process $\mathbf{X}$ such that $X(1) \sim \nu$. Some typical examples of Lévy processes are:

- A deterministic drift, i.e. $X(t) = t \cdot \gamma$ for some $\gamma \in \mathbb{R}^d$. The associated triplet is $(\gamma, 0, 0)$.
- A Brownian motion with drift γ and covariance matrix Σ. Here the triplet is $(\gamma, \Sigma, 0)$.
- A compound Poisson process with rate $\lambda \geq 0$ and jump distribution μ. The triplet is given by $(\gamma, 0, \Lambda)$ where $\Lambda = \lambda \cdot \mu$ and $\gamma = \int_{\{\|x\|\leq 1\}} x\, \Lambda(\mathrm{d}x)$.
- An α-stable process. That is, a Lévy process where the distribution of $X(1)$ is α-stable. See §IV.5 for further details.

Furthermore, the sum of two independent Lévy processes is again a Lévy process and its triplet is simply the sum of the two original triplets.

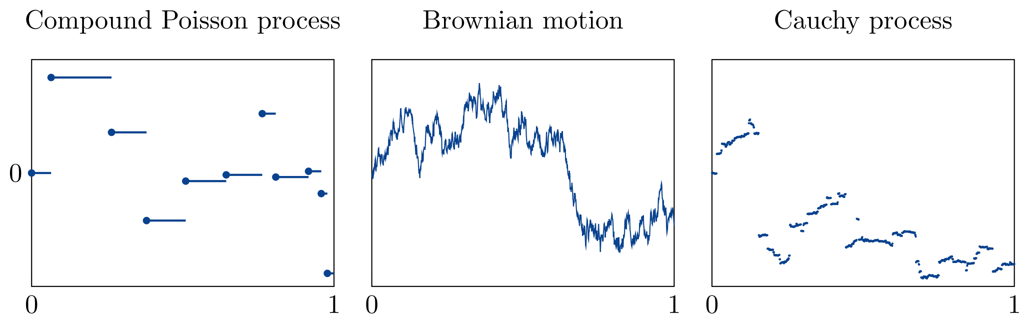


Figure IV.3: Examples of Lévy processes.

A Lévy process may be represented by its Lévy–Itô decomposition. More precisely, if $\mathbf{X}$ is a Lévy process with triplet $(\gamma, \Sigma, \Lambda)$ we can write it as $\mathbf{X} = \mathbf{J} + \mathbf{W}$, where $\mathbf{J}$ is a Lévy process with characteristic triplet $(0, 0, \Lambda)$, and $\mathbf{W}$ is a Brownian motion with drift γ and covariance matrix Σ which is independent of $\mathbf{J}$. Since $\mathbf{W}$ is a.s. everywhere continuous it is common to refer to $\mathbf{J}$ and $\mathbf{W}$ as respectively the jump part and the Brownian part of $\mathbf{X}$. Importantly, the terms $\mathbf{J}$ and $\mathbf{W}$ may be constructed from $\mathbf{X}$. A fact which will be used to prove multiple results in this paper.

IV.3 Conditional independence for the process

We want to discuss conditional independence for the components of a d-dimensional Lévy process. More precisely, for disjoint subsets $A, B, C \subseteq V$ we consider the statement

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C.$$

We remark that this is just one way of thinking about conditional independence for Lévy processes. Another option is to consider $X_A(t_0) \perp\!\!\!\perp X_B(t_0) \mid X_C(t_0)$ for some $t_0 > 0$. The primary difference between these two set-ups is the amount of information that we condition on since $\sigma(\mathbf{X}_C)$ is much larger than $\sigma(X_C(t_0))$. Indeed, we note that not only does $X_C(t_0)$ tell us nothing about the the future (i.e. $(X_C(t_0 + t) - X_C(t_0))_{t \geq 0}$) but it also provides very little information about the fluctuations and jumps up to time t_0. As such the idea of conditioning on the entire process $\mathbf{X}_C$ seems more interesting. In the following subsection we briefly discuss how these two types of conditional independence are related.

IV.3.1 Fixed times

An important question is whether the following equivalence holds.

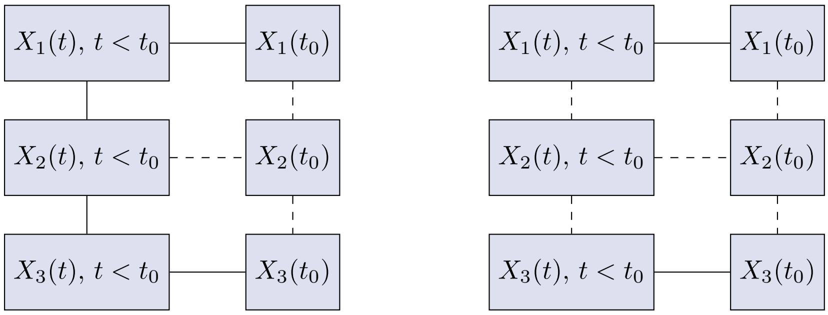
$$\text{There exists } t_0 > 0 \text{ such that } X_A(t_0) \perp\!\!\!\perp X_B(t_0) \mid X_C(t_0) \quad \Leftrightarrow \quad \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C. \tag{IV.2}$$

As the following example demonstrates we can rule out the implication from right to left in (IV.2). The issue is that $X_A(t_0)$ and $X_B(t_0)$ may have some dependence which is determined by the full history of $\mathbf{X}_C$ on $[0, t_0]$ and not only by $X_C(t_0)$.

Example IV.1. *Let $\mathbf{X}$ be a 3-dimensional compound Poisson process such that the jumps of $\mathbf{X}_2$ take the values -1 or 1 (both with positive probability). Moreover, $\mathbf{X}_1$ and $\mathbf{X}_3$ are independent, jumping precisely when $\Delta\mathbf{X}_2 = 1$ and $\Delta\mathbf{X}_2 = -1$ respectively. The jumps of $\mathbf{X}_1$ and $\mathbf{X}_3$ are all of size 1. Then $\mathbf{X}_1 \perp\!\!\!\perp \mathbf{X}_3 \mid \mathbf{X}_2$. For any $t_0 > 0$ the random variables $X_1(t_0)$ and $X_3(t_0)$ are not conditionally independent given $X_2(t_0)$. For example, conditionally on the event $\{X_2(t_0) = 0\}$ (which has positive probability) we have $X_1(t_0) = X_3(t_0)$ and this common value is not deterministic.*

The difference between conditioning on one value and conditioning on the entire path can also be illustrated visually. Suppose that $\mathbf{X}$ is 3-dimensional with $\mathbf{X}_1 \perp\!\!\!\perp \mathbf{X}_3 \mid \mathbf{X}_2$. Figure IV.4 shows part of the dependence structure. Observe that conditioning on $X_2(t_0)$ does not guarantee independence between $X_1(t_0)$ and $X_3(t_0)$ since there is still a potential connection via the past $X_2(t)$, $t < t_0$. The solid edges illustrate possible dependence while the dashed edges are the ones closed by the conditioning, i.e. those where at least one of the connected vertices is deterministic. Observe that conditioning on the entire path, i.e. $X_2(t)$, $t \leq t_0$, closes any possible connection between $X_1(t_0)$ and $X_3(t_0)$.

To obtain an implication similar to the one from left to right in (IV.2) we have to replace the left-hand side with a stronger statement. The following result provides a sufficient condition.



(a) Conditioning on $X_2(t_0)$. **(b)** Conditioning on $X_2(t),\ t \leq t_0$.

Figure IV.4: Illustration of the difference between conditioning on $X_2(t_0)$ and conditioning on $X_2(t),\ t \leq t_0$. To disconnect $X_1(t_0)$ and $X_3(t_0)$ it is not enough to condition on the present value $X_2(t_0)$.

Proposition IV.2. *There is the implication:*

$$X_A(t) \perp\!\!\!\perp X_B(t) \mid X_C(t) \text{ for all } t \geq 0 \quad \Rightarrow \quad \mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C.$$

The proof of Proposition IV.2 is in §IV.A.3 and does in fact not require that $X_A(t) \perp\!\!\!\perp X_B(t) \mid X_C(t)$ for all $t \geq 0$. Instead it is sufficient for this to hold for all $t \in (0, \epsilon)$ for some $\epsilon > 0$.

Assume for a moment that $\mathbf{X}$ is a stable Lévy process (we will come back to this in §IV.5). That is, for any $h > 0$ there exist $b > 0$ and $c \in \mathbb{R}^d$ such that $X(h) \stackrel{d}{=} bX(1) + c$. Then the left-hand side in Proposition IV.2 is equivalent to having the conditional independence $X_A(t_0) \perp\!\!\!\perp X_B(t_0) \mid X_C(t_0)$ for a single time point $t_0 > 0$. Whether this is also true without stability is unclear. Importantly, distributional properties of $X(t)$ may be time dependent. Sato (1999, §23) lists some examples of this such as unimodality and absolute continuity.

IV.3.2 The Lévy–Itô decomposition

As mentioned in the introduction we can write the process in its Lévy–Itô decomposition $\mathbf{X} = \mathbf{J} + \mathbf{W}$, where $\mathbf{J}$ is the jump part and $\mathbf{W}$ is the Brownian part. Interestingly, Proposition IV.3 below shows that we may study conditional independence for $\mathbf{J}$ and $\mathbf{W}$ separately.

Proposition IV.3. *There is the equivalence*

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C \quad \Leftrightarrow \quad \mathbf{J}_A \perp\!\!\!\perp \mathbf{J}_B \mid \mathbf{J}_C \text{ and } \mathbf{W}_A \perp\!\!\!\perp \mathbf{W}_B \mid \mathbf{W}_C.$$

While this result is not the most surprising it is also not as trivial as it might appear. Note for example that for independent random vectors Y and Z the implication

$$Y_A \perp\!\!\!\perp Y_B \mid Y_C \text{ and } Z_A \perp\!\!\!\perp Z_B \mid Z_C \quad \Rightarrow \quad Y_A + Z_A \perp\!\!\!\perp Y_B + Z_B \mid Y_C + Z_C$$

does not hold in general. The issue is that conditioning on $Y_C + Z_C$ is not guaranteed to provide enough information about the individual terms Y_C and Z_C. The proof of Proposition IV.3 is in §IV.A.4 and makes use of the fact that $\mathbf{J}_C$ and $\mathbf{W}_C$ may be constructed from $\mathbf{X}_C$.

The Brownian part $\mathbf{W}$ is stable (with $\alpha = 2$) so Proposition IV.2 and the subsequent discussion tells us that $\mathbf{W}_A \perp\!\!\!\perp \mathbf{W}_B \mid \mathbf{W}_C$ if and only if $W_A(t_0) \perp\!\!\!\perp W_B(t_0) \mid W_C(t_0)$ for some $t_0 > 0$. If Σ is invertible we know that the latter is equivalent to having $(\Sigma^{-1})_{ij} = 0$ for all $i \in A, j \in B$. In §IV.4 we study conditional independence for the jump part $\mathbf{J}$.

IV.3.3 The killed process

The concept of killing is central in the theory of Lévy processes. It does, for example, allow us to consider a setting where $\mathbf{X}$ is only observed until some time point $T > 0$. Here we briefly discuss it in the context of conditional independence.

For the Lévy process $\mathbf{X}$ and a fixed time $T > 0$ we define the *killed process* $\mathbf{X}^T$ by

$$X^T(t) = \begin{cases} X(t) & \text{for } t < T, \\ \dagger & \text{for } t \geq T. \end{cases}$$

Then we have the following simple result.

Proposition IV.4. *There is the equivalence:*

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C \quad \Leftrightarrow \quad \mathbf{X}_A^T \perp\!\!\!\perp \mathbf{X}_B^T \mid \mathbf{X}_C^T \textit{ for all } T > 0.$$

Extending Proposition IV.4 to the case where T can be any positive stopping time is not possible. The interesting implication is the one from the left, and in general it does not hold. For example, if $|A| = |B| = 1$ we can look at $T = \inf\{t \geq 1 \mid |X_A(t) - X_B(t)| \leq \epsilon\}$ for some $\epsilon > 0$. If we further assume that $\mathbf{X}_C$ is independent of $\mathbf{X}_{A\cup B}$ we easily see that $\mathbf{X}_A^T$ and $\mathbf{X}_B^T$ are not conditionally independent given $\mathbf{X}_C^T$.

For left-hand side in Proposition IV.4 to imply conditional independence for the killed process $\mathbf{X}^T$ it is sufficient to assume that T is $\sigma(\mathbf{X}_C)$-measurable. In fact, the original proof goes through without changes.

IV.4 Conditional independence for the jumps

We consider now a Lévy process $\mathbf{X}$ with characteristic triplet $(\gamma, 0, \Lambda)$, meaning that $\mathbf{X}$ has no Brownian component. Studying conditional independence for the jumps of $\mathbf{X}$ is straightforward if the the Lévy measure is finite. One simply considers $Y_A \perp\!\!\!\perp Y_B \mid Y_C$ where Y is a random vector with distribution $\Lambda/\Lambda(\mathbb{R}^d)$ (the jump distribution). However, if Λ has infinite mass this is obviously not possible, so it becomes necessary to find a different approach. What we will use is a generalized notion of conditional independence which was introduced by Engelke et al. (2022). This is inspired by the work of Engelke and Hitz (2020) who considered conditional independence for multivariate Pareto distributions. One of the main contributions of the present paper is Theorem IV.7 below which characterizes conditional independence for $\mathbf{X}$ in terms of its Lévy measure.

IV.4.1 Conditional independence for the Lévy measure

Recall the notation $V = \{1, \ldots, d\}$. We write

$$\mathcal{R}(\Lambda) = \{R = \times_{v\in V} R_v \mid R_v \in \mathcal{B}(\mathbb{R}), 0 \notin \bar{R}, \Lambda(R) > 0\}$$

for the class of product sets which have positive mass and are bounded away from the origin. We require each set $R \in \mathcal{R}(\Lambda)$ to have positive and finite mass since this means that it induces a probability measure $\mathbb{P}_R = \Lambda(\,\cdot\, \cap R)/\Lambda(R)$. The reason for restricting to product sets is that it is natural in the context of independence. Indeed, two random vectors can only be independent if the joint support is a product set. We will often consider a random variable Y with distribution $\mathbb{P}_R$. To simplify notation we typically think of Y as the identity such that $\mathbb{P}_R(Y \in A) = \mathbb{P}_R(A)$.

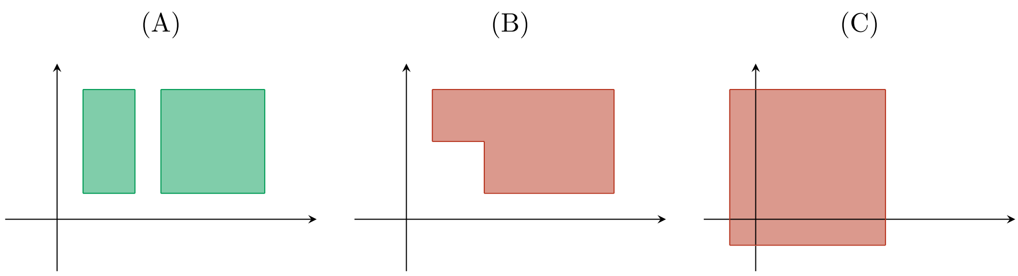


Figure IV.5: Examples of sets in $\mathbb{R}^2$. Figures (A) and (B) show two sets which are bounded away from 0 but only the former is a product set. Meanwhile, (C) shows a product set containing 0. Hence, only (A) shows a set which may be in $\mathcal{R}(\Lambda)$.

For a non-empty subset $D \subseteq V$ we may consider two measures on $\mathcal{E}^D = \mathbb{R}^D \setminus \{0_D\}$ defined by

$$\Lambda_D = \Lambda(\{x_D \in \cdot\,\}), \qquad \Lambda_D^0 = \Lambda(\{x_D \in \cdot\,, x_{V\setminus D} = 0_{V\setminus D}\}).$$

For $v \in V$ we shall use the simplified notation $\Lambda_v = \Lambda_{\{v\}}$. The distributions $\mathbb{P}_R$ are now used for defining conditional independence w.r.t. the measure Λ.

Definition IV.5. *Let $A, B, C \subseteq V$ be disjoint. We say that Λ admits conditional independence of A and B given C if and only if*

$$Y_A \perp\!\!\!\perp Y_B \mid Y_C \quad \textit{for} \quad Y_{A\cup B\cup C} \sim \mathbb{P}_{R_{A\cup B\cup C}} \quad \textit{for all} \quad R_{A\cup B\cup C} \in \mathcal{R}(\Lambda_{A\cup B\cup C}).$$

We denote this conditional independence by $A \perp B \mid C\,[\Lambda]$. For $C = \emptyset$ we say that Λ admits independence of A and B and this is denoted by $A \perp B\,[\Lambda]$.

