Convex minorants and the simulation of the extrema of Lévy processes

by

Jorge Ignacio González Cázares

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy in Statistics

Department of Statistics

December 2021

Contents

Acknowledgements

I want to thank my supervisor, Aleksandar Mijatović, for providing me with great guidance throughout my PhD and for his immense support from beginning to end. I thank David Bang, Vladimir Fomichov, Jevgenijs Ivanovs, Arturo Kohatsu–Higa and Gerónimo Uribe Bravo for the wonderful experience I had during our collaborations and for everything they taught me. The support provided by the Alan Turing Institute and Conacyt proved crucial in the development of this thesis and the multiple papers that I wrote during my PhD. I want to thank my family for their emotional support, especially during the difficult years of 2020 and 2021.

Declaration

Part of the work included in this thesis was developed in collaboration with Aleksandar Mijatović in [52, 53] and with Aleksandar Mijatović and Gerónimo Uribe Bravo in [56].

Abstract

In this thesis we will establish the stick-breaking representation of the convex minorant and the extrema of an arbitrary Lévy process. Our self-contained elementary proof is based on the analysis of piecewise linear convex functions and requires only very basic properties of Lévy processes. We then use the stick-breaking representation to create geometrically convergent simulation algorithm for the extrema of a Lévy process whose increments can be sampled. For processes whose increments cannot be sampled we develop a multilevel Monte Carlo algorithm using the stick-breaking representation. In all cases, the algorithms present in this thesis outperform the existing algorithms in the literature.

Chapter 1

Introduction

Consider a stochastic process S driven by a Lévy process X (e.g. with $S = S_0 + X$ or $S = S_0 \mathrm{e}^X$). Given a time horizon $T > 0$ consider the vector $\overline{\chi}_T(S) = (S_T, \overline{S}_T, \overline{\tau}_T(S))$ consisting of the position S_T, the supremum $\overline{S}_T$ of S over the interval $[0, T]$ and the first time $\overline{\tau}_T(S)$ the process S attains the supremum $\overline{S}_T$. This vector is of interest in applications, for instance, it appears in the buffer size in queuing theory [4, 79], insurance mathematics [65], mathematical finance [20, 80, 82] and in optimal stopping [11, 12]. Except for specific cases, the law of $\overline{S}_T$ (let alone the vector $\overline{\chi}_T(S)$) is typically intractable and even its simulation is hard to achieve. In this thesis we study this vector and, for simplicity, restrict our applications to the domain of mathematical finance.

We start in Chapter 2 with a study on the convex minorant C_T^X of X, the largest convex function dominated by X on $[0, T]$. We prove that the convex minorant C_T^X is piecewise linear and establish in Theorem 2.1 a general stochastic representation in terms of a stick-breaking process ℓ on $[0, T]$ (see Figure 1.1 below) and the increments of an independent Lévy process Y with the same law as X. The representation shows that the C_T^X has the same law as the unique piecewise linear convex function with infinitely many linear segments whose n-th linear segment (enumerated in some order) has length ℓ_n and height $\xi_n = Y_{L_{n-1}} - Y_{L_n}$, where $L_n = T - \sum_{k=1}^{n-1} \ell_k$ (see Figure 1.2 below).

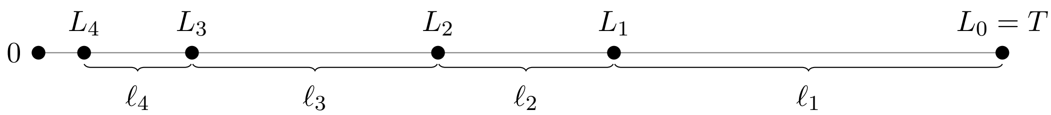

Figure 1.1: The figure illustrates the first $n = 4$ sticks of a stick-breaking process. The increments of Y in (2.3) are taken over the intervals $[L_k, L_{k-1}]$ of length ℓ_k.

Observe that these linear segments indeed completely determine the piecewise linear convex function since, in the graph of this function, the segments appear chronologically by increasing slope. Strikingly, the proof of Theorem 2.1 requires only elementary geometry and analysis, used to study the convergence of piecewise linear functions, and the fact that Lévy processes have stationary increments and right-continuous paths with left limits. In particular, this proof does not require the introduction of fluctuation theory, local times, excursion theory or even the Lévy-Khintchine formula.

Moreover, since the minimum of X and its temporal location can be obtained by adding the heights and lengths of the linear segments of C_T^X with negative height, we obtain a simple stochastic representation for the vector containing the state X_T, the infimum $\underline{X}_T$ and the first time the infimum is attained $\underline{\tau}_T(X)$ and, by a time reversal argument, for the vector $\overline{\chi}_T(X)$. In fact, Theorem 2.1 can be used to obtain novel and simple proofs of some of the classical and most prominent results in fluctuation theory including Rogozin's criterion and the Wiener–Hopf factorisation (see Section §2.6 below).

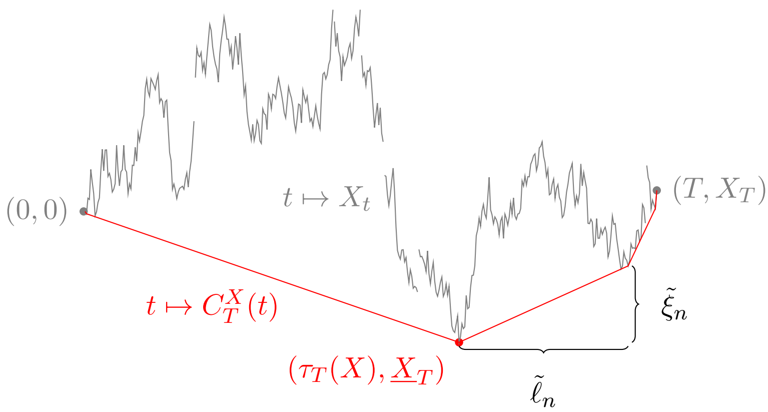

Figure 1.2: A sample path of a Lévy process X on the interval $[0,T]$, the graphs of the convex minorant C_T^X and the time and space position of their infimum $(\underline{\tau}_T(X), \underline{X}_T)$. The length $\tilde{\ell}_n$ and height $\tilde{\xi}_n$ of the chosen segment has the same joint law as ℓ_n and ξ_n.

In Chapter 3 we use the stick-breaking representation of the $\overline{\chi}_T(X)$ to develop the stick-breaking approximation (SBA) of this vector and SB-Alg, the corresponding geometrically convergent simulation algorithm. The algorithm is valid for any Lévy process whose marginals can be sampled exactly, comprising most of the Lévy processes used in practical models. The geometrically fast convergence of SB-Alg is a consequence of the fact that $\ell_n \to 0$ geometrically fast with $\mathbb{E}\ell_n = 2^{-n}T$ and makes our algorithm outperform all other existing algorithms, which only converge polynomially fast in the computational effort. We establish the geometrically fast convergence of the SBA under several metrics. Finally, the corresponding Monte Carlo and multilevel Monte Carlo algorithms are developed and analysed.

In Chapter 4 we consider Lévy processes whose marginals cannot be sampled, requiring a Gaussian approximation for its small-jump component. Such processes include empirically fit Lévy processes and other widely used models such as CGMY processes. We use the stick-breaking representation of the vector $\overline{\chi}_T(X)$ to analyse the distance between the vector $\overline{\chi}_T(X)$ and that of its Gaussian approximation and obtain sharp rates of convergence under several metrics. The stick-breaking representation is also used to develop SBG-Alg, a fast simulation algorithm for the vector of the Gaussian approximation, which yields a novel multilevel Monte Carlo algorithm for $\overline{\chi}_T(X)$. We analyse the complexity of this multilevel Monte Carlo algorithm and show that it is typically orders of magnitude faster than existing alternatives.

We stress here that Theorem 2.1 has more applications beyond those included within the scope of this thesis. For instance, the author of this thesis has used Theorem 2.1 to develop ε-strong simulation algorithms of stable meanders [55], an exact simulation algorithm of the extrema of stable processes [54] and Monte Carlo algorithms for the extrema of tempered stable processes [57]. The author has also used Theorem 2.1 to obtain new theoretical results, deriving results on the regularity of the law of the supremum of a stable process [27] and describing the asymptotic shape of the convex minorant of a Lévy process [9].

§1.1 Discussion on algorithms SB-Alg and SBG-Alg

The key role of $\overline{\chi}_T(X)$ in applied probability, combined with its intractability when X is *not* a jump diffusion (i.e., the linear combination of a compound Poisson process and a Brownian motion with drift), has lead to numerous works on its approximation over the last quarter of a century [7, 17, 21, 22, 30, 38, 39, 42, 43, 45, 49, 60, 70]. These approximations naturally yield Monte Carlo (MC) and Multilevel Monte Carlo (MLMC) algorithms for $\overline{\chi}_T(X)$. Without exception, the errors of these algorithms achieve *polynomial* decay in the computational cost. These algorithms, like the ones we will present in Chapters 3 and 4, are constructed under different sets of assumptions, jointly covering almost all Lévy models used in applications.

The approximations of the vector $\overline{\chi}_T(X)$ can be split into two classes, according to two assumptions: approximations reliant on the ability to simulate (exactly) the increments of the Lévy process X and approximations reliant on the ability to simulate the jumps of X or, equivalently, the ability to simulate from the restrictions of the Lévy measure of X to the subset $\mathbb{R} \setminus (-\varepsilon, \varepsilon)$ for any $\varepsilon > 0$. The former assumption is generally stronger and it typically provides faster convergence rates in the computational effort than the second assumption.

Approximations requiring the simulation of increments

Let us start with a brief discussion on the literature of approximations of $\overline{\chi}_T(X)$ under the baseline assumption that the increments (equivalently, its transitions) of X over intervals of length $t \in [0, T]$ can be sampled exactly with an expected cost that is bounded for $t \in [0, T]$. The random walk approximation (RWA) approximates $\overline{\chi}_T(X)$ with the corresponding vector of the skeleton $(X_{kT/n})_{k\in\{0,\dots,n\}}$ of the Lévy process X.

The RWA is a widely used method for approximating $\overline{\chi}_T(X)$ with computational cost proportional to the discretisation parameter n. In the case of Brownian motion, the asymptotic law of the error was studied in [7]. The papers [21, 22] (resp. [42, 43]) identified the dominant error term of the RWA for barrier and lookback options under the exponential Lévy models when X is a Brownian motion with drift (resp. jump diffusion). Based on Spitzer's identity, [30] developed bounds on the decay of the error in L^1 for general Lévy processes, extending the results of [42]. Ideas from [60] were employed in [17] to obtain sharper bounds on the convergence of the error of the RWA in L^p for general Lévy processes and any $p > 0$. Such results are useful for the analysis of MC and MLMC schemes based on the RWA, see [49] for the case of certain parametric Lévy models. We discuss these contributions in more detail in §3.1 as they are contrasted with the analogous results for the SBA.