To verify the conditional independence introduced in Definition IV.5 it is sufficient to consider certain sub-classes of product sets. In particular, we may look at sets on the form $R_{v,\epsilon} = \{|x_v| \geq \epsilon\}$ for $v \in V$ and $\epsilon > 0$. We shall use the simplified notation $\mathbb{P}_{v,\epsilon} = \mathbb{P}_{R_{v,\epsilon}}$. For a partition $A, B, C \subseteq V$ Engelke et al. (2022, Thm. 4.1) showed that the following statements are equivalent:

- $A \perp B \mid C\,[\Lambda]$.
- $Y_A \perp\!\!\!\perp Y_B \mid Y_C$ for $Y \sim \mathbb{P}_{v,\epsilon}$ for all $v \in V, \epsilon > 0$.
- $Y_A \perp\!\!\!\perp Y_B \mid Y_C$ for $Y \sim \mathbb{P}_{c,\epsilon}$ for all $c \in C, \epsilon > 0$, and $A \perp B\,[\Lambda^0_{A\cup B}]$.

Contained in this result is the often useful implication:

$$A \perp B \mid C\,[\Lambda] \quad \Rightarrow \quad A \perp B\,[\Lambda^0_{A\cup B}]$$

for a partition $A, B, C \subseteq V$.

IV.4.2 Distributional interpretation of the conditional independence

The conditional independence in Definition IV.5 concerns the Lévy measure and hence the jumps of the process. In this subsection we show how this translates to statements about the process $\mathbf{X}$ itself.

We begin by discussing the simpler case of independence. Assume that

$$\Lambda(\{x_v \neq 0\}) \in \{0, \infty\} \quad \text{for all } v \in V. \qquad \text{[IV.A1]}$$

Then, for a partition $A, B \subseteq V$, Engelke et al. (2022, Prop. 5.1) show that the independence $A \perp B\,[\Lambda]$ holds if and only if $\Lambda(\{x_A \neq 0_A, x_B \neq 0_B\}) = 0$. The latter is the same as saying that the Lévy measure is concentrated on the subfaces $\{x_A = 0_A\}$ and $\{x_B = 0_B\}$. Without [IV.A1] we still have that $\Lambda(\{x_A \neq 0_A, x_B \neq 0_B\}) = 0$ implies independence w.r.t. Λ. This leads to the following independence result for the Lévy process.

Lemma IV.6. *Let $A, B \subseteq V$ be a partition and assume [IV.A1]. Then there is the equivalence:*

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \quad \Leftrightarrow \quad A \perp B\,[\Lambda].$$

The implication from the left holds even if [IV.A1] is false.

Proof. The processes $\mathbf{X}_A$ and $\mathbf{X}_B$ are independent if and only if they have no simultaneous jumps, corresponding to $\Lambda(\{x_A \neq 0_A, x_B \neq 0_B\}) = 0$. □

Having dealt with independence we turn to the generally more complicated case of conditional independence. For this we need need a stronger assumption than before. More precisely, [IV.A1] is replaced with the following.

$$\Lambda^0_D(\{x_d \neq 0\}) \in \{0, \infty\} \quad \text{for all } D \subseteq V \text{ and all } d \in D. \qquad \text{[IV.A2]}$$

We will now see that Lemma IV.6 is just a special case of the more general Theorem IV.7 below. Intuitively it tells us that $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C$ if and only if the jumps of $\mathbf{X}$ exhibit the same conditional independence.

Theorem IV.7. *Let $A, B, C \subseteq V$ be a partition. Under [IV.A2] there is the equivalence:*

$$\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C \quad \Leftrightarrow \quad A \perp B \mid C\,[\Lambda].$$

The implication from the left holds without [IV.A2].

The proof of the result is postponed to §IV.A.6 but we can sketch the general idea, concentrating on the implication from the right. It is possible to write $\mathbf{X} = \mathbf{X}' + \mathbf{X}''$, where $\mathbf{X}', \mathbf{X}''$ are independent Lévy processes with Lévy measures Λ', Λ'' given by Λ restricted to $\{x_C = 0\}$ and $\{x_C \neq 0\}$ respectively. The idea is to show that both terms satisfy the conditional independence between components A and B given C. For $\mathbf{X}''$ we know that all jumps have a non-zero C-component. Hence, conditioning on $\mathbf{X}''_C$ fixes all jump times. We combine this with the conditional independence $A \perp B \mid C\,[\Lambda]$. For $\mathbf{X}'$ we note that it is essentially a $(d - |C|)$-dimensional process since the C-component is deterministic. The associated Lévy measure is $\Lambda^0_{V\setminus C}$ and due to [IV.A2] it satisfies [IV.A1]. Hence, we may apply Lemma IV.6 (recall that $A \perp B \mid C\,[\Lambda]$ implies $A \perp B\,[\Lambda^0_{V\setminus C}]$).

IV.4.3 Graphical models

As previously advertised we want to define (undirected) graphical models for Lévy processes. The main object is a graph $G = (V, E)$, where $V = \{1, \ldots, d\}$ is the index set and E is a subset of $V \times V$ with a pair $(i, j) \in E$ representing an edge between the vertices i and j. Since our focus is on undirected models we do not care about the ordering, i.e. (i, j) and (j, i) are seen as the same edge.

For sets $A, B, C \subseteq V$ we say that C separates A and B in G if every path from any vertex in A to any vertex in B goes through C. We say that the Lévy process $\mathbf{X}$ satisfies the *global Markov property* with respect to a graph G if $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C$ for any disjoint sets $A, B, C \subseteq V$ such that C separates A and B in G. In this case we call the pair $(\mathbf{X}, G)$ a graphical model.

These graphical models are also what we call *semi-graphoids*. That is, for a graphical model $(\mathbf{X}, G)$ and disjoint subsets $A, B, C, D \subseteq V$ the following four properties are satisfied:

(L1) If $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C$ then $\mathbf{X}_B \perp\!\!\!\perp \mathbf{X}_A \mid \mathbf{X}_C$.

(L2) If $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_{B\cup D} \mid \mathbf{X}_C$ then $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C$.

(L3) If $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_{B\cup D} \mid \mathbf{X}_C$ then $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_{C\cup D}$.

(L4) If $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C$ and $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_D \mid \mathbf{X}_{B\cup C}$ then $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_{B\cup D} \mid \mathbf{X}_C$.

Importantly, Lauritzen (1996, Prop. 3.4) uses these semi-graphoid properties to show that the global Markov property implies the local Markov property which further implies the pairwise Markov property.

Graphical models need not be based on the traditional notion of conditional independence. In fact, in the work by Engelke et al. (2022) the idea is to study graphical models based on conditional independence w.r.t. Λ as defined in Definition IV.5. One of many results is that the semi-graphoid properties are satisfied under [IV.A2]. A fact which may also be seen as a corollary of Theorem IV.7. This connection makes another argument for the need of [IV.A2] in the above.

IV.5 Stable processes

The class of so-called stable processes have particularly nice Lévy measures which allow for some interesting theory along with the possibility of consistently estimating certain graphical structures. We will now give a brief introduction to stable processes. A more detailed overview is provided by e.g. Sato (1999, §13–14).

A Lévy process $\mathbf{X}$ is said to be stable if for any $h > 0$ there exist $b > 0$ and $c \in \mathbb{R}^d$ such that

$$X(h) \stackrel{d}{=} bX(1) + c.$$

This turns out to be equivalent to a seemingly stronger statement. Namely that there exists $\alpha \in (0, 2]$ with the following property:

$$\begin{gathered}\text{For any } h > 0 \text{ there exists } c\colon [0,\infty) \to \mathbb{R}^d \text{ such that} \\ (X(ht))_{t\geq 0} \stackrel{d}{=} (h^{1/\alpha}X(t) + c(t))_{t\geq 0}.\end{gathered} \tag{IV.3}$$

The constant α is known as the stability index of $\mathbf{X}$.

If (IV.3) holds with $c \equiv 0$ for all values of $h > 0$ we say that $\mathbf{X}$ is strictly stable. In that case the property in (IV.3) is also called self-similarity with exponent $1/\alpha$.

The stability index is essential in describing stable processes. For $\alpha = 2$ the process $\mathbf{X}$ is α-stable if and only if it is a Brownian motion, i.e. its characteristic triplet is $(\gamma, \Sigma, 0)$. For $\alpha \in (0, 2)$, on the other hand, $\mathbf{X}$ is α-stable if and only if its triplet is of the form $(\gamma, 0, \Lambda)$ (i.e. no Brownian component) and the Lévy measure is $-\alpha$-homogeneous. The latter means that $\Lambda(hE) = h^{-\alpha}\Lambda(E)$ for any $h > 0$ and $E \in \mathcal{B}(\mathbb{R}^d)$. For any $v \in V$ the marginal measure Λ_v is also $-\alpha$-homogeneous. Consequently,

$$\Lambda_v(\mathrm{d}u) = \begin{cases} m_v^+ \alpha u^{-1-\alpha}\,\mathrm{d}u & \text{for } u > 0, \\ m_v^- \alpha |u|^{-1-\alpha}\,\mathrm{d}u & \text{for } u < 0, \end{cases}$$

where $m_v^\pm = \Lambda(\{\pm x_v \geq 1\})$. Note that we may have $\Lambda(\{x_v = 0\}) > 0$.

We remark that [IV.A2] is automatically satisfied if Λ is $-\alpha$-homogeneous. Indeed, if $\Lambda_D^0(\{x_d \neq 0\}) > 0$ for some $D \subseteq V$ and $d \in V$, then there is $n \in \mathbb{N}$ with $\Lambda_D^0(\{|x_d| \geq 1/n\}) > 0$. Using homogeneity we find that

$$\begin{aligned}\Lambda_D^0(\{x_d \neq 0\}) &= \lim_{m\to\infty} \Lambda_D^0(\{|x_d| \geq 1/m\}) \\ &= \lim_{m\to\infty} \Lambda(\{|x_d| \geq n/(nm), x_{V\setminus D} = 0_{V\setminus D}\}) \\ &= \lim_{m\to\infty} (n/m)^{-\alpha} \Lambda_D^0(\{|x_d| \geq 1/n\}) \\ &= \infty.\end{aligned}$$

IV.5.1 The kernel representation

For the rest of this section we assume that $\mathbf{X}$ is a stable Lévy process. We will utilize the homogeneity of Λ to derive the existence of certain random vectors which will become essential in describing the structure of the jumps.

For a non-empty subset $C \subseteq V$ there is a Λ_C-unique probability kernel $\nu_C \colon \mathcal{E}^C \times \mathcal{B}(\mathbb{R}^{V\setminus C}) \to [0,1]$ such that for any $R \in \mathcal{R}(\Lambda)$ with $0_C \notin R_C$ we have that

$$\Lambda(R) = \int_{R_C} \nu_C(x_C, R_{V\setminus C})\, \Lambda_C(\mathrm{d}x_C),$$

see Engelke et al. (2022, Lem. 4.3). Furthermore, for any $c \in C$ and $\epsilon > 0$ with $\Lambda(R_{c,\epsilon}) > 0$ there is the identity

$$\nu_C(x_C, R_{V\setminus C}) = \mathbb{P}_{c,\epsilon}(Y_{V\setminus C} \in R_{V\setminus C} \mid Y_C = x_C) \tag{IV.4}$$

for Λ_C-almost all x_C with $|x_c| \geq \epsilon$.

We will focus on the case where C contains just one element c. In this case $\nu_{\{c\}}$ admits a rather particular representation.

Lemma IV.8. *For any $c \in V$ there exist two d-dimensional random vectors $\xi^{(c,+)}, \xi^{(c,-)}$ such that $\xi_c^{(c,\pm)} = \pm 1$ almost surely, and for Λ_c-almost all $h \neq 0$*

$$\nu_{\{c\}}(h, \cdot\,) = \mathbb{P}(|h|\xi^{(c,\pm)}_{V\setminus\{c\}} \in \cdot\,) \quad \textit{for } \pm h > 0.$$

According to (IV.4) we may think of $\nu_{\{c\}}(h, \cdot\,)$ as the conditional distribution of $Y_{V\setminus\{c\}}$ given $Y_c = h$, where $Y \sim \mathbb{P}_{c,\epsilon}$ for some $\epsilon > 0$. Hence, the intuitive interpretation of Lemma IV.8 is that sampling Y conditionally on $Y_c = h \neq 0$ is the same as sampling $\xi^{(c,+)}$ or $\xi^{(c,-)}$ (depending on the sign of h) followed by multiplication with $|h|$.

Lemma IV.9. *The random vectors $\xi^{(c,\pm)}$ satisfy $\mathbb{E}[\|\xi^{(c,\pm)}\|^\alpha] < \infty$ and every such pair of vectors may arise from some -α-homogeneous Lévy measure. Furthermore, for any $v \in V\setminus\{c\}$,*

$$\int_{\{x_c \in (0,1)\}} x_v^2\, \Lambda(\mathrm{d}x) = \frac{\alpha m_c^+}{2-\alpha}\mathbb{E}[(\xi_v^{(c,+)})^2],$$

and

$$\int_{\{x_c \in (-1,0)\}} x_v^2\, \Lambda(\mathrm{d}x) = \frac{\alpha m_c^-}{2-\alpha}\mathbb{E}[(\xi_v^{(c,-)})^2].$$

Assume that Λ is concentrated on $[0, \infty)^d$. Since $\xi^{(c,-)}$ is not of interest in this situation we will use the slightly simpler notation $\xi^{(c)} = \xi^{(c,+)}$. If $\Delta^{(c)}$ is a random vector with distribution $\mathbb{P}_{c,\epsilon}$ it follows from Lemma IV.8 that

$$\Delta^{(c)} \overset{d}{=} P \cdot \xi^{(c)} \tag{IV.5}$$

where P is independent of $\xi^{(c)}$ and absolutely continuous with density proportional to $x \mapsto x^{-1-\alpha}$ on $[\epsilon, \infty)$. The latter is the same as saying that P follows a Pareto distribution with scale ϵ and shape α.

IV.5.2 Tree models

In this subsection we assume that Λ satisfies the global Markov property w.r.t. a tree $T = (V, E)$. The idea is that this is a graph where we condition on only one vertex at a time. Lemma IV.8 tells us that conditioning on a single vertex c can be expressed using the random vectors $\xi^{(c,+)}, \xi^{(c,-)}$.

Given a pair of vertices $i, j \in V$ we define $\xi^{\pm}_{(i,j)} = \xi^{(i,\pm)}_j$. For $c \in V$ and $\epsilon > 0$ we have

$$\mathbb{P}_{c,\epsilon}(Y_v \in |h|F \mid Y_c = h) = \mathbb{P}(\xi^{\mathrm{sgn}(h)}_{(c,v)} \in F) \quad \text{for all } F \in \mathcal{B}(\mathbb{R}) \tag{IV.6}$$

for Λ_c-almost all $|h| \geq \epsilon$, where $\mathrm{sgn}(h) = +$ for $h \geq 0$ and $\mathrm{sgn}(h) = -$ for $h < 0$.

We further want to know the conditional distribution of Y_j given $Y_i = h$ under $\mathbb{P}_{c,\epsilon}$ in the particular case where c and j are separated by i in the tree T.

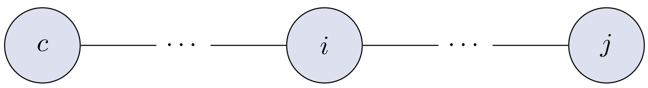


Figure IV.6: Graphical representation of $\{c\} \perp \{j\} \mid \{i\} \, [\Lambda]$.

Lemma IV.10. *Suppose that $\{c\} \perp \{j\} \mid \{i\} \, [\Lambda]$ as illustrated in Figure IV.6. Then for any $\epsilon > 0$ and $\mathbb{P}_{c,\epsilon}(Y_i \in \cdot\,)$-almost all $h \neq 0$*

$$\mathbb{P}_{c,\epsilon}(Y_j \in |h|F \mid Y_i = h) = \mathbb{P}(\xi^{\mathrm{sgn}(h)}_{(i,j)} \in F) \quad \textit{for all } F \in \mathcal{B}(\mathbb{R}). \tag{IV.7}$$

Proof. For $\epsilon, \delta > 0$ and $F \in \mathcal{B}(\mathbb{R})$ we find that

$$\begin{aligned}
\Lambda(\{|x_c| \geq \epsilon, x_i \geq \delta, x_j \in x_i F\}) &= \int_\delta^\infty \mathbb{P}_{i,\delta}(|Y_c| \geq \epsilon, Y_j \in hF \mid Y_i = h) \, \Lambda_i(\mathrm{d}h) \\
&= \int_\delta^\infty \mathbb{P}_{i,\delta}(|Y_c| \geq \epsilon \mid Y_i = h) \mathbb{P}_{i,\delta}(Y_j \in hF \mid Y_i = h) \, \Lambda_i(\mathrm{d}h) \\
&= \Lambda(\{|x_c| \geq \epsilon, x_i \geq \delta\}) \mathbb{P}(\xi^+_{(i,j)} \in F)
\end{aligned}$$

using (IV.4) and the assumed conditional independence. Dividing by $\Lambda(\{|x_c| \geq \epsilon\})$ yields

$$\begin{aligned}
\mathbb{P}(\xi^+_{(i,j)} \in F) \mathbb{P}_{c,\epsilon}(Y_i \geq \delta) &= \mathbb{P}_{c,\epsilon}(Y_i \geq \delta, Y_j \in Y_i F) \\
&= \int_\delta^\infty \mathbb{P}_{c,\epsilon}(Y_j \in hF \mid Y_i = h) \, \mathbb{P}_{c,\epsilon}(Y_i \in \mathrm{d}h).
\end{aligned}$$

Via standard arguments we have now established (IV.7) for $\mathbb{P}_{c,\epsilon}(Y_i \in \cdot\,)$-almost all $h > 0$. For $h < 0$ we consider $\Lambda(\{|x_c| \geq \epsilon, x_i \leq -\delta, x_j \in |x_i|F\})$ and apply similar arguments. $\square$

For a fixed vertex $c \in V$ we define a directed tree $T^{(c)} = (V, E^{(c)})$ by letting $E^{(c)}$ consist of all edges in E which are pointing away from c in the original tree. For $i \in V$ we let $\mathrm{ph}_T(c,i)$ denote the set of edges in the unique path from c to i in $T^{(c)}$. If $i = c$ we have $\mathrm{ph}_T(c,i) = \emptyset$.