Exploiting the the Wiener-Hopf factorisation, [70] introduced the Wiener-Hopf approximation (WHA) of $(X_T, \overline{X}_T)$. The WHA is given by $(X_{G_n}, \overline{X}_{G_n})$, where G_n is the sum of n independent exponential random variables with mean T/n, so that $\mathbb{E}G_n = T$ with variance T^2/n. Implementing the WHA requires the ability to sample the increment and supremum of X at an independent exponential time, which is only done approximately for a specific parametric class of Lévy processes with exponential moments and arbitrary path variation [70]. The computational cost of the WHA is proportional to n. The decay of the bias and the MLMC version of the WHA were later studied in [45]. As observed in [49, §1], the WHA currently cannot be directly applied to various parametric models used in practice possessing increments that can be simulated exactly (e.g. the variance gamma process).

Approximations requiring the simulation of jumps

The Gaussian approximation (GA) introduced in [6] proposed approximating the small-jump martingale containing the jumps of X whose size is smaller than $\kappa \in (0,1]$ with an independent Brownian motion with the same variance. The resulting approximating process $X^{(\kappa)}$ is a diffusion with finitely many jumps, so the vector $\overline{\chi}^{(\kappa)} = (X^{(\kappa)}, \overline{X}^{(\kappa)}, \tau_T(X^{(\kappa)}))$ may be sampled exactly. The paper [52], by the author of the current thesis, was the first to present a complete complexity analysis for this vector. In particular, sharp bounds on the bias and a multilevel simulation algorithm were first introduced in [52]. The approach used there combines the GA and the SBA to produce the SBGA (stick-breaking Gaussian approximation). A more detailed description of the properties of this approximation will be provided in Chapter 4.

The jump-adapted Gaussian approximation (JAGA), introduced in [38, 39] to approximate Lipschitz functions in the supremum norm of Lévy-driven stochastic differential equations with Lipschitz coefficients, can be used to estimate the extrema of Lévy processes. The algorithm is based on an approximation of the skeleton $\{X_{t_k}\}_{k=1}^n$ where the time grid includes the times of the jumps of X whose magnitude is larger than some cutoff level κ and the small-jump component of X is approximated by an additional Brownian motion. Typically, the cost and bias of the JAGA are proportional to $n + \kappa^{-\beta}$ and $(n^{-1/2} + n^{1/4}\kappa)\sqrt{\log n}$, respectively, where β is the Blumenthal-Getoor index, see (3.14) for its definition. The complexity of the MLMC version of the JAGA for Lipschitz functions of $(X_T, \overline{X}_T)$ is compared with that of the SBA in §3.2.4 of Chapter 3.

In contrast with Theorem 3.1 for the SBA, the laws of the errors of all the other algorithms discussed in the present subsection are intractable. The error of the SBA χ_n in (3.1) decays geometrically in L^p (see Theorem 3.3 below) as opposed to the polynomial decay for the other algorithms (see §3.2.1 below). The error in L^p of the SBA applied to locally Lipschitz and barrier-type functions arising in applications also decays geometrically (see Propositions 3.6 & 3.7 below). Such errors have not been analysed for algorithms other than the RWA, which has polynomial decay (see §3.2.2 for details) and the SBG (see details in Chapter 4). The rate of the decay of the bias is directly linked to the computational complexity of MC and MLMC estimates. Indeed, if the mean squared error is to be at most $\epsilon > 0$, the MC algorithm based on the SBA has (near optimal) complexity of order $\mathcal{O}(\epsilon^{-2}\log\epsilon)$. The MLMC scheme based on Algorithm 1. SB-Alg has (optimal) complexity of order $\mathcal{O}(\epsilon^{-2})$, which is in general neither the case for the RWA [49] nor the WHA [45] (see details in §3.2.4).

Chapter 2

Lévy processes: extrema and convex minorants

Notation

In this thesis, standard knowledge and notation of probability theory (particularly the first 20 chapters of the monograph [62]) is assumed. Some notation may vary between chapters to avoid the use of overloaded notation. However, all such notation is introduced and kept consistent within each chapter. Moreover, the following notation is also fixed throughout: for any $x, y \in \mathbb{R}$, we denote

$$x \wedge y = \min\{x, y\}, \quad x \vee y = \max\{x, y\}, \quad x^+ = \max\{x, 0\}, \quad \text{and} \quad x^- = \max\{-x, 0\}.$$

Given two functions f, g, we write $f(x) = \mathcal{O}(g(x))$ (resp. $f(x) = \mathrm{o}(g(x))$; $f(x) = \Theta(g(x))$; $f(x) \sim g(x)$) as $x \to a$, if $\limsup_{x \to a} f(x)/g(x) < \infty$ (resp. $f(x)/g(x) \to 0$; $f(x) = \mathcal{O}(g(x))$ and $g(x) = \mathcal{O}(f(x))$; $f(x)/g(x) \to 1$), where a is usually taken in the set $\{0, \infty\}$.

For any càdlàg function $f : [0, \infty) \to \mathbb{R}$ (i.e. right-continuous with left-hand limits), we denote by $\overline{f}_t = \sup\{f_s : s \in [0, t]\}$ its supremum over the interval $[0, t]$ and by $\tau_t(f) = \inf\{s \in [0, t] : \overline{f}_s = \overline{f}_t\}$ the first time the supremum $\overline{f}_t$ is attained.

§2.1 Lévy processes

A Lévy process $X = (X_t)_{t \geq 0}$ is a stochastic process with independent and stationary increments exhibiting càdlàg paths, i.e., paths that are right-continuous with left-hand limits. Simple examples of Lévy processes include Brownian motion, Poisson process and compound Poisson process. Given independent Lévy process, any linear combination is still a Lévy process. In particular, the linear combination of a Brownian motion and a compound Poisson process is termed *jump diffusion*. We will assume throughout a basic understanding of Lévy processes. The interested reader is referred to the monograph [91] for a thorough treatment of these processes.

Under weak assumptions it is possible to simulate the entire path of a compound Poisson process since it has only finitely many jumps. Moreover, since the law of the extrema of a Brownian motion is very well understood, it is clear that many properties of the extrema of jump diffusions can also be attained. In particular, its exact simulation is possible. This, however, is generally not the case for Lévy processes.

The main difficulty in studying Lévy processes arises when considering Lévy processes with an infinite amount of jumps on any nonempty finite interval. This property is only dependent on the Lévy measure ν, which specifies the expected number of jumps on any given measurable set. Moreovoer, for such processes, even if it is possible to simulate its jumps, this is not enough to sample X_T exactly, let alone its extrema. This thesis presents multiple approximations of the vector $\overline{\chi}_T = (X_T, \overline{X}_T, \tau_T(X))$ under different assumptions and analyse their computational complexities and convergence speeds. These are typically described in terms of the Blumenthal–Getoor index.

§2.2 Stick-breaking representation of convex minorants

The main goal of this chapter is to establish Theorem 2.1, which will trivially imply (2.2). Theorem 2.1 gives a stick-breaking representation of the convex minorant of a Lévy process, first established in [53, Thm 12] and, for Lévy processes with diffuse marginals, in [87, Thm 1]. The representation provided by Theorem 2.1 can be used to obtain short proofs of the Rogozin's criterion for the regularity of X at its starting point, Spitzer's formula for the supremum of X, the Wiener-Hopf identities and the continuity of the law of the triplet. All these results are easy corollaries of Theorem 2.1 and basic properties of X. In particular, our approach circumvents local times and excursion theory used in other probabilistic proofs of fluctuation identities [13, 73] and the continuity of the law of the triplet [28] (see [53] for a detailed account).

We will provide an simple proof of Theorem 2.1, relying entirely on elementary geometrical arguments (given in §2.3) to control the convergence of the piecewise linear convex functions of the approximating random walks to the convex function whose law is equal to that of the convex minorant of an arbitrary Lévy process.

Considering the convex minorant as a whole, rather than face-by-face, was crucial both in finding the correct formulation of Theorem 2.1, which generalises [87, Thm 1], and a proof that circumvented local times, excursion theory and the Skorokhod topology.

Given any càdlàg function $x : [0, T] \to \mathbb{R}$, its convex minorant, denoted by C_T^x, is the largest convex function that is pointwise smaller than x. The goal of this section is to prove our main result, Theorem 2.1, which (when applied to $-X$) clearly yields (2.2) (see details in the paragraph following the statement of Theorem 2.1).

Theorem 2.1. *Let X be a Lévy process and fix $T > 0$. Let $(\ell_n)_{n\in\mathbb{N}}$ be a uniform stick-breaking process on $[0, T]$ independent of $Y \overset{d}{=} X$. Then the convex minorant C_T^X of X has the same law (in the space of continuous functions on $[0, T]$) as the piecewise linear convex function on $[0, T]$ given by the formula*

$$\begin{aligned} t \mapsto \sum_{n=1}^{\infty} \xi_n \min\{(t - a_n)^+/\ell_n, 1\}, \quad \textit{where} \quad \xi_n = Y_{L_{n-1}} - Y_{L_n} \textit{ and} \\ a_n = \sum_{k=1}^{\infty} \ell_k \mathbb{1}_{\{\xi_k/\ell_k<\xi_n/\ell_n\}} + \sum_{k=1}^{n-1} \ell_k \mathbb{1}_{\{\xi_k/\ell_k=\xi_n/\ell_n\}}, \quad n \in \mathbb{N}. \end{aligned} \tag{2.1}$$

In particular, the face of the piecewise linear function with horizontal length ℓ_n has vertical height ξ_n.

The maximality of C_T^{-X} implies that C_T^{-X} and $-X$ have the same initial, final and minimal values and that the times where they first reach their minima on $[0, T]$ agree. Moreover, the minimum (resp. the first time of the minimum is attained) of C_T^{-X} can be calculated by adding all the heights (resp. lengths) of the faces of C_T^{-X} with negative height. Since the location and value of the infimum of $-X$ can be used to recover the location and value of the supremum of X, Theorem 2.1 (applied to $-X$) yields the following distributional equality:

$$\overline{\chi}_T = (X_T, \overline{X}_T, \tau_T(X)) \overset{d}{=} \sum_{n=1}^{\infty} \big(\xi_n, \xi_n^+, \ell_n \mathbb{1}_{\{\xi_n>0\}}\big). \tag{2.2}$$

By possibly extending the probability space, we may use [62, Thm 6.10] and the fact that the Skorokhod space $\mathcal{D}[0, T]$ of right-continuous functions on $[0, T]$ with left-hand limits (see [14, p. 109]) is Polish under the J_1-topology [14, p. 112] and thus a Borel space [62, Thm A1.2], to assume (under a coupling of (X, Y, ℓ))

that the equality in (2.2) holds with probability 1:

$$\overline{\chi}_T = (X_T, \overline{X}_T, \tau_T(X)) = \sum_{n=1}^{\infty} \left(\xi_n, \xi_n^+, \ell_n \mathbb{1}_{\{\xi_n>0\}}\right) \text{ a.s.} \tag{2.3}$$

This coupling will be assumed throughout Chapters 3–4, as it is crucial in providing both, the required tool to study $\overline{\chi}_T(X)$ and a probability space where we may study strong errors. Note that, in particular, the coupling satisfies $Y_T = X_T$ a.s.