Assume that the random variables $\xi_e^{\pm}, e \in E^{(c)}$ are independent. For $\epsilon > 0$ we further let P be an independent random variable with distribution given by Λ_c restricted to $(-\infty, -\epsilon] \cup [\epsilon, \infty)$ and normalized. We can now define a random vector Z recursively by

$$Z_c = P \quad \text{and} \quad Z_j = |Z_i| \cdot \xi_{(i,j)}^{\mathrm{sgn}(Z_i)} \quad \text{for } j \neq c, \tag{IV.8}$$

where $i \in V$ is the unique vertex such that $(i,j) \in \mathrm{ph}_T(c,j)$.

Theorem IV.11. *The random vector Z defined in (IV.8) has distribution $\mathbb{P}_{c,\epsilon}$.*

This result tells us that we can 'build' a vector with distribution $\mathbb{P}_{c,\epsilon}$ by starting with P at vertex c and then expanding to the rest of the tree by multiplying with the random variables $\xi_e^{\pm}$.

If we assume that Λ is concentrated on $[0,\infty)^d$ we do not have to worry about the sign of Z_i since it will always be positive. Therefore it is possible to write down Z in a slightly simpler form. In this case we have that

$$Z_j = P \cdot \prod_{e \in \mathrm{ph}_T(c,j)} \xi_e,$$

where the empty product is defined to be 1. Here we use the simplified notation $\xi_e = \xi_e^+$. This is nearly identical to the formula given by Engelke and Volgushev (2022, Prop. 1).

We conclude the section with Proposition IV.12 below which tells us how the distributions of $\xi_{(i,j)}$ and $\xi_{(j,i)}$ are related. In other words, it describes what happens when we reverse the direction of an edge. Something which is relevant when the root of the directed tree $T^{(c)}$ is changed.

Proposition IV.12. *Assume that Λ has no mass outside $[0,\infty)^d$. For distinct $i, j \in V$ we have that*

$$\mathbb{P}(\xi_{(j,i)} \geq h) = \frac{m_i}{m_j} \mathbb{E}[\mathbf{1}_{\{1/\xi_{(i,j)} \geq h\}} \, \xi_{(i,j)}^{\alpha}], \qquad h > 0,$$

and

$$\mathbb{P}(\xi_{(j,i)} \neq 0) = \lim_{h \downarrow 0} \mathbb{P}(\xi_{(j,i)} \geq h) = \frac{m_i}{m_j} \mathbb{E}[\xi_{(i,j)}^{\alpha}],$$

where $m_i = \Lambda(\{x_i \geq 1\}), m_j = \Lambda(\{x_j \geq 1\})$.

We may combine this result with Theorem IV.11 to obtain the identity

$$\frac{m_h}{m_t} = \prod_{(u,v) \in \mathrm{ph}_T(h,t)} \frac{m_u}{m_v}, \qquad h, t \in V. \tag{IV.9}$$

This will be used in the proof of Proposition IV.13 below.

IV.6 Structure learning for trees

Assume that the conditional independence structure of the Lévy process $\mathbf{X}$ is given by a tree $T = (V, E)$. We are interested in learning the tree structure from discrete observations $X(k/n)$ where $n \in \mathbb{N}$ and $k = 1, \ldots, n$. We further assume that $\mathbf{X}$ is α-stable with $\alpha \in (0, 2)$.

We rely on the work of Engelke and Volgushev (2022, §4) who consider structure learning for extremal tree models. In extreme value theory it is standard to study dependence after normalization to unit Pareto marginals, yielding the measure Λ with the marginals $\Lambda(\{x_i \geq u\}) = u^{-1}$. Normalization of the Lévy measure is uncommon as it would drastically change the Lévy process. Hence we only assume stability and make no further restriction to identical marginals. Recall that the marginals of the $-\alpha$-homogeneous Lévy measure are given by

$$\overline{\Lambda}_i(u) := \Lambda(\{x_i \geq u\}) = m_i u^{-\alpha}, \qquad u > 0,$$

where $m_i = \Lambda(\{x_i \geq 1\})$. We mainly focus on the case when Λ is concentrated on $[0,\infty)^d$ and give comments about the general case at the end.

IV.6.1 Tree recovery from bivariate summary statistics

We begin with the theoretical foundations for learning the underlying tree T. Afterwards we discuss how the estimation is carried out in practice, culminating with Theorem IV.17 which provides a step-by-step procedure.

Mimicking the definition of the extremal correlation coefficient we let

$$\begin{aligned}\chi_{ij} :=& \frac{\Lambda(\{x_i \geq \overline{\Lambda}_i^{-1}(q), x_j \geq \overline{\Lambda}_j^{-1}(q)\})}{\Lambda(\{x_i \geq \overline{\Lambda}_i^{-1}(q)\})} \\ =& \Lambda(\{x_i \geq m_i^{1/\alpha}, x_j \geq m_j^{1/\alpha}\}),\end{aligned} \tag{IV.10}$$

which is independent of $q > 0$. The latter equality follows by homogeneity and the fact that $\Lambda(\{x_i \geq m_i^{1/\alpha}\}) = 1$. This is also the extremal correlation coefficient of the multivariate Pareto vector characterized by Λ. Note also that χ is symmetric, i.e. $\chi_{ij} = \chi_{ji}$. Due to homogeneity we may switch to strict inequalities in (IV.10) if desired.

The following result is crucial for showing that χ_{ij} can be used to recover the underlying tree T as the minimal spanning tree. Recall that $\mathrm{ph}_T(h,t)$ denotes the set of edges on the path from h to t in the directed tree $T^{(h)}$.

Proposition IV.13. *For any edge $(i,j) \in \mathrm{ph}_T(h,t)$ it holds that*

$$\chi_{ht} \leq \chi_{ij}. \tag{IV.11}$$

Note that because of symmetry (IV.11) holds for any undirected edge (i,j) on the path between h and t in the undirected tree.

Proof. From the definition of χ_{ht} and the factorial representation in Theorem IV.11 we find that

$$\chi_{ht} = m_h \int_{m_h^{1/\alpha}}^{\infty} \alpha x^{-\alpha-1} \mathbb{P}\Big(x \prod_{(u,v)\in \mathrm{ph}_T(h,t)} \xi_{(u,v)} \geq m_t^{1/\alpha}\Big)\, \mathrm{d}x.$$

Apply Tonelli's theorem and (IV.9) to obtain

$$\begin{aligned}\chi_{ht} &= m_h \mathbb{E}\Big[\max\Big(m_h^{1/\alpha}, m_t^{1/\alpha} / \prod_{(u,v)\in \mathrm{ph}_T(h,t)} \xi_{(u,v)}\Big)^{-\alpha}\Big] \\ &= \mathbb{E}\Big[\min\Big(1, \prod_{(u,v)\in \mathrm{ph}_T(h,t)} \frac{m_u}{m_v} \xi^{\alpha}_{(u,v)}\Big)\Big].\end{aligned}$$

According to Proposition IV.12 we have that

$$\mathbb{E}\Big[\frac{m_u}{m_v}\xi^{\alpha}_{(u,v)}\Big] = \mathbb{P}(\xi_{(v,u)} \neq 0) \leq 1,$$

and so it is left to apply the bound $\mathbb{E}[\min(1, Z_1 Z_2)] \le \mathbb{E}[\min(1, Z_1)]$ for independent Z_1 and Z_2 with $\mathbb{E}[Z_2] \le 1$.

□

The following theorem assumes that the inequality in (IV.11) is strict for all $(i,j) \in \mathrm{ph}_T(h,t)$ when $(h,t) \notin E$. A sufficient condition for this to hold is the following: For any $(i,j) \in E$ and every $\epsilon > 0$ we have $\mathbb{P}(\frac{m_i}{m_j}\xi_{(i,j)} > 1) > 0$ and $\mathbb{P}(1-\epsilon < \frac{m_i}{m_j}\xi_{(i,j)} < 1) > 0$. See Lemma IV.24 in §IV.A.1 for further details. This condition is quite similar to the one given by Hu et al. (2022, Prop. 2.5). It is satisfied if e.g. $\mathrm{supp}(\xi_{(i,j)}) = [0,\infty)$ for all $(i,j) \in E$. The latter is the condition given by Engelke and Volgushev (2022, just after Prop. 5).

Theorem IV.14. *Let g be a strictly decreasing function on $[0,1]$, and $\widehat{\chi}_{ij}^{(n)}$ be a consistent estimator of χ_{ij}. Assume also that the inequalities in (IV.11) are strict when $(h,t) \notin E$. Then the minimal spanning tree for the edge weights $g(\widehat{\chi}_{ij}^{(n)})$ coincides with T with probability tending to 1.*

Here one may use $g = -\log$ to guarantee the positivity of the weights if desired. This result readily follows using the arguments of Engelke and Volgushev (2022), see the proofs of Proposition 5 and Theorem 2 there. Importantly, the assumption of strict inequalities implies that for all $(i,j) \in E$ we have $\chi_{ij} > 0$, and so $\{i\} \perp \{j\}\ [\Lambda]$ cannot happen.

Remark IV.15. *It is important to note that homogeneity of Λ is crucial for the above theory in at least two ways. Firstly, it underlies the factorization identity leading to the inequalities for χ. Secondly, it allows for consistent estimation of χ_{ij} on a constant interval $[0,1]$. Indeed, the number of large jumps does not grow with n and hence we must rely on some structural assumption allowing for estimation of our summary statistic using small jumps.*

IV.6.2 Estimation of the coefficients

Initially we will assume that $\mathbf{X}$ is *strictly* α-stable. In §IV.6.4 we present an easy adaptation which covers the general case of stable processes.

For the estimation we use the increments $\Delta(n,1), \ldots, \Delta(n,n)$ given by

$$\Delta(n,k) = X(k/n) - X((k-1)/n).$$

Additionally for each $i \in V$ we let $\widehat{F}_i^{(n)}$ denote the empirical CDF of $\Delta_i(n,1), \ldots, \Delta_i(n,n)$. This is defined as

$$\widehat{F}_i^{(n)}(x) = \frac{1}{n+1}\sum_{k=1}^{n} \mathbf{1}_{\{\Delta_i(n,k) \le x\}}, \qquad x \in \mathbb{R}.$$

For $q \in (0,1)$ we define an estimator $\widehat{\chi}_{ij}^{(n)}(q)$ by

$$\widehat{\chi}_{ij}^{(n)}(q) = \frac{1}{n(1-q)}\sum_{k=1}^{n} \mathbf{1}_{\{\widehat{F}_i^{(n)}(\Delta_i(n,k))>q,\, \widehat{F}_j^{(n)}(\Delta_j(n,k))>q\}}\,. \tag{IV.12}$$

Note that computing $\widehat{\chi}_{ij}^{(n)}$ does not involve estimation of any other parameters even though the definition of χ_{ij} uses both m_i, m_j and α. Consistency of the estimator is provided by the following result.

Proposition IV.16. *Let (q_n) be a sequence in $(0,1)$ such that $q_n \to 1$ and $n(1-q_n) \to \infty$ as $n \to \infty$. Then*

$$\widehat{\chi}_{ij}^{(n)}(q_n) \xrightarrow{\mathbb{P}} \chi_{ij} \quad \textit{for any } i,j \in V.$$

We can combine all of the above into Theorem IV.17 which describes the two steps of the estimation.

Theorem IV.17. *Assume that the inequalities in (IV.11) are strict for $(h,t) \notin E$. Let (q_n) be a sequence of positive numbers satisfying $q_n \to 1$ and $n(1-q_n) \to \infty$. Then the following procedure correctly estimates the tree T with probability going to 1 as $n \to \infty$.*

1. *For each $i, j \in V$ compute $\widehat{\chi}_{ij}^{(n)}(q_n)$ according to (IV.12).*
2. *Calculate the edge weights $\widehat{\rho}_{ij}^{(n)} = -\log(\widehat{\chi}_{ij}^{(n)}(q_n))$ and compute the corresponding minimum spanning tree $\widehat{T}_n$.*

Above we require that $q_n \to 1$ as $n \to 0$ but we further assume that $1 - q_n$ does not vanish too quickly. However, for a given n this does not tell us how to pick q_n in good way. The simulation study in §IV.7.3 will illustrate how the recovery probability can depend on the choice of q_n.

A crucial ingredient in the proof of Proposition IV.16 is the following asymptotic behavior which follows from a classical convergence result (Sato, 1999, Cor. 8.9). For any $E \subseteq \mathbb{R}^d$ which is measurable and bounded away from zero

$$\mathbb{P}(X(h) \in E) \sim h\Lambda(E) \quad \text{as } h \downarrow 0. \tag{IV.13}$$

To prove the convergence we will consider the random vectors $Y(k) = X(k) - X(k-1)$ for $k \in \mathbb{N}$. Denote the empirical CDF of $Y_i(1), \ldots, Y_i(n)$ by $\widehat{G}_i^{(n)}$. By self-similarity of $\mathbf{X}$ we obtain the distributional identity

$$\widehat{\chi}_{ij}^{(n)}(q_n) \overset{d}{=} \frac{1}{n(1-q_n)} \sum_{k=1}^{n} \mathbf{1}_{\{\widehat{G}_i^{(n)}(Y_i(k)) > q_n, \widehat{G}_j^{(n)}(Y_j(k)) > q_n\}}.$$

Lemma IV.18. *For each $i \in V$ let G_i denote the CDF of $Y_i(1)$ and let (q_n) be a sequence as in Proposition IV.16. Then*

$$p_n = \mathbb{P}(G_i(Y_i(1)) > q_n, G_j(Y_j(1)) > q_n)) \sim (1-q_n)\chi_{ij} \quad \text{as } n \to \infty$$

for any $i, j \in V$.

Proof. First, let $h_{n,i} = G_i^{-1}(q_n)$ and note that $h_{n,i} \to \infty$ as $n \to \infty$. It follows from self-similarity and (IV.13) that

$$1 - q_n = \mathbb{P}(Y_i(1) > h_{n,i}) = \mathbb{P}(X_i(h_{n,i}^{-\alpha}) > 1) \sim h_{n,i}^{-\alpha} m_i \quad \text{as } n \to \infty. \tag{IV.14}$$

The probability of interest can be rewritten as

$$p_n = \mathbb{P}(X_i(h_{n,i}^{-\alpha}) > 1, X_j(h_{n,i}^{-\alpha}) > \tfrac{h_{n,j}}{h_{n,i}}).$$

Another application of (IV.13) yields the asymptotic equivalence

$$\begin{aligned} p'_n &= \mathbb{P}(X_i(h_{n,i}^{-\alpha}) > 1, X_j(h_{n,i}^{-\alpha}) > (\tfrac{m_j}{m_i})^{1/\alpha}) \\ &\sim h_{n,i}^{-\alpha} \Lambda(\{x_i > 1, x_j > (\tfrac{m_j}{m_i})^{1/\alpha}\}) \\ &\sim (1-q_n) \frac{\Lambda(\{x_i > 1, x_j > (\frac{m_j}{m_i})^{1/\alpha}\})}{\Lambda(\{x_i > 1\})} \\ &= (1-q_n)\chi_{ij}, \end{aligned}$$

where the two last steps use (IV.14) and homogeneity of Λ. Finally, for any $\epsilon > 0$ it follows from (IV.14) that $|(\frac{m_j}{m_i})^{1/\alpha} - \frac{h_{n,j}}{h_{n,i}}| \leq \epsilon$ for large enough n. Hence,

$$\begin{aligned} |p_n - p_n'| &\leq \mathbb{P}(X_j(h_{n,i}^{-\alpha}) \text{ is between } (\tfrac{m_j}{m_i})^{1/\alpha} \text{ and } \tfrac{h_{n,j}}{h_{n,i}}) \\ &\leq \mathbb{P}(X_j(h_{n,i}^{-\alpha}) \in [(\tfrac{m_j}{m_i})^{1/\alpha} - \epsilon, (\tfrac{m_j}{m_i})^{1/\alpha} + \epsilon]) \\ &\sim h_{n,i}^{-\alpha}\Lambda(\{x_j \in [(\tfrac{m_j}{m_i})^{1/\alpha} - \epsilon, (\tfrac{m_j}{m_i})^{1/\alpha} + \epsilon]\}) \\ &\sim (1-q_n)m_i^{-1}\Lambda(\{x_j \in [(\tfrac{m_j}{m_i})^{1/\alpha} - \epsilon, (\tfrac{m_j}{m_i})^{1/\alpha} + \epsilon]\}) \end{aligned}$$

for large enough n. This concludes the proof as $\Lambda(\{x_j \in [(\frac{m_j}{m_i})^{1/\alpha} - \epsilon, (\frac{m_j}{m_i})^{1/\alpha} + \epsilon]\})$ can be made arbitrarily small by choosing $\epsilon > 0$ small enough. □

Proving Proposition IV.16 is now mostly an exercise in applying the Glivenko–Cantelli theorem which states that for each $i \in V$

$$\sup_{x\in\mathbb{R}} |\widehat{G}_i^{(n)}(x) - G_i(x)| \to 0 \quad \text{almost surely,} \tag{IV.15}$$

see e.g. Kallenberg (2021, Prop. 5.24).