Overview of the proof of Theorem 2.1. Theorem 2.1 connects two worlds: (**W1**) X and its convex minorant C_T^X on the interval $[0, T]$ and (**W2**) a random piecewise linear convex function. We first establish a convergence result within (**W2**) for a sequence of piecewise linear convex functions, see §2.3. This crucial step in the proof requires only elementary geometric manipulations of piecewise linear convex functions. In §2.4, using [1, Thm 1], we establish a bridge between (**W1**) and (**W2**) for random walks. We recall the 3214 path transformation [1] for random walks and provide a short proof, based on the convergence results in §2.3, of the connection between (**W1**) and (**W2**) for random walks with general increments, see Theorem 2.6 below. In §2.5, we establish Theorem 2.1 by taking the limit of the convex minorant of the random-walk skeleton of X in (**W1**) and, using the convergence results of §2.3, the corresponding limit in (**W2**).

We stress that the proof of Theorem 2.1 given in this chapter is self-contained, requiring only rudimentary real analysis and the fact that X has stationary and independent increments and right-continuous paths with left limits. In particular, we make no use of the Lévy measure, the Lévy-Khintchine formula for X or weak convergence in the J_1-topology on the Skorokhod space.

§2.3 Convex minorants and piecewise linear functions

We denote $[\![n]\!] = \{1, \ldots, n\}$ for $n \in \mathbb{N}$ and adopt the convention $[\![\infty]\!] = \mathbb{N}$. We say that a function $f : [0, T] \to \mathbb{R}$ is piecewise linear if there exists a set consisting of $N \in \overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ pairwise disjoint non-degenerate subintervals $\{(a_n, b_n) : n \in [\![N]\!]\}$ of $[0, T]$ such that $\sum_{n=1}^{N}(b_n - a_n) = T$ and f is linear on each (a_n, b_n). The face of f corresponding to the subinterval (a_n, b_n), has length $l_n = b_n - a_n > 0$, height $h_n = f(b_n) - f(a_n) \in \mathbb{R}$ and slope h_n/l_n. Note that, if f is continuous and of finite variation $\sum_{n=1}^{N} |f(b_n) - f(a_n)| < \infty$, the following representation holds:

$$f(t) = f(0) + \sum_{n=1}^{N} h_n \min\{(t - a_n)^+/l_n, 1\}, \qquad t \in [0, T]. \tag{2.4}$$

The number N in representation (2.4) is not unique in general as any face may be subdivided into two faces with the same slope. Moreover, for a fixed f and N, the set of intervals $\{(a_n, b_n) : n \in [\![N]\!]\}$ need not be unique. Furthermore we stress that the sequence of faces in (2.4) does not necessarily respect the chronological order. Put differently, the sequence $(a_n)_{n\in[\![N]\!]}$ need not be increasing. Throughout, we use the convention $\sum_{k=n}^{m} = 0$ when $n > m$.

Lemma 2.2. *Fix $T > 0$, $N \in \overline{\mathbb{N}}$ and let $l = (l_n)_{n=1}^N$ be a sequence of positive lengths with $\sum_{n=1}^N l_n = T$.*
(a) *For any sequence of heights $h = (h_n)_{n=1}^N$ with $\sum_{n=1}^N |h_n| < \infty$, the function*

$$\begin{aligned} F_{l,h}(t) &= \sum_{n=1}^{N} h_n \min\{(t-a_n)^+/l_n, 1\}, \qquad t \in [0,T], \quad \textit{where} \\ a_n &= \sum_{k=1}^{N} l_k \cdot \mathbb{1}_{\{h_k/l_k < h_n/l_n\}} + \sum_{k=1}^{n-1} l_k \cdot \mathbb{1}_{\{h_k/l_k = h_n/l_n\}}, \qquad n \in [\![N]\!], \end{aligned} \tag{2.5}$$

is piecewise linear and convex with $F_{l,h}(0) = 0$. Differently put, $F_{l,h}$ is linear on each interval $(a_n, a_n + l_n)$ with length l_n and height h_n. Moreover, any piecewise linear convex function started at zero whose faces have lengths l and heights h must equal $F_{l,h}$.
(b) *Suppose $N < \infty$. Given two sequences of heights $h = (h_n)_{n=1}^N$ and $h' = (h'_n)_{n=1}^N$, denote the corresponding functions in (2.5) by $F_{l,h}$ and $F_{l,h'}$ with sequences $(a_n)_{n=1}^N$ and $(a'_n)_{n=1}^N$ of the left endpoints of the intervals on which these functions are linear, respectively. Define the function*

$$G_{l,h,h'}(t) = \sum_{n=1}^{N} h_n \min\{(t-a'_n)^+/l_n, 1\}, \qquad t \in [0,T].$$

Then, we have

$$\begin{aligned} &\max\{\|F_{l,h} - F_{l,h'}\|_\infty, \|F_{l,h'} - G_{l,h,h'}\|_\infty\} \\ &\quad \le \max\left\{\sum_{n=1}^{N}(h_n - h'_n)^+, \sum_{n=1}^{N}(h'_n - h_n)^+\right\}, \end{aligned} \tag{2.6}$$

where $\|f\|_\infty = \sup_{t\in[0,T]} |f(t)|$ denotes the supremum norm.

The piecewise linear function $G_{l,h,h'}$ need not be convex. However, it can be easily compared (in all cases, including $N = \infty$) with $F_{l,h'}$, because the intervals of linearity for $F_{l,h'}$ and $G_{l,h,h'}$ coincide. The function $G_{l,h,h'}$ will play a key bridging role in the proof of Proposition 2.4 below.

Proof of Lemma 2.2. (a) The lengths of the subintervals $(a_n, a_n + l_n)$, $n \leq N$, of $[0, T]$ sum up to $\sum_{n=1}^{N} l_n = T$. By comparing the respective slopes in the definition of a_n, it follows that these intervals are pairwise disjoint. Moreover, $F_{l,h}$ is convex on $[0, T]$ and linear on every $(a_n, a_n + l_n)$. Indeed, since a function is convex if and only if it has a non-decreasing right-derivative a.e., $F_{l,h}$ is convex. Any other piecewise linear convex function with the same faces must have the same derivative as $F_{l,h}$. Furthermore, if such a function also starts at 0, it must equal $F_{l,h}$.