Proof of Proposition IV.16. For $n \in \mathbb{N}$ and $k \leq n$ we let

$$Z_{n,k} = \mathbf{1}_{\{\widehat{G}_i^{(n)}(Y_i(k))>q_n, \widehat{G}_j^{(n)}(Y_j(k))>q_n\}}\,.$$

For any $\epsilon > 0$ we use Chebyshev's inequality to obtain the estimate

$$\mathbb{P}(|\widehat{\chi}_{ij}^{(n)} - (1-q_n)^{-1}\mathbb{E}[Z_{n,1}]| > \epsilon) \leq \epsilon^{-2}\mathrm{Var}(\widehat{\chi}_{ij}^{(n)}).$$

It is sufficient to show that the variance on the right-hand side vanishes since combining Lemma IV.18 and (IV.15) yields the convergence $(1-q_n)^{-1}\mathbb{E}[Z_{n,1}] \to \chi_{ij}$. The variance can be written as

$$\mathrm{Var}(\widehat{\chi}_{ij}^{(n)}) = \frac{1}{n^2(1-q_n)^2}\sum_{k=1}^{n}\sum_{\ell=1}^{n}\mathrm{Cov}(Z_{n,k}, Z_{n,\ell}).$$

There are $n(n-1)$ off-diagonal terms (those with $k \neq \ell$) for which $(1-q_n)^{-2}\mathrm{Cov}(Z_{n,k}, Z_{n,\ell}) \to 0$ uniformly in k, ℓ. Here we use the asymptotics from Lemma IV.18. Since we further divide by n^2 the total contribution of these terms vanish in the limit. The sum of the n terms with $k = \ell$ does also vanish as $n \to \infty$. To see this note that $\mathrm{Var}(Z_{n,k})$ is approximately $p_n(1-p_n)$ by (IV.15), where p_n is as in Lemma IV.18. To conclude simply apply the asymptotics of p_n along with the assumption that $n(1-q_n) \to \infty$. □

IV.6.3 Allowing for jumps in all directions

Suppose now that Λ is not restricted to the positive orthant, and let the marginal measures be given by $\Lambda(\{\pm x_i \geq u\}) = m_i^{\pm}u^{-\alpha}$ for $u > 0$. We still assume that Λ is globally Markov with respect to a tree T. It is easy to see that conditional independence statements remain to be true for Λ restricted to any orthant, see also Engelke et al. (2022, Lem. A.5). Hence one may apply the above described method to a single orthant to recover the underlying tree T, given that no independence statements arise because of this restriction.

The above approach is based on a small portion of information only (jumps in the chosen orthant). Thus we propose to consider a combined summary statistic over the four quadrants:

$$\chi_{ij} = \chi_{ij}^{++} + \chi_{ij}^{+-} + \chi_{ij}^{-+} + \chi_{ij}^{--},$$

where $\chi_{ij}^{+-} = \Lambda(\{x_i \geq (m_i^+)^{1/\alpha}, -x_j \geq (m_j^-)^{1/\alpha}\})$ and the other quantities are defined similarly. Importantly, Proposition IV.13 continues to hold in this setting. This can be seen by summing up the inequalities for all 2^d orthants encoded by $(s_1, \ldots, s_d)$ with $s_i \in \{+, -\}$:

$$\Lambda(\{s_h x_h \geq (m_h^{s_h})^{1/\alpha}, s_t x_t \geq (m_t^{s_t})^{1/\alpha}, s_v x_v > 0 \;\forall v \neq h, t\})$$
$$\leq \Lambda(\{s_i x_i \geq (m_i^{s_i})^{1/\alpha}, s_j x_j \geq (m_j^{s_j})^{1/\alpha}, s_v x_v > 0 \;\forall v \neq i, j\}).$$

Next, a consistent estimator of χ_{ij} is obtained by summing up four estimators of the type (IV.12), corresponding to each quadrant. This estimator is then used in step 2. of the procedure in Theorem IV.17.

One way to obtain an estimator for χ_{ij}^{+-} is to take the estimator in (IV.12) with $\Delta_j(n,k)$ replaced by $-\Delta_j(n,k)$. Note that $\widehat{F}_j^{(n)}$ must then be the empirical distribution function of $-\Delta_j(n,1), \ldots, -\Delta_j(n,n)$. Equivalently, one can keep the original increments and simply replace the inequality $\widehat{F}_i^{(n)}(\Delta_j(n,k)) > q$ with $\widehat{F}_i^{(n)}(\Delta_j(n,k)) < 1-q$. The other estimators are obtained by similar modifications of $\widehat{\chi}_{ij}^{(n)}$.

IV.6.4 Generalizing to non-strict stability

As mentioned we also want a method for structure estimation when $\mathbf{X}$ is a general α-stable process. A well-known idea used by e.g. Zolotarev (1986, §4.3) is to replace in the increments by the 'increments of increments' $\Delta^{(2)}(n,1), \ldots, \Delta^{(2)}(n, \lfloor n/2 \rfloor)$ which are given by $\Delta^{(2)}(n,k) = \Delta(n,2k) - \Delta(n,2k-1)$. These can be seen as increments of a strictly α-stable process with Lévy measure $\tilde{\Lambda}$ given by $\tilde{\Lambda}(\mathrm{d}x) = \Lambda(\mathrm{d}x) + \Lambda(\mathrm{d}(-x))$. The two terms obviously have identical dependence structures. Hence, $\tilde{\Lambda}$ will have the same dependence structure. Thus, estimation of the underlying tree is carried out by simply performing the estimation procedure in Theorem IV.17 using $\Delta^{(2)}(n,1), \ldots, \Delta^{(2)}(n, \lfloor n/2 \rfloor)$ in place of the original increments. Notice that $\tilde{\Lambda}$ will have mass outside $[0,\infty)^d$ so one should do quadrant-wise estimation as explained in §IV.6.3.

IV.7 Examples and simulations

Until now we have taken a fairly theoretical approach with few examples. In this section we give an example of a parametric class of stable Lévy processes for which the conditional independence structure can be read off from a single matrix. We also discuss a method for approximate simulation of Lévy processes which is useful when one can simulate from the distributions $\mathbb{P}_{c,\epsilon}$. Finally we present a simulation study, demonstrating the estimation procedure from §IV.6.

IV.7.1 Hüsler–Reiss type jumps

Let Γ be a symmetric $d \times d$ strictly conditionally negative definite matrix Γ (i.e. $u^\top \Gamma u < 0$ for any $u \in \mathbb{R}^d \setminus \{0\}$ with $1^\top u = 0$) with $\mathrm{diag}(\Gamma) = 0$ and non-negative entries A d-dimensional random vector Y is said to be Hüsler–Reiss distributed with variogram Γ if it is concentrated on $\{x \geq 0, \|x\|_\infty \geq 1\}$ and

$$Y^{(c)} \stackrel{d}{=} P \cdot \exp(U - U_c - \tfrac{1}{2}\Gamma_{\cdot c}) \tag{IV.16}$$

for each $c \in V$, where $Y^{(c)}$ has the distribution of $Y \mid Y_c \geq 1$. Here P is a standard Pareto random variable and U is an independent centered normal vector with variogram Γ (i.e. $\mathbb{E}[(U_i - U_j)^2] = \Gamma_{ij}$ for any $i, j \in V$).

The Hüsler–Reiss distributions belong to the larger class of multivariate Pareto distributions. This means that if Y is Hüsler–Reiss distributed we have that

$$\mathbb{P}(Y \in \cdot\,) = \frac{\Lambda(\,\cdot\, \cap \{\|x\|_\infty \geq 1\})}{\Lambda(\{\|x\|_\infty \geq 1\})},$$

where Λ is the so-called exponent measure. This measure satisfies $\Lambda(\{0\}) = 0$, it is -1-homogeneous and $\Lambda(\{x_i \geq 1\})$ is the same for all $i \in V$. Because of the latter we may freely assume unit marginals, i.e. $\Lambda(\{x_i \geq 1\}) = 1$ for all i.

We want to extend the notion of Hüsler–Reiss distributions. Specifically we want to include any $-\alpha$-homogeneous measure (where $\alpha \in (0,2)$) and we want to generalize to different marginal masses. For $(\alpha, m_1, \ldots, m_d) \in (0,2) \times (0,\infty)^d$ we define

$$\tilde{\Lambda}(\{x_1 \geq u_1, \ldots, x_d \geq u_d\}) = \Lambda(\{x_1 \geq m_1^{-1}u_1^\alpha, \ldots, x_d \geq m_d^{-1}u_d^\alpha\}), \qquad u_1, \ldots, u_d > 0,$$

where Λ is a Hüsler–Reiss exponent measure as described above. The result is a $-\alpha$-homogeneous measure with $\tilde{\Lambda}(\{x_i \geq 1\}) = m_i$ for each i. Consider now the random vectors

$$Y^{(c)} \sim \frac{\Lambda(\,\cdot\, \cap \{x_c \geq m_c^{-1}\epsilon^\alpha\})}{\Lambda(\{x_c \geq m_c^{-1}\epsilon^\alpha\})} \quad \text{and} \quad \Delta^{(c)} \sim \frac{\tilde{\Lambda}(\,\cdot\, \cap \{x_c \geq \epsilon\})}{\tilde{\Lambda}(\{x_c \geq \epsilon\})}.$$

As in (IV.16) we can write $Y^{(c)} \overset{d}{=} P \cdot \exp(U - U_c - \frac{1}{2}\Gamma_{\cdot c})$ where U is as before and P is now Pareto distributed with scale $m_c^{-1}\epsilon^\alpha$ and shape 1. Moreover, we note that $\Delta^{(c)} \overset{d}{=} (mY^{(c)})^{1/\alpha}$. This readily leads to a similar representation for $\Delta^{(c)}$. Namely,

$$\Delta^{(c)} \overset{d}{=} \tilde{P} \cdot \exp\{\tfrac{1}{\alpha}(U + \log(m) - (U_c + \log(m_c)) - \tfrac{1}{2}\Gamma_{\cdot c})\},$$

where $\tilde{P} = (m_c P)^{1/\alpha}$. Note that $\tilde{P}$ is independent of U and that $\tilde{P}$ is Pareto distributed with scale ϵ and shape α. If $\tilde{\Lambda}$ is the Lévy measure of $\mathbf{X}$ we say that the process has Hüsler–Reiss type jumps with stability index α, variogram Γ and marginals $m_1, \ldots, m_d$.

Since $\tilde{\Lambda}$ has the same conditional independence structure as Λ we conclude that it depends only on the variogram matrix Γ and not on the remaining parameters $\alpha, m_1, \ldots, m_d$. This allows us to directly apply results known from extreme value theory. For each $k \in V$ we define the matrix

$$\Sigma^{(k)} = \tfrac{1}{2}\{\Gamma_{ik} + \Gamma_{jk} - \Gamma_{ij}\}_{i,j \neq k}.$$

This is a positive definite matrix and we denote its inverse by $\Theta^{(k)}$. These matrices are used to determine the conditional independence structure. To be precise, for $i \neq j$ and $k \in V \setminus \{i,j\}$ we have that

$$\{i\} \perp \{j\} \mid V \setminus \{i,j\}\,[\Lambda] \quad \Leftrightarrow \quad \Theta_{ij}^{(k)} = 0,$$

see Engelke and Hitz (2020, Prop. 3). Since $\Theta_{ij}^{(k)} = \Theta_{ij}^{(k')}$ for any $k, k' \in V \setminus \{i,j\}$ one may define the so-called Hüsler–Reiss precision matrix, see Hentschel et al. (2022, Def. 3.2), by

$$\Theta_{ij} = \Theta_{ij}^{(k)} \quad \text{for some } k \in V \setminus \{i,j\}.$$

We see that $\{i\} \perp \{j\} \mid V \setminus \{i,j\}\,[\Lambda]$ if and only if $\Theta_{ij} = 0$.

Remark IV.19 (Symmetry)**.** *Assume that* $\mathbf{X}$ *has Hüsler–Reiss type jumps with parameters* $\alpha, \Gamma, m_1, \ldots, m_d$. *Then for each* $c \in V$ *the random vector* $\xi^{(c)} = \xi^{(c,+)}$ *from Lemma IV.8 can be represented as*

$$\xi^{(c)} \overset{d}{=} \exp\{\tfrac{1}{\alpha}(U + \log(m) - (U_c + \log(m_c)) - \tfrac{1}{2}\Gamma_{\cdot c})\}, \tag{IV.17}$$

where U is any centered normal random vector with variogram Γ. Recalling the notation $\xi_{(i,j)} = \xi_j^{(i)}$ we see that

$$\xi_{(j,i)} \overset{d}{=} \big(\tfrac{m_i}{m_j}\big)^{2/\alpha} \xi_{(i,j)}.$$

This shows that Hüsler–Reiss type jumps have built-in symmetry between the distributions of $\xi_{(i,j)}$ and $\xi_{(j,i)}$ which is simpler than the general result in Proposition IV.12.

IV.7.2 Approximate simulation

Consider a Lévy process $\mathbf{X}$ with characteristic function

$$\mathbb{E}[e^{i\langle u, X(1)\rangle}] = \exp\Big(i\langle u, \gamma\rangle - \frac{1}{2}\langle u, \Sigma u\rangle + \int_{\mathbb{R}^d} e^{i\langle u,x\rangle} - 1 - i\langle u, x\rangle\, \mathbf{1}_{\{\|x\|_\infty \le 1\}}\, \Lambda(\mathrm{d}x)\Big), \qquad u \in \mathbb{R}^d.$$

Exact simulation of $\mathbf{X}$ at fixed times $0 \le t_0 < t_1 < \ldots < t_n$ is easy if one can simulate increments of the process. When this is not possible it is typical to simulate an approximation of the process. A simple approach for approximating the jump part of $\mathbf{X}$ consists of picking a small $\epsilon > 0$ and simulating a compound Poisson process $\mathbf{N}^{(\epsilon)}$ with Lévy measure given by $\Lambda(\,\cdot \cap \{\|x\|_\infty \ge \epsilon\})$. This involves simulating from $\mathbb{P}_\epsilon = \Lambda(\,\cdot \cap \{\|x\|_\infty \ge \epsilon\})/\Lambda(\{\|x\|_\infty \ge \epsilon\})$, i.e. the jump distribution of $\mathbf{N}^{(\epsilon)}$. It might not be clear how to do this but sometimes it is easier to simulate from the distributions $\mathbb{P}_{v,\epsilon} = \Lambda(\,\cdot \cap \{|x_v| \ge \epsilon\})/\Lambda(\{|x_v| \ge \epsilon\})$. An example where this is clearly the case is when $\mathbf{X}$ has Hüsler–Reiss type jumps. In general, the following rejection sampling procedure lets us simulate from $\mathbb{P}_\epsilon$ if we can simulate from $\mathbb{P}_{v,\epsilon}$ for each $v \in V$.

Proposition IV.20. *The following procedure generates a random vector with distribution $\mathbb{P}_\epsilon$.*

1. *Simulate a V-valued random variable I with $\mathbb{P}(I = v)$ proportional to $m_v = \Lambda(\{|x_v| \ge 1\})$.*
2. *Simulate a random vector Z such that $(Z \mid I = v) \sim \mathbb{P}_{v,\epsilon}$ for any $v \in V$.*
3. *Accept Z with probability $(\#\{v \in V \mid |Z_v| \ge \epsilon\})^{-1}$. Otherwise, repeat from step 1.*

Approximation of the jump part using only a compound Poisson process is not ideal since it ignores the small jump activity. To address this issue one may (under certain assumptions) use a Brownian motion to approximate the small jumps. Here we assume for simplicity that $\mathbf{X}$ is α-stable for some $\alpha \in (0, 2)$. For a fixed $\epsilon > 0$ we then define

$$\gamma^{(\epsilon)} = \gamma - \int_{\{\epsilon \le \|x\|_\infty \le 1\}} x\, \Lambda(\mathrm{d}x), \qquad S = \int_{\{\|x\|_\infty \le 1\}} x x^\top\, \Lambda(\mathrm{d}x).$$

The approximating process $\mathbf{X}^{(\epsilon)}$ is given by

$$X^{(\epsilon)}(t) = \gamma^{(\epsilon)} t + \epsilon^{1-\alpha/2} W(t) + N^{(\epsilon)}(t),$$

where $\mathbf{W}$ is a centered Brownian motion with covariance matrix S and $\mathbf{N}^{(\epsilon)}$ is as above. Now, when S is invertible Cohen and Rosiński (2007, Thm. 3.1) proved the convergence $\mathbf{X}^{(\epsilon)} \overset{d}{\to} \mathbf{X}$ as $\epsilon \downarrow 0$.

IV.7.3 Simulation study

A simple simulation study has been conducted with the purpose of demonstrating how the estimation procedure described above recovers the true tree structure with large probability. At the same time we also investigate how this probability depends on the threshold q. All of this is done using the `R` programming language and the code is available at https://github.com/jakobdt/levy_graphical_models.git. The actual tree is depicted in Figure IV.7 below.

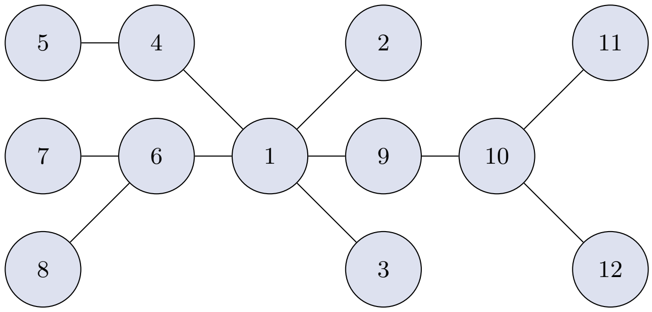


Figure IV.7: The tree T used for the simulation.

For $n \in \mathbb{N}$ we simulate $X(1/n), \ldots, X(1)$ approximately using the procedure discussed in §IV.7.2. Three different situations are considered:

(A) Λ is concentrated on $[0, \infty)^d$.

(B) Λ is again concentrated on $[0, \infty)^d$ but the dependence between components is weaker than in (A).

(C) Λ has mass on all orthants in $\mathbb{R}^d$.

The exact details of the models and the simulation are given below. From the simulated realisations we perform the estimation procedure described above. In each scenario we use the approach from §IV.6.3, meaning that in (A) and (B) we do not use knowledge about Λ having no mass outside $[0, \infty)^d$.

We estimate the probability of recovering the underlying tree by running the experiment 1000 times for each $n = 200, 500, 1000, 2000, 5000$ and $q = \frac{1}{2} + k \cdot 10^{-2}$ with $k = 0, \ldots, 49$. The results are presented in Figure IV.8 below. In (A) we find that $n \geq 500$ observations are sufficient for high recovery probability. Unsurprisingly the weaker dependence in (B) means that more observations are needed. We further see that a larger n is required in the two-sided scenario (C). This is to be expected since the fact that the jumps are divided among far more orthants (2^d versus only 1).

In all three scenarios we clearly see that the assumptions on the sequence (q_n) in Proposition IV.16 are necessary. Indeed, the probability of recovering T drops when q is far from 1 or too close to 1.

The simulated processes have stability index $\alpha = 3/2$. The simulation is performed using the approximate simulation method discussed in §IV.7.2. We use the threshold $\epsilon = 10^{-3}$ and the Lévy measure Λ is scaled such that $\Lambda(\{\|x\|_\infty \geq \epsilon\}) = 10\,000$. We approximate the quantities $\gamma^{(\epsilon)}$ and S using numerical methods. In all simulations the processes are strictly stable. This is means that $\gamma = -\int_{\{\|x\|_\infty > 1\}} x \, \Lambda(\mathrm{d}x)$. In the three scenarios the models are given as follows.

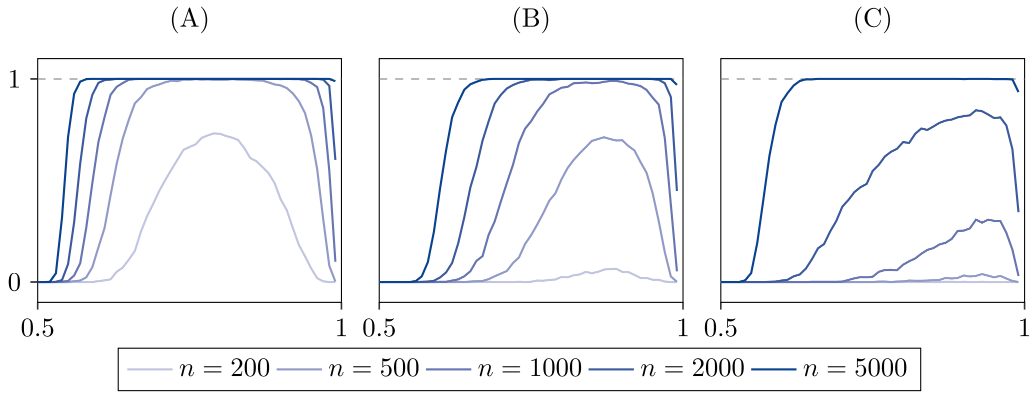


Figure IV.8: Estimated probabilities of recovering the correct tree as a function of $q \in [\frac{1}{2}, 1)$ for each of the three scenarios (A), (B), and (C).

(A) The process has Hüsler–Reiss type jumps with identical marginals and variogram Γ, where Γ_{ij} is given by the number of edges in the shortest path from from vertex i to vertex j in the tree T. That is, for $\Delta^{(c)} \sim \mathbb{P}_{c,\epsilon}$ we have the representation

$$\Delta^{(c)} \stackrel{d}{=} P \cdot \exp\{\tfrac{1}{\alpha}(U - U_c - \tfrac{1}{2}\Gamma_{\cdot c})\},$$

where U is a centered normal random vector with variogram Γ. This vector is constructed by letting $\{U_1\} \cup \{U_i - U_j \mid (i,j) \in E\}$ be a collection of independent standard normals.

(B) The process has Hüsler–Reiss type jumps with different marginals and the variogram is given by 4Γ, where Γ is the variogram from (A).

(C) Let $\Lambda^{(\mathrm{A})}$ denote the Lévy measure of the process in (A). In this scenario the Lévy measure Λ is a 'symmetric version' of $\Lambda^{(\mathrm{A})}$. That is, for any $s \in \{-1,1\}^d$ and $h \in [0,\infty)^d$,

$$\Lambda(\times_{i\in V}\{s_i x_i \geq h_i\}) = 2^{-d}\Lambda^{(\mathrm{A})}(\times_{i\in V}\{x_i \geq h_i\}).$$

Hence, if $\Delta^{(\mathrm{A})} \sim \mathbb{P}_\epsilon^{(\mathrm{A})}$ is a jump of the compound Poisson process simulated in (A), and σ is an independent random vector which is uniform in $\{-1,1\}^d$, then the jumps in this scenario can be represented as

$$\Delta_i = \sigma_i \Delta_i^{(\mathrm{A})}, \qquad i \in V.$$

IV.8 Application to stock prices

To further demonstrate the estimation method from §IV.6 we consider daily stock prices for several American companies during the period April 1st 2010 until December 31st 2015. Table IV.1 below contains all the stocks including their respective sectors. For the analysis we assume that the prices can be modelled by the exponential of a strictly stable Lévy process. Hence, the estimation procedure in is carried out using the log prices. The data and R code is available at `https://github.com/jakobdt/levy_graphical_models.git`.

Company	Ticker	Sector
Home Depot	HD	Consumer Discretionary
Target	TGT	Consumer Discretionary
The Coca-Cola Company	KO	Consumer Staples
PepsiCo Inc.	PEP	Consumer Staples
Chevron Corp.	CVX	Energy
Occidental Petroleum	OXY	Energy
JPMorgan Chase & Co.	JPM	Financials
Wells Fargo	WFC	Financials
FedEx Corporation	FDX	Industrials
United Parcel Service	UPS	Industrials
Apple Inc.	AAPL	Information Technology
Microsoft Corp.	MSFT	Information Technology

Table IV.1: Stocks included in the analysis.

IV.8.1 Results

For the estimation procedure we have to choose the parameter q. Based on Figure IV.12 we will use $q = 0.94$. This choice is further explained below. The tree in Figure IV.9 is the result of the estimation procedure. We see that much of the structure is rather intuitive. Indeed, companies from the same sector are generally located near each other.

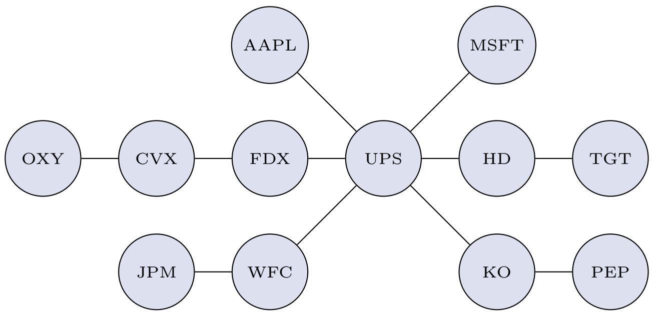


Figure IV.9: Result of the estimation procedure with $q = 0.94$.

We further choose to do the estimation above on subsamples of the data. More precisely, we randomly sample 754 data points (half of the original data) without replacement and estimate the tree. This is done 10 000 times and the color map in Figure IV.10 shows the proportion of times each edge was selected by the estimation procedure. To avoid putting too much importance on the choice of q we choose, for each subsample, to sample q uniformly in $[0.925, 0.95]$. Again, the choice of this range is explained by the observations in Figure IV.12 below. We observe that certain edges between companies from the same sector are selected almost every time.

If we denote the values displayed in Figure IV.10 by w_{ij} then we can compute the minimum spanning tree for the edge weights $-w_{ij}$. This tree is displayed in Figure IV.11 below. We observe that the tree is not identical to the one in Figure IV.9 but much of the general structure is the same. Indeed, eight out of eleven edges are preserved.

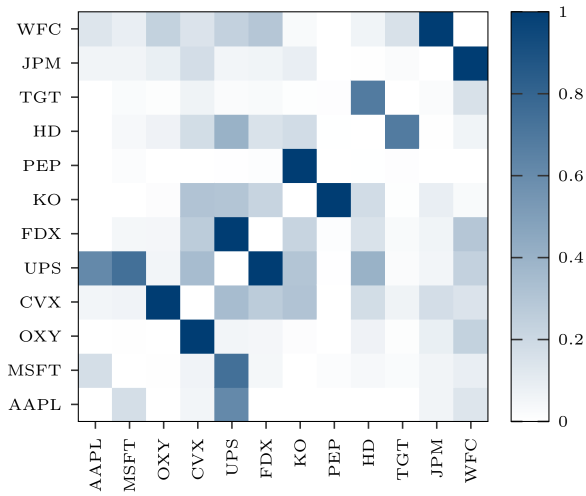


Figure IV.10: Result of running the estimation procedure on random subsamples.

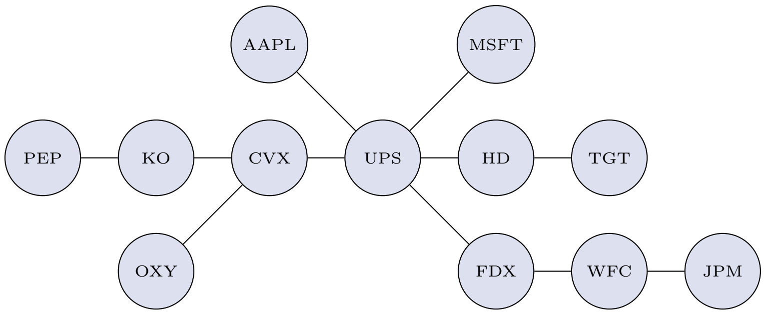


Figure IV.11: Result of running the estimation procedure on random subsamples.

IV.8.2 Choosing q

In order to decide on a value for the hyperparameter q we look at $-\log \widehat{\chi}_{ij}^{(n)}(q)$ as a function of q. Figure IV.12 below shows the calculated values for different values of q. Here we use $i = 1$ which corresponds to the AAPL ticker. We observe that the interval $[0.925, 0.95]$ is a somewhat stable region.

IV.8.3 Conclusion

This is a rather basic study but it shows that the method has potential applications to actual data. Indeed, the stronger dependence observed between companies from the same sector suggests that such an analysis is able to, at least partially, uncover the dependence structure.

Assuming that the dependence structure is given by a tree is of course quite strict. A natural next step would be to somehow extend to a larger class of graphs. Additionally, the process of log prices might not be jointly strictly stable. Even if one believes that each marginal is strictly stable it is not necessarily that easy to verify joint stability. Therefore, it would be sensible to study what can be done in the case where this is not satisfied.

It is reasonable to expect that we may have both positive and negative jumps. Looking at Figure IV.8 (C) we note that the number of observations in this study (approximately 1500) might not actually be sufficient to have a good chance of recovering the 'correct tree'. For this reason one might want to increase the sampling frequency to e.g. hourly rather than

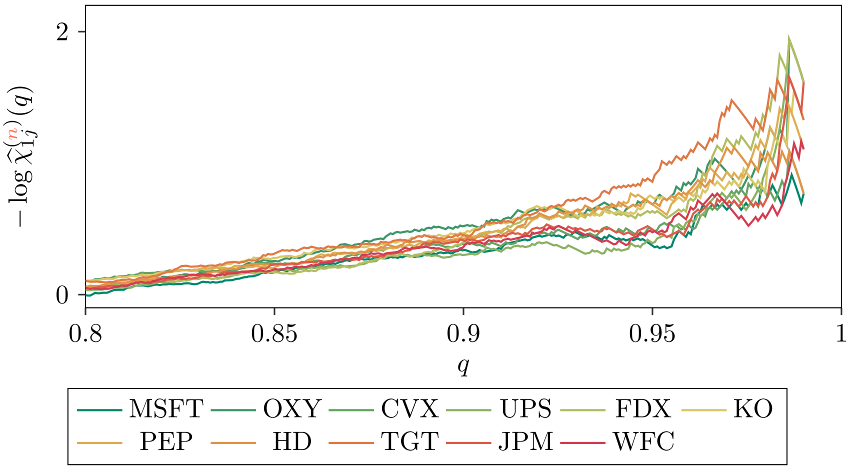


Figure IV.12: Estimation of $-\log \chi_{1j}$ for the stock data, where the vertex 1 corresponds to AAPL.

daily.

IV.9 Small-jump approximation

In this section we again assume that the Lévy process $\mathbf{X}$ is α-stable with $\alpha \in (0, 2)$. Above we discussed approximate simulation where one simulates a compound Poisson process containing the large jumps. The small jumps are then approximated with a suitable Brownian motion. Previously a jump was considered 'large' if its sup norm was above some threshold ϵ. Instead we will now fix a non-empty subset $U \subseteq V$ and say that a jump is large if it exceeds ϵ in absolute value in at least one component of U. Note that $U = V$ corresponds to the previous notion of large jumps.

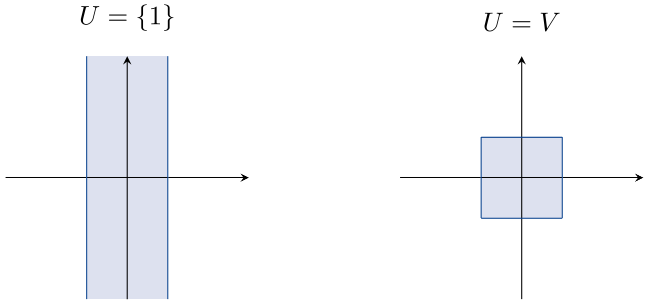


Figure IV.13: Jumps in the blue area are considered small.

We may represent $\mathbf{X}$ as an independent sum $\mathbf{X} = \mathbf{X}^{(\epsilon)} + \mathbf{W}^{(\epsilon)} + \gamma^{(\epsilon)}$, where $\gamma^{(\epsilon)}$ is a linear drift, $\mathbf{W}^{(\epsilon)}$ is a martingale and a Lévy process with Lévy measure given by Λ restricted to the set $\{|x_u| \leq \epsilon \, \forall u \in U\}$, and $\mathbf{X}^{(\epsilon)}$ is a compound Poisson process with jumps exceeding ϵ in absolute value in at least one component of U.

In the following we assume that

$$\int_{\{|x_u| \leq 1 \, \forall u \in U\}} x_i^2 \, \Lambda(\mathrm{d}x) < \infty \quad \text{for all } i \notin U, \tag{IV.18}$$

and note that this property readily extends to all $i \in V$ from the properties of Λ. It is important to note that this assumption implies that

$$\Lambda(\{x_U = 0_U\}) = 0,$$

because the Lévy measure $\Lambda^0_{V\setminus U}$ is $-\alpha$-homogeneous and integrates $x_{V\setminus U} \mapsto x_i^2$ for all $i \notin U$. Now we may define the matrix

$$\Sigma = \int_{\{|x_u|\leq 1\ \forall u\in U\}} xx^\top \Lambda(\mathrm{d}x), \tag{IV.19}$$

where existence of the integral on the right-hand side is ensured by the assumption above.