(b) A termwise comparison shows that

$$-\sum_{n=1}^{N}(h_n - h'_n)^+ \leq F_{l,h'} - G_{l,h,h'} \leq \sum_{n=1}^{N}(h'_n - h_n)^+,$$

pointwise. Thus, it remains to show the inequality for $\|F_{l,h} - F_{l,h'}\|_\infty$, which requires two steps.

Step 1. First assume there exists $m \in [\![N]\!]$ such that $h'_m \neq h_m$ and $h_n = h'_n$ for $n \in [\![N]\!] \setminus \{m\}$. By symmetry we may assume $h'_m > h_m$. For all $n \in [\![N]\!]$, define the slopes $s_n = h_n/l_n$ and $s'_n = h'_n/l_n$. Thus $s'_m > s_m$ and, if $n \neq m$, we have $s_n = s'_n$. Since

$$a_n = \sum_{k=1}^{N} l_k \cdot \mathbb{1}_{\{s_k < s_n\}} + \sum_{k=1}^{n-1} l_k \cdot \mathbb{1}_{\{s_k = s_n\}}, \quad a'_n = \sum_{k=1}^{N} l_k \cdot \mathbb{1}_{\{s'_k < s'_n\}} + \sum_{k=1}^{n-1} l_k \cdot \mathbb{1}_{\{s'_k = s'_n\}},$$

the right-derivatives $f_{l,h}$ and $f_{l,h'}$ of $F_{l,h}$ and $F_{l,h'}$, respectively, are piecewise constant non-decreasing functions satisfying $f_{l,h} \leq f_{l,h'}$ on $[0, T]$. Since $F_{l,h}(0) = 0 = F_{l,h'}(0)$, we deduce that $F_{l,h'} - F_{l,h} \geq 0$.

By construction, $F_{l,h'} \geq G_{l,h,h'}$ pointwise (in fact, termwise) and $\|F_{l,h'} - G_{l,h,h'}\|_\infty = h'_m - h_m$. Put $b_n = a_n + l_n$ and $b'_n = a'_n + l_n$ for $n \in [\![N]\!]$ and note that, since $s_m \leq s'_m$, we have $a_m \leq a'_m$ and

$$\begin{aligned} F_{l,h}(t) &= G_{l,h,h'}(t) = F_{l,h'}(t), && \text{for } t \in [0, a_m], \\ F_{l,h}(t) &= G_{l,h,h'}(t) \leq F_{l,h'}(t) \leq G_{l,h,h'}(t) + (h'_m - h_m), && \text{for } t \in [b'_m, T]. \end{aligned}$$

Thus, to establish (2.6) in this case, it suffices to prove that $F_{l,h'}(t) - F_{l,h}(t) \leq h'_m - h_m$ on $t \in [a_m, b'_m]$.

By construction of a'_m, the right-derivative $f_{l,h'}$ is smaller or equal to $s'_m =$

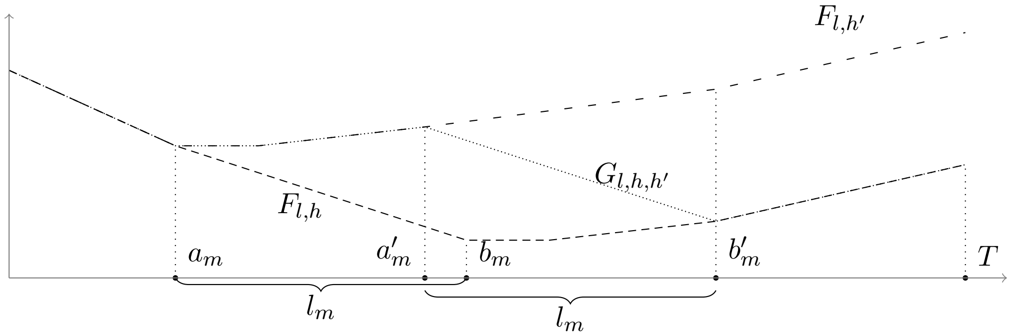

Figure 2.1: Comparison between $F_{l,h}$, $F_{l,h'}$ and $G_{l,h,h'}$.

h'_m/l_m on (a_m, b'_m). Since $F_{l,h'}(a_m) = F_{l,h}(a_m)$, for $t \in [a_m, b_m]$ we have

$$F_{l,h'}(t) - F_{l,h}(t) = \int_{a_m}^{t} f_{l,h'}(u)du - s_m(t - a_m)$$
$$\leq (s'_m - s_m)(t - a_m) \leq (s'_m - s_m)l_m = h'_m - h_m.$$

For $t \in [b_m, b'_m]$ we have $t - l_m \in [a_m, a'_m]$ and thus $F_{l,h}(t) - h_m = G(t - l_m) = F_{l,h'}(t - l_m)$. Hence

$$F_{l,h'}(t) - F_{l,h}(t) = F_{l,h'}(t) - F_{l,h'}(t - l_m) - h_m$$
$$= \int_{t-l_m}^{t} f_{l,h'}(u)du - h_m \leq \int_{t-l_m}^{t} s'_m du - h_m = h'_m - h_m.$$

Thus, $F_{l,h'} - F_{l,h} \leq h'_m - h_m$ on $[a_m, b'_m]$, proving (2.6) in this case.

Step 2. Consider the general case. For $k \in \{0, \ldots, N\}$, let $h^{(k)} = (h_n^{(k)})_{n \in [\![N]\!]}$ be given by $h_n^{(k)} = h_n \cdot \mathbb{1}_{\{n>k\}} + h'_n \cdot \mathbb{1}_{\{n \leq k\}}$ for $n \in [\![N]\!]$. Note that $h' = h^{(0)}$ and $h = h^{(N)}$. Since the sequences $h^{(k)}$ and $h^{(k-1)}$ only differ in the coordinate $h_k^{(k)} \neq h_k^{(k-1)}$, the identity $F_{l,h'} - F_{l,h} = \sum_{k=1}^{N} (F_{l,h^{(k-1)}} - F_{l,h^{(k)}})$ and **Step 1** imply (2.6), completing the proof. □

Lemma 2.3. *Let $(N_k)_{k \in \mathbb{N}}$ be a sequence in $\mathbb{N}$ with a limit $N_k \to N_\infty \in \mathbb{N}$. For each $j \in \overline{\mathbb{N}}$, let $(l_{j,n})_{n \in [\![N_j]\!]}$ be positive numbers satisfying $\sum_{n=1}^{N_j} l_{j,n} = T$, $(h_{j,n})_{n \in [\![N_j]\!]}$ real numbers and C_j the piecewise linear convex function defined in* (2.5) *with lengths $(l_{j,n})_{j \in [\![N_j]\!]}$ and heights $(h_{j,n})_{j \in [\![N_j]\!]}$. Suppose $l_{k,n} \to l_{\infty,n}$ and $h_{k,n} \to h_{\infty,n}$ as $k \to \infty$ for all $n \in [\![N_\infty]\!]$. Then $\|C_\infty - C_k\|_\infty \to 0$ as $k \to \infty$.*

Proof. The convergence $N_k \to N_\infty$ as $k \to \infty$ implies $N_k = N_\infty = N$ for all sufficiently large k. Thus, we assume without loss of generality that $N_j = N$ for all

$j \in \overline{\mathbb{N}}$. Define $s_{j,n} = h_{j,n}/l_{j,n}$ for $j \in \overline{\mathbb{N}}$ and $n \in [\![N]\!]$ and note that $s_{k,n} \to s_{\infty,n}$ as $k \to \infty$ for all $n \in [\![N]\!]$. Thus, for all sufficiently large k, if the inequality $s_{\infty,n} < s_{\infty,m}$ holds, then $s_{k,n} < s_{k,m}$. Thus, we assume this property holds for all $k \in \mathbb{N}$. Moreover, we assume without loss of generality, by relabeling if necessary, that $s_{\infty,1} \leq \cdots \leq s_{\infty,N}$.

We will next introduce a sequence of convex functions F_k satisfying the limits $\|C_\infty - F_k\|_\infty \to 0$ and $\|F_k - C_k\|_\infty \to 0$ as $k \to \infty$. These convex functions will replace each "block" of faces of C_k with a given common *limiting* slope, with a single face with the mean slope.

Let $M \leq N$ be the number of distinct slopes in $\{s_{\infty,n} : n \in [\![N]\!]\}$ and note that $s_{\infty,i_1} < \cdots < s_{\infty,i_M}$, where we set

$$i_1 = 1 \quad \text{and} \quad i_{n+1} = \min\{m \in \{i_n + 1, \ldots, N\} : s_{\infty,m} > s_{\infty,i_n}\} \quad \text{for} \quad n \in [\![M-1]\!].$$

Note that $s_{k,m} \to s_{\infty,i_n}$ as $k \to \infty$ for all $m \in \{i_n, \ldots, i_{n+1} - 1\}$. Define the sums $L_{j,n} = \sum_{m=i_n}^{i_{n+1}-1} l_{j,m}$ and $H_{j,n} = \sum_{m=i_n}^{i_{n+1}-1} h_{j,m}$ for $n \in [\![M]\!]$ and $j \in \overline{\mathbb{N}}$, where $i_{M+1} = N + 1$. Furthermore, for $j \in \overline{\mathbb{N}}$, let $(a_{j,n})_{n \in [\![N]\!]}$ be the left endpoints of the intervals in (2.5) on which C_j is linear. Note that C_∞ admits the representation

$$C_\infty(t) = \sum_{n=1}^{M} H_{\infty,n} \min\{(t - a_{\infty,i_n})^+/L_{\infty,n}, 1\} \quad \text{for} \quad t \in [0, T],$$

and define the convex functions $F_k(t) = \sum_{n=1}^{M} H_{k,n} \min\{(t - a_{k,i_n})^+/L_{k,n}, 1\}$ for $k \in \mathbb{N}$. The limits $l_{k,n} \to l_{\infty,n}$ and $h_{k,n} \to h_{\infty,n}$ imply $a_{k,i_n} \to a_{\infty,i_n}$, $L_{k,n} \to L_{\infty,n}$ and $H_{k,n} \to H_{\infty,n}$ as $k \to \infty$ for $n \in [\![M]\!]$. Thus, we have the pointwise (in fact, termwise) convergence $F_k \to C_\infty$. Since the functions are convex, the pointwise convergence implies $\|C_\infty - F_k\|_\infty \to 0$ as $k \to \infty$.