Lemma IV.21. *Assume* (IV.18)*. Then with* $a_\epsilon = \epsilon^{1-\alpha/2}$ *it holds that*

$$\mathbf{W}^{(\epsilon)}/a_\epsilon \xrightarrow{d} \mathbf{W} \quad \text{as } \epsilon \downarrow 0, \tag{IV.20}$$

where $\mathbf{W}$ *is a drift-less Brownian motion with covariance matrix* Σ *defined in* (IV.19)*.*

The next step is to establish conditions such that conditional independence for Λ transfers to the classical conditional independence for the limiting Brownian motion $\mathbf{W}$ in Lemma IV.21. We start with the simpler case of independence.

Lemma IV.22. *Assume that* $A \perp B\,[\Lambda]$ *for a partition* $A, B \subseteq V$*, and that* (IV.18) *is satisfied. Then the Brownian motions* $\mathbf{W}_A$ *and* $\mathbf{W}_B$ *are independent.*

Proof. Recall that [IV.A2] is satisfied because the process is stable. According to Engelke et al. (2022, Prop. 1) we have that $\Lambda(\{x_i \neq 0, x_j \neq 0\}) = 0$ for $i \in A, j \in B$. Hence, $\Sigma_{ij} = 0$ and the result follows. □

The case of conditional independence is much more subtle, and it requires several strong assumptions.

Theorem IV.23. *Assume that* U *satisfies* (IV.18)*, and that* $A \perp B \mid C\,[\Lambda]$ *for a partition* $A, B, C \subseteq V$*. Moreover, assume that*

- $|C| = 1$.
- $\Lambda(\{x_C < 0\}) = 0$.
- U *is a subset of either* $A \cup C$ *or* $B \cup C$*.*
- Σ *defined in* (IV.19) *is invertible.*

Then $(\Sigma)^{-1}_{ij} = 0$ *for all* $i \in A, j \in B$*, and hence* $W_A(1) \perp\!\!\!\perp W_B(1) \mid W_C(1)$*.*

IV.A Proofs and other technical details

IV.A.1 A sufficient condition

We are in the setting of §IV.6. In particular, we assume that Λ satisfies the global Markov property with respect to a Tree $T = (V, E)$.

Lemma IV.24. *Assume that for all* $(i,j) \in E$ *every* $\epsilon > 0$ *we have*

$$\mathbb{P}(\frac{m_i}{m_j}\xi^{\alpha}_{(i,j)} > 1) > 0 \quad \textit{and} \quad \mathbb{P}(1-\epsilon < \frac{m_i}{m_j}\xi^{\alpha}_{(i,j)} < 1) > 0. \tag{IV.21}$$

Then for $h, t \in V$ *with* $(h,t) \notin E$ *we have the inequality*

$$\chi_{ht} < \chi_{ij}$$

for all $(i,j) \in \mathrm{ph}_T(h,t)$.

Proof. In the proof of Prop IV.13 we saw that

$$\chi_{ht} = \mathbb{E}\Big[\min\Big(1, \prod_{(u,v)\in \mathrm{ph}_T(h,t)} \frac{m_u}{m_v}\xi^{\alpha}_{(u,v)}\Big)\Big]. \tag{IV.22}$$

It remains to prove that $\mathbb{E}[\min(1, Z_1 Z_2)] < \mathbb{E}[\min(1, Z_1)]$, where

$$Z_1 = \frac{m_i}{m_j}\xi^{\alpha}_{(i,j)} \quad \text{and} \quad Z_2 = \prod_{\substack{(u,v)\in \mathrm{ph}_T(h,t)\\ (u,v)\neq(i,j)}} \frac{m_u}{m_v}\xi^{\alpha}_{(u,v)}.$$

Independence of Z_1 and Z_2 allows us to write

$$\mathbb{E}[\min(1, Z_1 Z_2)] = \int_{[0,\infty)} \mathbb{E}[\min(1, z_1 Z_2)]\, \mathbb{P}_{Z_1}(\mathrm{d}z_1).$$

Generally, Jensen's inequality tells us that $\mathbb{E}[\min(1, z_1 Z_2)] \leq \min(1, z_1\mathbb{E}[Z_2])$. The first assumption of the lemma ensures the existence of $\epsilon > 0$ such that $\mathbb{P}(Z_2 \in (1/z_1, \infty)) > 0$ for all $z_1 \in (1-\epsilon, 1)$. For such z_1 we have

$$\begin{aligned}
\mathbb{E}[\min(1, z_1 Z_2)] &= \int_{[0,1/z_1]} z_1 z_2\, \mathbb{P}_{Z_2}(\mathrm{d}z_2) + \int_{(1/z_1,\infty)} 1\, \mathbb{P}_{Z_2}(\mathrm{d}z_2) \\
&< \int_{[0,\infty)} z_1 z_2\, \mathbb{P}_{Z_2}(\mathrm{d}z_2) \\
&= z_1 \mathbb{E}[Z_2] \\
&= \min(1, z_1\mathbb{E}[Z_2]),
\end{aligned}$$

where the final identity uses $\mathbb{E}[Z_2] \leq 1$ which follows from Proposition IV.12. Using the second assumption we conclude that

$$\mathbb{E}[\min(1, Z_1 Z_2)] < \int_{[0,\infty)} \min(1, z_1\mathbb{E}[Z_2])\, \mathbb{P}_{Z_1}(\mathrm{d}z_1) \leq \mathbb{E}[\min(1, Z_1)],$$

where we again used $\mathbb{E}[Z_2] \leq 1$. □

IV.A.2 Decompositions and conditioning

The Lévy process $\mathbf{X}$ may be represented by its Lévy–Itô decomposition $\mathbf{X} = \mathbf{J} + \mathbf{W}$, where $\mathbf{J}$ is a Lévy process with characteristic triplet $(0, 0, \Lambda)$, and $\mathbf{W}$ is a Brownian motion with drift γ and covariance matrix Σ which is independent of $\mathbf{J}$. Note that $\mathbf{J}$ contains all the jumps of $\mathbf{X}$. For $C \subseteq \{1, \ldots, d\}$ the processes $\mathbf{J}_C, \mathbf{W}_C$ may be obtained as almost sure limits of processes $\mathbf{J}^{(n)}_C, \mathbf{W}^{(n)}_C$ created from $\mathbf{X}_C$, see e.g. Applebaum (2009, §2.4). Thus, $\mathbf{J}_C, \mathbf{W}_C$ are $\bar{\sigma}(\mathbf{X}_C)$-measurable, where $\bar{\sigma}(\mathbf{X}_C)$ denotes the $\mathbb{P}$-completion of $\sigma(\mathbf{X}_C)$. For $F \in \sigma(\mathbf{J}_C, \mathbf{W}_C)$

we may therefore write $F = F' \cup N$, where N is a null-set and $F' \in \sigma(\mathbf{X}_C)$. For an integrable random variable Z one immediately finds that

$$\mathbb{E}[\mathbb{E}[Z \mid \mathbf{X}_C]\,\mathbf{1}_F] = \mathbb{E}[Z\,\mathbf{1}_F].$$

Since $\mathbb{E}[Z \mid \mathbf{X}_C]$ is $\sigma(\mathbf{J}_C, \mathbf{W}_C)$-measurable it follows that

$$\mathbb{E}[Z \mid \mathbf{X}_C] = \mathbb{E}[Z \mid \mathbf{J}_C, \mathbf{W}_C]$$

almost surely.

In addition to the Lévy–Itô decomposition we may represent $\mathbf{X}$ in other ways. Suppose for simplicity that $\mathbf{X}$ is a Lévy process with triplet $(\gamma, 0, \Lambda)$. Define $\Lambda^{(1)} = \Lambda(\,\cdot\, \cap \{x_C = 0_C\})$ and $\Lambda^{(2)} = \Lambda(\,\cdot\, \cap \{x_C \neq 0_C\})$. Then $\mathbf{X}$ may be represented as the independent sum of two Lévy processes $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}$ with triplets $(\gamma, 0, \Lambda^{(1)})$ and $(0, 0, \Lambda^{(2)})$. The processes $\mathbf{X}^{(1)}_C, \mathbf{X}^{(2)}_C$ are constructed from $\mathbf{X}_C$ and with arguments similar to the above we find that

$$\mathbb{E}[Z \mid \mathbf{X}_C] = \mathbb{E}[Z \mid \mathbf{X}^{(1)}_C, \mathbf{X}^{(2)}_C]$$

almost surely for any integrable random variable Z.

IV.A.3 Proof of Proposition IV.2

Assume that $X_A(t) \perp\!\!\!\perp X_B(t) \mid X_C(t)$ for all $t \geq 0$. We need to show that

$$\{X_A(t_i)\}_{i=1}^k \perp\!\!\!\perp \{X_B(t_i)\}_{i=1}^k \mid \mathbf{X}_C \tag{IV.23}$$

for any $k \in \mathbb{N}$ and $0 = t_1 < \ldots < t_k$.

For each $n \geq k$ we let $S^{(n)} = \{s^{(n)}_j\}_{j=1}^n$ be a collection of time points such that

- For any $n \geq k$ and $j = 1, \ldots, n-1$ we have $0 \leq s^{(n)}_j < s^{(n)}_{j+1} \leq t_k$.
- For any $n \geq k$ and $i = 1, \ldots, k$ we have $t_i \in S^{(n)}$.
- The sequence of sets $(S^{(n)})$ is increasing.
- There is the convergence $\max_{j=1,\ldots,n-1}(s^{(n)}_{j+1} - s^{(n)}_j) \to 0$.

Now, for each $n \geq k$ and $j = 1, \ldots, n-1$ we let $I^{(n)}(j) = X(s^{(n)}_{j+1}) - X(s^{(n)}_j)$ denote the increment of $\mathbf{X}$ between times $s^{(n)}_j$ and $s^{(n)}_{j+1}$. Using stationarity and independence of the increments along with the initial conditional independence assumption we deduce that

$$\{I^{(n)}_A(j)\}_{j=1}^{n-1} \perp\!\!\!\perp \{I^{(n)}_B(j)\}_{j=1}^{n-1} \mid \{I^{(n)}_C(j)\}_{j=1}^{n-1}.$$

From the increments we may construct $\{X_A(t_i)\}_{i=1}^k$ and $\{X_B(t_i)\}_{i=1}^k$, and we therefore have that

$$\{X_A(t_i)\}_{i=1}^k \perp\!\!\!\perp \{X_B(t_i)\}_{i=1}^k \mid \{I^{(n)}_C(j)\}_{j=1}^{n-1} \quad \text{for all} \quad n \geq k.$$

It is sufficient to prove (IV.23) with $\mathbf{X}_C$ replaced by $(X_C(s))_{0 \leq s \leq t_k}$ since conditioning on the process $(X_C(t_k + s) - X_C(t_k))_{s \geq 0}$ is easily added using independence. By the above conditional independence it is enough to show the identity

$$\sigma(X_C(s), 0 \leq s \leq t_k) = \mathcal{G},$$

where $\mathcal{G} = \cup_{n \geq k} \sigma(I^{(n)}_C(j), j = 1, \ldots, n-1)$. The inclusion $\supseteq$ is obvious. For the other inclusion it suffices to show that $X_C(s)$ is the limit of a sequence of $\mathcal{G}$-measurable random vectors for any $s \in [0, t_k]$. For each $n \geq k$ we pick an index j_n such that the sequence $(s^{(n)}_{j_n})$ converges to s from the right. Then $X_C(s^{(n)}_{j_n})$ converges to $X_C(s)$, and since each $X_C(s^{(n)}_{j_n})$ is $\mathcal{G}$-measurable we are done. □

IV.A.4 Proof of Proposition IV.3

Lemma IV.25. *Let $(X_n), (Y_n)$ be sequences of random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, taking values in Polish spaces S_X, S_Y, and assume that there exist random variables X, Y, Z such that $X_n \to X$ a.s., $Y_n \to Y$ a.s. and $X_n \perp\!\!\!\perp Y_n \mid Z$ for all $n \geq 1$. Then $X \perp\!\!\!\perp Y \mid Z$.*

Proof. The collection of random variables $((X_n), (Y_n), X, Y)$ takes values in a Polish space, thus establishing the existence of a regular conditional distribution given Z. Denoting this probability kernel by μ we note that $\mathbb{P}$-almost surely

$$\mu(Z, \{x_n \in A, y_n \in B\}) = \mu(Z, \{x_n \in A\}) \cdot \mu(Z, \{y_n \in B\})$$

for all $A \in \mathcal{B}(S_X), B \in \mathcal{B}(S_Y)$, where e.g. $\{x_n \in A\} = \{((x_m), (y_m), x, y) \in S_X^{\mathbb{N}} \times S_Y^{\mathbb{N}} \times S_X \times S_Y \mid x_n \in A\}$. Furthermore, $\mu(Z, \{x_n \to x, y_n \to y\}) = 1$ almost surely. Now, since independence is preserved under almost sure convergence the result follows. □

Proof of Proposition IV.3. Assume that $\mathbf{J}_A \perp\!\!\!\perp \mathbf{J}_B \mid \mathbf{J}_C$ and $\mathbf{W}_A \perp\!\!\!\perp \mathbf{W}_B \mid \mathbf{W}_C$. Recall from §IV.A.2 that conditioning on $\mathbf{X}_C$ is the same as conditioning on $(\mathbf{J}_C, \mathbf{W}_C)$. We combine this with the independence of $\mathbf{J}$ and $\mathbf{W}$ to obtain the identity

$$\begin{aligned}&\mathbb{E}[\mathbf{1}_{E_A \times F_A}(\mathbf{J}_A, \mathbf{W}_A)\, \mathbf{1}_{E_B \times F_B}(\mathbf{J}_B, \mathbf{W}_B) \mid \mathbf{X}_C] \\ &\quad = \mathbb{E}[\mathbf{1}_{E_A \times E_B}(\mathbf{J}_A, \mathbf{J}_B) \mid \mathbf{J}_C]\mathbb{E}[\mathbf{1}_{F_A \times F_B}(\mathbf{W}_A, \mathbf{W}_B) \mid \mathbf{W}_C]\end{aligned}$$

for any Borel sets E_A, E_B, F_A, F_B. The conditional expectations on the right-hand side factorize due to the assumed conditional independence. Applying the above identity twice with appropriately chosen E_A, E_B, F_A, F_B yields

$$\begin{aligned}&\mathbb{E}[\mathbf{1}_{E_A \times F_A}(\mathbf{J}_A, \mathbf{W}_A)\, \mathbf{1}_{E_B \times F_B}(\mathbf{J}_B, \mathbf{W}_B) \mid \mathbf{X}_C] \\ &\quad = \mathbb{E}[\mathbf{1}_{E_A \times F_A}(\mathbf{J}_A, \mathbf{W}_A) \mid \mathbf{X}_C]\mathbb{E}[\mathbf{1}_{E_B \times F_B}(\mathbf{J}_B, \mathbf{W}_B) \mid \mathbf{X}_C].\end{aligned}$$

Hence, the pairs $(\mathbf{J}_A, \mathbf{W}_A)$ and $(\mathbf{J}_B, \mathbf{W}_B)$ are conditionally independent given $\mathbf{X}_C$. The stated conditional independence follows immediately.

Assume instead that $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C$. The processes $\mathbf{J}_A, \mathbf{W}_A$ can be constructed as almost sure limits of functions of $\mathbf{X}_A$, and similarly for the B-component. According to Lemma IV.25 we then have

$$\mathbf{J}_A \perp\!\!\!\perp \mathbf{J}_B \mid \mathbf{X}_C \quad \text{and} \quad \mathbf{W}_A \perp\!\!\!\perp \mathbf{W}_B \mid \mathbf{X}_C.$$

As previously discussed, conditioning on $\mathbf{X}_C$ is the same as conditioning on the pair $(\mathbf{J}_C, \mathbf{W}_C)$. Now the claimed conditional independence follows by since $\mathbf{J}$ and $\mathbf{W}$ are independent. □

IV.A.5 Proof of Proposition IV.4

For any $T > 0$ we note that conditioning on $\mathbf{X}_C$ is the same as conditioning on the pair $\{\mathbf{X}_C^T, (X_C(T+t) - X_C(T))_{t \geq 0}\}$. Here we use the fact that $\mathbf{X}$ is a.s. continuous at time T.

Assume that $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C$ and let $T > 0$. Then $\mathbf{X}_A^T \perp\!\!\!\perp \mathbf{X}_B^T \mid \mathbf{X}_C$. Since $\mathbf{X}^T$ is independent of $(X_C(T+t) - X_C(T))_{t \geq 0}$ we also have conditional independence given just $\mathbf{X}_C^T$.

For the opposite implication we assume that $\mathbf{X}_A^T \perp\!\!\!\perp \mathbf{X}_B^T \mid \mathbf{X}_C^T$ for all $T > 0$. For $k \in \mathbb{N}$ and $0 \leq t_1 < \ldots < t_k$ we pick $T > t_k$ and note that $X_A(t_i) = X_A^T(t_i)$ for all $i = 1, \ldots, k$ (and similarly for the B-component). Hence, $\{X_A(t_i)\}_{i=1}^k \perp\!\!\!\perp \{X_B(t_i)\}_{i=1}^k \mid \mathbf{X}_C^T$. Finally we employ independence to further condition on the process $(X_C(T+t) - X_C(T))_{t \geq 0}$. □

IV.A.6 Proof of Theorem IV.7

Part 1: Assume that $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C$. Let $c \in C$ and $\epsilon > 0$ We may write $\mathbf{X}$ as the independent sum $\mathbf{X} = \mathbf{X}^{\geq\epsilon} + \mathbf{X}^{<\epsilon}$, where $\mathbf{X}^{\geq\epsilon}$ is a compound Poisson process consisting of the jumps of $\mathbf{X}$ in $\{|x_c| \geq \epsilon\}$. Then $\mathbf{X}_A^{\geq\epsilon}$ is a function of $\mathbf{X}_A, \mathbf{X}_C$ and similarly for $\mathbf{X}_B^{\geq\epsilon}$. It follows that $\mathbf{X}_A^{\geq\epsilon} \perp\!\!\!\perp \mathbf{X}_B^{\geq\epsilon} \mid \mathbf{X}_C$. Since $\mathbf{X}_C^{\geq\epsilon}, \mathbf{X}_C^{<\epsilon}$ are functions of $\mathbf{X}_C$ we have $\mathbf{X}_A^{\geq\epsilon} \perp\!\!\!\perp \mathbf{X}_B^{\geq\epsilon} \mid \mathbf{X}_C^{\geq\epsilon}, \mathbf{X}_C^{<\epsilon}$ which is the same as $\mathbf{X}_A^{\geq\epsilon} \perp\!\!\!\perp \mathbf{X}_B^{\geq\epsilon} \mid \mathbf{X}_C^{\geq\epsilon}$ by independence. We may write

$$X^{\geq\epsilon}(t) = \sum_{n=1}^{N(t)} Y(n), \qquad t \geq 0,$$

where $\mathbf{N}$ is a Poisson process with rate $\Lambda(\{|x_c| \geq \epsilon\})$ and $(Y(n))$ is a sequence of i.i.d. random vectors independent of $\mathbf{N}$ with $Y(1) \sim \mathbb{P}_{c,\epsilon} = \Lambda(\,\cdot\, \cap \{|x_c| \geq \epsilon\})/\Lambda(\{|x_c| \geq \epsilon\})$. Since $\mathbf{N}$ and $(Y(n))$ can be obtained from $\mathbf{X}^{\geq\epsilon}$ we find that $Y_A(1) \perp\!\!\!\perp Y_B(1) \mid Y_C(1)$. According to Engelke et al. (2022, Thm. 4.1) it remains to prove that $A \perp B\,[\Lambda^0_{A\cup B}]$. We consider a different decomposition $\mathbf{X} = \mathbf{X}^{=0} + \mathbf{X}^{\neq 0}$, where $\mathbf{X}^{=0}, \mathbf{X}^{\neq 0}$ are independent Lévy processes with Lévy measures given by Λ restricted to $\{x_C = 0_C\}$ and $\{x_C \neq 0_C\}$ respectively. The process $\mathbf{X}^{=0}$ can be constructed as an almost sure limit of $\sigma(\mathbf{X})$-measurable processes (we refer to the discussion in §IV.A.2). By applying Lemma IV.25 we find that $\mathbf{X}_A^{=0} \perp\!\!\!\perp \mathbf{X}_B^{=0} \mid \mathbf{X}_C^{=0}$, and in fact we have $\mathbf{X}_A^{=0} \perp\!\!\!\perp \mathbf{X}_B^{=0}$ since $\mathbf{X}_C^{=0}$ is deterministic. Then the independence $A \perp B\,[\Lambda^0_{A\cup B}]$ follows from Lemma IV.6 as we note that $\mathbf{X}_{A\cup B}^{=0}$ has Lévy measure $\Lambda^0_{A\cup B}$.

Part 2: Assume that $A \perp B \mid C\,[\Lambda]$. First we consider the case where Λ is a finite Lévy measure such that $\Lambda(\{x_C = 0_C\}) = 0$. Then $\mathbf{X}$ is the sum of a linear drift and a compound Poisson process. To be precise,

$$X(t) = t\gamma + \sum_{n=1}^{N(t)} Y(n), \qquad t \geq 0,$$

where $\mathbf{N}$ is a Poisson process with rate $\Lambda(\mathbb{R}^d)$ and $(Y(n))$ is a sequence of i.i.d. random variables, independent of $\mathbf{N}$ and with $Y(1) \sim \Lambda/\Lambda(\mathbb{R}^d)$. We note that conditioning on $\mathbf{X}_C$ will also fix $\mathbf{N}$ since $\Lambda(\{x_C = 0_C\}) = 0$. For each $n \in \mathbb{N}$ we further find that the random variables $Y_A(n)$ and $Y_B(n)$ are conditionally independent given $\mathbf{X}_C$. Now we apply the ideas from the proof of Proposition IV.3 to conclude that $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C$.

Now we no longer assume that Λ is finite. Instead we assume that $\Lambda(\{x_c = 0\}) = 0$ for some $c \in C$. We may view $\mathbf{X}$ as an almost sure limit of a sequence $(\mathbf{X}^{(n)})$ of Lévy processes, where $(\mathbf{X}^{(n)})$ has *finite* Lévy measure given by $\Lambda^{(n)} = \Lambda(\,\cdot\, \cap \{1/n < |x_c|\})$. Since $A \perp B \mid C\,[\Lambda]$ we also have $A \perp B \mid C\,[\Lambda^{(n)}]$ for every $n \in \mathbb{N}$. Then, by the previous paragraph, we get that $\mathbf{X}_A^{(n)} \perp\!\!\!\perp \mathbf{X}_B^{(n)} \mid \mathbf{X}_C$ since conditioning on $\mathbf{X}_C$ or $\mathbf{X}_C^{(n)}$ has the same effect on $\mathbf{X}^{(n)}$. Using Lemma IV.25 the conditional independence $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C$ follows.

Finally we can consider a general Lévy measure Λ. We proceed with induction in $k = |C|$. From Lemma IV.6 we know that the result holds for $k = 0$. Now, assume that it holds when C has $k-1$ elements, and fix some $c \in C$. We write $\mathbf{X} = \mathbf{X}' + \mathbf{X}''$, where $\mathbf{X}'$ and $\mathbf{X}''$ are independent Lévy processes with Lévy measures $\Lambda' = \Lambda(\,\cdot\, \cap \{x_c = 0\})$ and $\Lambda'' = \Lambda(\,\cdot\, \cap \{x_c \neq 0\})$. We have $A \perp B \mid C\,[\Lambda']$ and since Λ' is concentrated on $\{x_c = 0\}$ it follows that $A \perp B \mid C \setminus \{c\}\,[\Lambda'_{V\setminus\{c\}}]$. Importantly, $\Lambda'_{V\setminus\{c\}}$ satisfies [IV.A2], so by the induction hypothesis we have that $\mathbf{X}'_A \perp\!\!\!\perp \mathbf{X}'_B \mid \mathbf{X}'_{C\setminus\{c\}}$. We note that this still holds if we condition on all of $\mathbf{X}'_C$ instead because $\mathbf{X}'_c$ is deterministic. We further have $A \perp B \mid C\,[\Lambda'']$, and since $\Lambda''(\{x_c = 0\}) = 0$ we have $\mathbf{X}''_A \perp\!\!\!\perp \mathbf{X}''_B \mid \mathbf{X}''_C$ by the paragraph above. To conclude we make use of the fact that conditioning on $\mathbf{X}_C$ is the same as conditioning on the pair

$(\mathbf{X}'_C, \mathbf{X}''_C)$ as discussed in §IV.A.2. By combining this with the conditional independence statements for the two terms we arrive at the conditional independence $\mathbf{X}_A \perp\!\!\!\perp \mathbf{X}_B \mid \mathbf{X}_C$. For the last step we again use the ideas from the proof of Proposition IV.3. □

IV.A.7 Proof of Lemma IV.8

For any Borel set $E_{V\setminus\{c\}} \subseteq \mathbb{R}^{V\setminus\{c\}}$ we introduce an auxiliary set $\Delta^+_{E_{V\setminus\{c\}}} = \{(h, hx_{V\setminus\{c\}}) \mid h \geq 1, x_{V\setminus\{c\}} \in E_{V\setminus\{c\}}\}$ as illustrated in Figure IV.14 below. We can then define a probability measure μ_c^+ by

$$\mu_c^+(E_{V\setminus\{c\}}) = \Lambda(\Delta^+_{E_{V\setminus\{c\}}})/\Lambda_c([1,\infty)), \qquad E_{V\setminus\{c\}} \in \mathcal{B}(\mathbb{R}^{V\setminus\{c\}}).$$

Now, let $\xi^{(c,+)}$ be a d-dimensional random vector such that $\xi^{(c,+)}_{V\setminus\{c\}} \sim \mu_c^+$ and $\xi_c^{(c,+)} = 1$. For any $\epsilon > 0$ and $h > 0$ we have

$$\begin{aligned}\int_{[\epsilon,\infty)} \nu_{\{c\}}(h, hE_{V\setminus\{c\}})\,\Lambda_c(\mathrm{d}h) &= \Lambda(\epsilon\Delta^+_{E_{V\setminus\{c\}}}) \\ &= \epsilon^{-\alpha}\Lambda(\Delta^+_{R_{A\cup B}}) \\ &= \epsilon^{-\alpha}\Lambda_c([1,\infty))\mu_c^+(E_{V\setminus\{c\}}) \\ &= \Lambda_c([\epsilon,\infty))\mu_c^+(E_{V\setminus\{c\}}).\end{aligned}$$

Hence, $\nu_{\{c\}}(h, hE_{V\setminus\{c\}}) = \mu_c^+(E_{V\setminus c})$ for Λ_c-almost all $h \geq \epsilon$. Using standard arguments we may extend this to

$$\nu_{\{c\}}(h, hE_{V\setminus\{c\}}) = \mu_c^+(E_{V\setminus\{c\}}) \quad \text{for all } E_{V\setminus\{c\}} \in \mathcal{B}(\mathbb{R}^{V\setminus\{c\}})$$

for Λ_c-almost all $h > 0$. Hence,

$$\nu_{\{c\}}(h, E_{V\setminus\{c\}}) = \mu_c^+(h^{-1}E_{V\setminus\{c\}}) = \mathbb{P}(h\xi^{(c,+)} \in E_{V\setminus\{c\}}) \quad \text{for all } E_{V\setminus\{c\}} \in \mathcal{B}(\mathbb{R}^{V\setminus\{c\}})$$

for Λ_c-almost all $h > 0$.

The case of negative h is similar. Here we define $\mu_c^-(E_{V\setminus\{c\}}) = \Lambda(\Delta^-_{E_{V\setminus\{c\}}})/\Lambda_c((-\infty,-1])$, where $\Delta^-_{E_{V\setminus\{c\}}} = \{(h, |h|x_{V\setminus\{c\}}) \mid h \leq -1, x_{V\setminus\{c\}} \in E_{V\setminus\{c\}}\}$. □

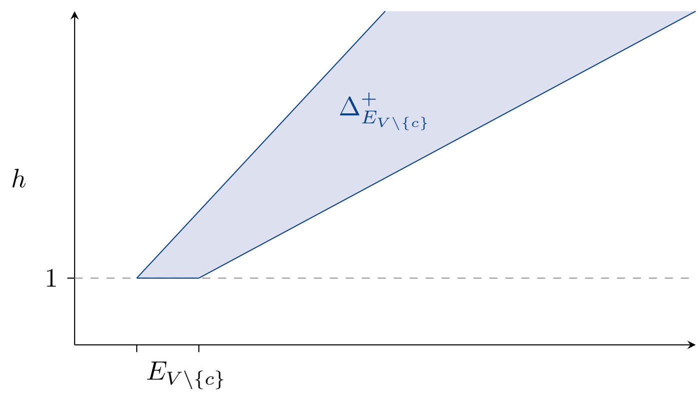


Figure IV.14: Illustration of the set $\Delta^+_{E_{V\setminus\{c\}}}$.

IV.A.8 Proof of Lemma IV.9

For any $v \in V \setminus \{c\}$ consider

$$\begin{aligned}\Lambda(\{x_v > 1, x_c > 0\}) &= \mathbb{E}[\int_0^\infty \mathbf{1}_{\{x_c \xi_v^{(c,+)} > 1\}} \alpha m_c^+ x_c^{-\alpha-1} \, \mathrm{d}x_c] \\ &= m_c^+ \mathbb{E}[|\xi_v^{(c,+)}|^\alpha \, \mathbf{1}_{\{\xi_v^{(c,+)} > 0\}}],\end{aligned}$$

which must be finite. We find that $\mathbb{E}[|\xi_v^{(c,+)}|^\alpha \, \mathbf{1}_{\{\xi_v^{(c,+)} < 0\}}] < \infty$ by considering $\Lambda(\{x_v < -1, x_c > 0\})$. Hence $\xi^{(c,+)}$ is in $\mathcal{L}^\alpha$ as claimed. Similar calculations show that $\xi^{(c,-)}$ is in $\mathcal{L}^\alpha$.

Now, consider a d-dimensional random vector $\xi^{(c,+)}$ with $\xi_c^{(c,+)} = 1$ a.s. We can define

$$\Lambda^+(E) = \mathbb{E}[\int_0^\infty \mathbf{1}_{\{x_c \xi^{(c,+)} \in E\}} \, x_c^{-\alpha-1} \, \mathrm{d}x_c], \qquad E \in \mathcal{B}(\mathbb{R}^d).$$

One easily checks that Λ^+ is a $-\alpha$-homogeneous measure with $\Lambda^+(\{0\}) = 0$. If we further assume that $\xi^{(c,+)}$ is in $\mathcal{L}^\alpha$ we find that

$$\int_0^\infty (1 \wedge \|x\|^2) \, \Lambda^+(\mathrm{d}x) < \infty.$$

Similarly, for a random vector $\xi^{(c,-)}$ with $\xi_c^{(c,-)} = -1$ a.s. we define

$$\Lambda^-(E) = \mathbb{E}[\int_{-\infty}^0 \mathbf{1}_{\{|x_c| \xi^{(c,-)} \in E\}} |x_c|^{-\alpha-1} \, \mathrm{d}x_c], \qquad E \in \mathcal{B}(\mathbb{R}^V),$$

and if $\xi^{(c,-)}$ is in $\mathcal{L}^\alpha$ this again defines a $-\alpha$-homogeneous Lévy measure on $\{x_c < 0\}$. Now it is easy to see that $\Lambda = \Lambda^+ + \Lambda^-$ is a $-\alpha$-homogeneous Lévy measure giving rise to the vectors $\xi^{c,\pm}$.

The final calculations are trivial. □

IV.A.9 Proof of Theorem IV.11

The conditional independence structures for Z and for Y under $\mathbb{P}_{c,\epsilon}$ are both given by the same tree T. Hence, by conditioning according to the tree structure one finds that it is sufficient to show the following:

(a) $Z_c \sim \mathbb{P}_{c,\epsilon}(Y_c \in \cdot\,)$.

(b) $(Z_i \mid Z_c = h) \sim \mathbb{P}_{c,\epsilon}(Y_i \in \cdot \mid Y_c = h)$ for $\mathbb{P}_{c,\epsilon}(Y_c \in \cdot\,)$-a.a. $|h| \geq \epsilon$.

(c) $(Z_j \mid Z_i = h) \sim \mathbb{P}_{c,\epsilon}(Y_j \in \cdot \mid Y_i = h)$ for $\mathbb{P}_{c,\epsilon}(Y_i \in \cdot\,)$-a.a. $h \in \mathbb{R}$ for any $(i,j) \in E^{(c)}$.

Firstly, it is clear that Z satisfies (a). Secondly, (b) follows directly from (IV.6). Now, Lemma IV.10 establishes (c) for $h \neq 0$ since $\{c\} \perp \{j\} \mid \{i\} \, [\Lambda]$ when $(i,j) \in E^{(c)}$. Finally, it might be that 0 is an atom for $\mathbb{P}_{c,\epsilon}(Y_i \in \cdot\,)$. However, in that case we simply note that

$$\mathbb{P}_{c,\epsilon}(Y_j = 0 \mid Y_i = 0) = 1$$

since $\Lambda(\{x_c \geq \epsilon, x_j \neq 0, x_i = 0\}) = 0$ according to Engelke et al. (2022, Cor. 6.3). □

IV.A.10 Proof of Proposition IV.12

We prove the first equality in the result and note that everything else follows immediately from there. Assume w.l.o.g. that $d = 2$, $i = 1$ and $j = 2$. For $v = 1, 2$ we let $\Delta^{(v)}$ be a random vector with distribution $\Lambda(\cdot \cap \{x_v \geq 1\})/\Lambda(\{x_v \geq 1\})$. Hence, $\Delta^{(1)} \overset{d}{=} P \cdot (1, \xi_{(1,2)})$ and $\Delta^{(2)} \overset{d}{=} P \cdot (\xi_{(2,1)}, 1)$, where P is independent of $\xi_{(1,2)}, \xi_{(2,1)}$ and has density $x \mapsto \alpha x^{-1-\alpha} \mathbf{1}_{\{x>1\}}$.

For $h \geq 1$ the set $\Theta_h = \{x_2 \geq 1, x_1 \geq h x_2\}$ is contained in $[1, \infty)^2$. We find that

$$\begin{aligned}
\mathbb{P}(\xi_{(2,1)} \geq h) &= \mathbb{P}(\Delta^{(2)} \in \Theta_h) \\
&= \frac{\Lambda(\Theta_h)}{\Lambda(\{x_2 \geq 1\})} \\
&= \frac{\Lambda(\Theta_h)}{\frac{m_2}{m_1}\Lambda(\{x_1 \geq 1\})} \\
&= \frac{m_1}{m_2}\mathbb{P}(\Delta^{(1)} \in \Theta_h).
\end{aligned}$$

To calculate the probability on the right-hand side we first condition on P. Note that we may rewrite $\Theta_h = \{x_1 \geq h, 1 \leq x_2 \leq x_1/h\}$. We get

$$\begin{aligned}
\mathbb{P}(\Delta^{(1)} \in \Theta_h) &= \int_h^\infty \alpha x_1^{-1-\alpha}\mathbb{P}(1 \leq x_1\xi_{(1,2)} \leq x_1/h)\,\mathrm{d}x_1 \\
&= \mathbb{E}\Big[\int_0^\infty \alpha x_1^{-1-\alpha}\,\mathbf{1}_{\{x_1 \geq h, 1 \leq x_1\xi_{(1,2)} \leq x_1/h\}}\,\mathrm{d}x_1\Big] \\
&= \mathbb{E}\Big[\int_0^\infty \alpha y^{-1-\alpha}\xi_{(1,2)}^\alpha\,\mathbf{1}_{\{y \geq h\xi_{(1,2)}, 1 \leq y, \xi_{(1,2)} \leq 1/h\}}\,\mathrm{d}y\Big] \\
&= \mathbb{E}[\mathbf{1}_{\{1/\xi_{(1,2)} \geq h\}}\,\xi_{(1,2)}^\alpha],
\end{aligned}$$

where the third equality comes from defining $y = x_1\xi_{(1,2)}$.

For $h \in (0, 1)$ we let $\Psi_h = \{x_1 \geq 1, 1/h \leq x_2 \leq x_1/h\}$. We deduce that

$$\begin{aligned}
h^\alpha\mathbb{P}(\xi_{(2,1)} \geq h) &= \mathbb{P}(P \geq 1/h, \xi_{(2,1)} \geq h) \\
&= \mathbb{P}(\Delta^{(2)} \in \Psi_h) \\
&= \frac{m_1}{m_2}\mathbb{P}(\Delta^{(1)} \in \Psi_h),
\end{aligned}$$

where the last equality follows as before using the fact that $\Psi_h \subseteq [1, \infty)^2$. With calculations similar to the above we can show that

$$\mathbb{P}(\Delta^{(1)} \in \Psi_h) = h^\alpha\mathbb{E}[\mathbf{1}_{\{1/\xi_{(1,2)} \geq h\}}\,\xi_{(1,2)}^\alpha].$$

□

IV.A.11 Proof of Proposition IV.20

The target distribution $\mathbb{P}_\epsilon$ has a density f w.r.t. Λ given by

$$f(x) = (\Lambda(\{\|x\|_\infty \geq \epsilon\}))^{-1}\,\mathbf{1}_{\{\|x\|_\infty \geq \epsilon\}}(x), \qquad x \in \mathbb{R}^d,$$

and the random vector Z has a density g w.r.t. Λ given by

$$g(x) = \frac{\epsilon^\alpha}{m}\#\{v \in V \mid |x_v| \geq \epsilon\}, \qquad x \in \mathbb{R}^d,$$

where $m = m_1 + \cdots + m_d$. To see this consider a set of coordinates $W \subseteq V$ and a Borel measurable subset $S \subseteq \{|x_v| \geq \epsilon \,\forall v \in W, |x_v| < \epsilon \,\forall v \in V \setminus W\}$. Then

$$\begin{aligned}\mathbb{P}(Z \in S) &= \sum_{v \in W} \mathbb{P}(Z \in S \mid I = v) \cdot \mathbb{P}(I = v) \\ &= \sum_{v \in W} \frac{\Lambda(S)}{\Lambda(\{|x_v| \geq \epsilon\})} \cdot \frac{m_v}{m} \\ &= \sum_{v \in W} \Lambda(S) \frac{\epsilon^\alpha}{m},\end{aligned}$$

where the last equality uses homogeneity of Λ. Note that the final sum coincides with $\int_S g(x)\, \Lambda(\mathrm{d}x)$.

We must bound the ratio f/g. For $x \in \mathbb{R}^d$ with $\|x\|_\infty \geq \epsilon$ we have

$$\begin{aligned}\frac{f(x)}{g(x)} &= \frac{(\Lambda(\{\|x\|_\infty \geq \epsilon\}))^{-1}}{\frac{\epsilon^\alpha}{m} \#\{v \in V \mid |x_v| \geq \epsilon\}} \\ &= \frac{m}{\Lambda(\{\|x\|_\infty \geq 1\}) \#\{v \in V \mid |x_v| \geq \epsilon\}} \\ &\leq \frac{m}{\Lambda(\{\|x\|_\infty \geq 1\})} \\ &=: M,\end{aligned}$$

where we once again use homogeneity of Λ. Finally, it is well-known that we obtain the target distribution by accepting the proposed Z with probability $f(Z)/(Mg(Z))$. This is exactly the quantity $(\#\{v \in V \mid |Z_v| \geq \epsilon\})^{-1}$ from step 3 of the procedure. □

IV.A.12 Proof of Lemma IV.21

Letting $\Lambda^{(\epsilon)}$ be the Lévy measure of the pre-limit process $\mathbf{W}^{(\epsilon)}/a_\epsilon$ we find

$$\begin{aligned}\int xx^\top \Lambda^{(\epsilon)}(\mathrm{d}x) &= a_\epsilon^{-2} \int_{\{|x_u| \leq \epsilon \,\forall u \in U\}} xx^\top \Lambda(\mathrm{d}x) \\ &= \int_{\{|x_u| \leq 1 \,\forall u \in U\}} xx^\top \Lambda(\epsilon\, \mathrm{d}x) \epsilon^\alpha \\ &= \Sigma,\end{aligned}$$

where the last equality follows from homogeneity of Λ.

Next we show that $\Lambda^{(\epsilon)} \xrightarrow{v} 0$ on $\overline{\mathbb{R}^d} \setminus \{0\}$. That is, for every $h > 0$ and $i \in V$ we have the convergence $\Lambda^{(\epsilon)}(\{|x_i| \geq h\}) \to 0$ as $\epsilon \downarrow 0$. Using homogeneity again we find that

$$\begin{aligned}\Lambda^{(\epsilon)}(\{|x_i| \geq h\}) &= \Lambda(\{|x_u| \leq \epsilon \,\forall u \in U, |x_i| \geq h a_\epsilon\}) \\ &= \epsilon^{-\alpha} \Lambda(\{|x_u| \leq 1 \,\forall u \in U, |x_i| \geq h\epsilon^{-\alpha/2}\}).\end{aligned}$$

On the latter set we have $\epsilon^{-\alpha} \leq h^{-2} x_i^2$ and so we find that

$$\Lambda^{(\epsilon)}(\{|x_i| \geq h\}) \leq h^{-2} \int_{|x_u| \leq 1 \,\forall u \in U, |x_i| \geq h\epsilon^{-\alpha/2}} x_i^2\, \Lambda(\mathrm{d}x) \to 0,$$

where we used dominated convergence with dominating function $x \mapsto x_i^2 \mathbf{1}_{\{|x_u| \leq 1 \,\forall u \in U\}}$.

According to Kallenberg (2021, Thm. 7.7 & Thm. 16.14) it is left to show

$$\int_{\{\|x\| \leq 1\}} xx^\top \Lambda^{(\epsilon)}(\mathrm{d}x) \to \Sigma \quad \text{and} \quad \gamma_\epsilon \to 0,$$

where γ_ϵ is obtained from the martingale requirement, i.e. $\gamma_\epsilon + \int_{\{\|x\|>1\}} x\,\Lambda^{(\epsilon)}(\mathrm{d}x) = 0$. Both are implied by

$$\int_{\{\|x\|>1\}} x_i^2\,\Lambda^{(\epsilon)}(\mathrm{d}x) = \int_{\{|x_u|\le 1\,\forall u\in U,\|x\|\ge\epsilon^{-\alpha/2}\}} x_i^2\,\Lambda(\mathrm{d}x) \to 0,$$

derived using the same calculations as above. For the second we also employ the vague convergence to the zero measure.

IV.A.13 Proof of Theorem IV.23

We start by proving the following auxiliary result.

Lemma IV.26. *Consider a positive semi-definite block matrix of the form*

$$M = \begin{pmatrix} \delta & \delta m_1^\top & m_2^\top \\ \delta m_1 & \delta M_{11} & m_1 m_2^\top \\ m_2 & m_2 m_1^\top & M_{22} \end{pmatrix},$$

where δ is a scalar, M_{11}, M_{22} are square matrices, and m_1, m_2 are vectors of appropriate dimensions. If M is positive-definite (i.e. invertible) then the $(2,3)$ and $(3,2)$ blocks of M^{-1} are zero.

Proof. For a positive definite block matrix $X = \begin{pmatrix} P & Q \\ R & S\end{pmatrix}$ with square blocks P, S it is well known that

$$X^{-1} = \begin{pmatrix} P^{-1} + P^{-1}QYRP^{-1} & -P^{-1}QY \\ -YRP^{-1} & Y \end{pmatrix}, \qquad Y = (S - RP^{-1}Q)^{-1}.$$

Firstly, we have

$$\begin{pmatrix} 1 & m_1^\top \\ m_1 & M_{11}\end{pmatrix}^{-1} = \begin{pmatrix} \cdot & \cdot \\ -(M_{11} - m_1 m_1^\top)^{-1} m_1 & (M_{11} - m_1 m_1^\top)^{-1}\end{pmatrix},$$

where the first row is not important to us. Secondly, we partition M so that $S = M_{22}$ and obtain the following representation of the $(2,3)$ block of M^{-1}:

$$-\delta^{-1}[-(M_{11} - m_1 m_1^\top)^{-1} m_1, (M_{11} - m_1 m_1^\top)^{-1}] \begin{pmatrix} m_2^\top \\ m_1 m_2^\top \end{pmatrix} Y = 0.$$

□

Proof of Theorem IV.23. Without loss of generality we may assume that $C = \{1\}$, the elements of A are smaller than the elements of B, and $U \subseteq B \cup \{1\}$. According to Lemma IV.26 it is sufficient to establish that Σ has the form of M.

The marginal measure is given by

$$\Lambda_1(\mathrm{d}x_1) = m_1 \alpha x_1^{-\alpha-1}\,\mathrm{d}x_1, \qquad x_1 > 0,$$

where $m_1 = \Lambda(\{x_1 > 1\})$. For $x_1 > 0$ consider the kernel $\nu_{\{1\}}(x_1, \cdot)$. Recall that Lemma IV.8 gives the existence of a d-dimensional random vector $\xi^{(1)}$ such that $\xi^{(1)}_1 = 1$ a.s. and

$$\nu_{\{1\}}(x_1, \cdot\,) = \mathbb{P}(x_1 \xi^{(1)}_{V\setminus\{1\}} \in \cdot\,)$$

for Λ_1-almost all $x_1 > 0$.

For $i \in A \cup \{1\}$ and $j \in B \cup \{1\}$ we may compute

$$\begin{aligned}\Sigma_{ij} &= \int_{\{|x_u| \le 1\,\forall u \in U\}} x_i x_j \,\Lambda(\mathrm{d}x) \\ &= m_1 \alpha \mathbb{E}\Big[\int_{\{x_1 > 0, |x_1 \xi_u^{(1)}| \le 1\,\forall u \in U\}} (x_1 \xi_i^{(1)})(x_1 \xi_j^{(1)}) x_1^{-\alpha-1}\,\mathrm{d}x_1\Big] \\ &= \frac{m_1 \alpha}{2-\alpha} \mathbb{E}\big[\xi_i^{(1)} \xi_j^{(1)} \min_{u \in U} |\xi_u^{(1)}|^{\alpha-2}\big].\end{aligned}$$

For the second equality we used that $x_1 = 0$ has no contribution to Σ. If $i = 1$ or $j = 1$ this is obvious, and if $i \in A, j \in B$ we recall that $A \perp B \mid \{1\}\,[\Lambda]$ implies $\Lambda^0_{V\setminus\{1\}}(\{x_i \neq 0, x_j \neq 0\}) = 0$, see Engelke et al. (2022, Lem. 3.3 & Prop. 5.1).

According to Engelke et al. (2022, Thm. 4.4) the conditional independence $A \perp B \mid \{1\}\,[\Lambda]$ implies $\xi_A^{(1)} \perp\!\!\!\perp \xi_B^{(1)}$. Letting $\delta = \frac{m_1\alpha}{2-\alpha}\mathbb{E}\left[\min_{u\in U}|\xi_u^{(1)}|^{\alpha-2}\right]$ we find that

$$\begin{aligned}\Sigma_{11} &= \delta, \\ \Sigma_{1i} &= \delta \mathbb{E}[\xi_i^{(1)}], \\ \Sigma_{1j} &= \frac{m_1\alpha}{2-\alpha}\mathbb{E}[\xi_j \min_{u\in U}|\xi_u^{(1)}|^{\alpha-2}], \\ \Sigma_{ij} &= \frac{m_1\alpha}{2-\alpha}\mathbb{E}[\xi_i^{(1)}]\mathbb{E}[\xi_j \min_{u\in U}|\xi_u^{(1)}|^{\alpha-2}].\end{aligned}$$

This shows that Σ has the form of M in Lemma IV.26.

□

References

D. Applebaum. *Lévy processes and stochastic calculus*, volume 116 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2009. ISBN 978-0-521-73865-1.

J. Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996. ISBN 0-521-56243-0.

S. Cohen and J. Rosiński. Gaussian approximation of multivariate Lévy processes with applications to simulation of tempered stable processes. *Bernoulli*, 13(1):195–210, 2007. ISSN 1350-7265.

V. Didelez. Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 70(1):245–264, 2008. ISSN 13697412, 14679868.

I. Eder and C. Klüppelberg. Pareto Lévy measures and multivariate regular variation. *Adv. in Appl. Probab.*, 44(1):117–138, 2012. ISSN 0001-8678.

M. Eichler. Graphical modelling of multivariate time series. *Probability Theory and Related Fields*, 153(1):233–268, 2012.

S. Engelke and A. S. Hitz. Graphical models for extremes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 82(4):871–932, 2020. ISSN 1369-7412. With discussions.

S. Engelke and S. Volgushev. Structure learning for extremal tree models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 84(5):2055–2087, 2022. ISSN 1369-7412,1467-9868.

S. Engelke, J. Ivanovs, and K. Strokorb. Graphical models for infinite measures with applications to extremes and Lévy processes. *arXiv preprint arXiv:2211.15769*, 2022.

C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969. ISSN 00129682, 14680262.

M. Hentschel, S. Engelke, and J. Segers. Statistical inference for Hüsler-Reiss graphical models through matrix completions. *arXiv preprint arXiv:2210.14292*, 2022.

S. Hu, Z. Peng, and J. Segers. Modelling multivariate extreme value distributions via markov trees. *arXiv preprint arXiv:2208.02627*, 2022.

J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003. ISBN 3-540-43932-3.

O. Kallenberg. *Foundations of modern probability*, volume 99 of *Probability Theory and Stochastic Modelling*. Springer, Cham, third edition, 2021. ISBN 978-3-030-61871-1; 978-3-030-61870-4.

J. Kallsen and P. Tankov. Characterization of dependence of multidimensional Lévy processes using Lévy copulas. *J. Multivariate Anal.*, 97(7):1551–1572, 2006. ISSN 0047-259X.

S. L. Lauritzen. *Graphical models*, volume 17 of *Oxford Statistical Science Series*. The Clarendon Press, Oxford University Press, New York, 1996. ISBN 0-19-852219-3. Oxford Science Publications.

N. Misra and E. E. Kuruoglu. Stable graphical models. *Journal of Machine Learning Research*, 17(168):1–36, 2016.

S. W. Mogensen and N. R. Hansen. Graphical modeling of stochastic processes driven by correlated noise. *Bernoulli*, 28(4):3023 – 3050, 2022.

K.-i. Sato. *Lévy processes and infinitely divisible distributions*, volume 68 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1999. ISBN 0-521-55302-4. Translated from the 1990 Japanese original, Revised by the author.

V. M. Zolotarev. *One-dimensional stable distributions*, volume 65 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1986. ISBN 0-8218-4519-5. Translated from the Russian by H. H. McFaden, Translation edited by Ben Silver.