To prove that $\|F_k - C_k\|_\infty \to 0$ as $k \to \infty$, note that for $a, c \in \mathbb{R}$ and $b, d > 0$ satisfying $a/b \leq c/d$, we have $a/b \leq (a+c)/(b+d) \leq c/d$. Thus, $H_{k,n}/L_{k,n}$ lies between the smallest and largest values of $S_{k,n} = \{h_{k,i_n}/l_{k,i_n}, \ldots, h_{k,i_{n+1}-1}/l_{k,i_{n+1}-1}\}$. Since all the slopes in $S_{k,n}$ converge to s_{∞,i_n}, by the triangle inequality, we have $\max_{s \in S_{k,n}} |H_{k,n}/L_{k,n} - s| \leq \max_{s,s' \in S_{k,n}} |s' - s| \leq b_k = 2\max_{m \in [\![N]\!]} |s_{k,m} - s_{\infty,m}| \to 0$ as $k \to \infty$. Hence, the right-derivative of F_k is at most b_k away from the right-derivative of C_k, implying $\|F_k - C_k\|_\infty \leq b_k T \to 0$ as $k \to \infty$, completing the proof. □

Proposition 2.4. *Let N_k and N_∞ be $\overline{\mathbb{N}}$-valued random variables with $N_k \to N_\infty$ a.s. as $k \to \infty$. Let $(l_{j,n})_{n=1}^{N_j}$, $j \in \overline{\mathbb{N}}$, be random sequences of positive numbers* *satisfying* $\sum_{n=1}^{N_j} l_{j,n} = T$ *and* $(h_{j,n})_{n=1}^{N_j}$, $j \in \overline{\mathbb{N}}$, *sequences of random variables with* $\sum_{n=1}^{N_j} |h_{j,n}| < \infty$ *a.s. Let* C_j *be the piecewise linear convex function in* (2.5) *with sequences of lengths* $(l_{j,n})_{n=1}^{N_j}$ *and heights* $(h_{j,n})_{n=1}^{N_j}$ *for* $j \in \overline{\mathbb{N}}$. *Suppose* $l_{k,n} \to l_{\infty,n}$ *a.s. and* $h_{k,n} \to h_{\infty,n}$ *a.s. as* $k \to \infty$ *for all* $n < N_\infty + 1$ *and*

$$\lim_{M\to\infty} \limsup_{k\to\infty} \mathbb{E} \min\left\{1, \sum_{n=M}^{N_k} |h_{k,n}|\right\} = 0. \tag{2.7}$$

Then $\|C_\infty - C_k\|_\infty \xrightarrow{\mathbb{P}} 0$ *as* $k \to \infty$.

Proof. On the event $\{N_\infty < \infty\}$, by Lemma 2.3 we have $\|C_\infty - C_k\|_\infty \to 0$ a.s. as $k \to \infty$ (and (2.7) holds by our summing convention). Assume we are on the event $\{N_\infty = \infty\}$. For each $M \in \mathbb{N}$ and $j \in \overline{\mathbb{N}}$, let $C_{j,M}$ be the piecewise linear convex function in (2.5) with lengths $(l_{j,n})_{n=1}^{N_j}$ and heights $(h_{j,n}\mathbb{1}_{\{n<M\}})_{n=1}^{N_j}$. Denote $a \wedge b = \min\{a, b\}$ for any $a, b \in \mathbb{R}$. For each $j \in \overline{\mathbb{N}}$, define

$$a_{j,n} = \sum_{m=1}^{N_j} l_{j,m} \cdot \mathbb{1}_{\{h_{j,m}/l_{j,m} < h_{j,n}/l_{j,n}\}} + \sum_{m=1}^{n-1} l_{j,m} \cdot \mathbb{1}_{\{h_{j,m}/l_{j,m} = h_{j,n}/l_{j,n}\}}, \qquad n \in [\![N_j]\!],$$

end the function $G_{j,M}(t) = \sum_{m=1}^{N_j \wedge (M-1)} h_{j,n} \min\{(t - a_{j,n})^+/l_{j,n}, 1\}$, $t \in [0, T]$. Note that C_j and $G_{j,M}$ are linear on every interval $(a_{j,n}, a_{j,n} + l_{j,n})$, $n \in [\![N_j]\!]$, but $C_{j,M}$ may have different intervals of linearity. Since $1 \wedge (x + y) \le 1 \wedge x + 1 \wedge y$ for all $x, y \ge 0$, the triangle inequality implies

$$1 \wedge \|C_\infty - C_k\|_\infty \le A_{\mathbf{(I)}} + A_{\mathbf{(II)}} + A_{\mathbf{(III)}} + A_{\mathbf{(IV)}} + A_{\mathbf{(V)}}, \tag{2.8}$$

where $A_{\mathbf{(I)}} = 1 \wedge \|C_\infty - G_{\infty,M}\|_\infty$, $A_{\mathbf{(II)}} = 1 \wedge \|G_{\infty,M} - C_{\infty,M}\|_\infty$, $A_{\mathbf{(III)}} = 1 \wedge \|C_{\infty,M} - C_{k,M}\|_\infty$, $A_{\mathbf{(IV)}} = 1 \wedge \|C_{k,M} - G_{k,M}\|_\infty$ and $A_{\mathbf{(V)}} = 1 \wedge \|G_{k,M} - C_k\|_\infty$. As $\zeta_n \xrightarrow{\mathbb{P}} 0$ as $n \to \infty$ if and only if $\mathbb{E}[1 \wedge |\zeta_n|] \to 0$, it suffices to prove that the expectation of each of the terms in (2.8) converges to 0 as we take $\limsup_{k\to\infty}$ and then $M \to \infty$.

(I)&**(V)**. By construction of C_j and $G_{j,M}$ we have $\|C_j - G_{j,M}\|_\infty \le \sum_{n=M}^{N_j} |h_{j,n}|$ for all $j \in \overline{\mathbb{N}}$. Thus, we have $\|C_\infty - G_{\infty,M}\|_\infty \to 0$ a.s. and hence $\mathbb{E}A_{\mathbf{(I)}} =$

$\mathbb{E}[1 \wedge \|C_\infty - G_{\infty,M}\|_\infty] \to 0$ as $M \to \infty$. Moreover, by assumption in (2.7),

$$\limsup_{k\to\infty} \mathbb{E}A_{(\mathbf{V})} = \limsup_{k\to\infty} \mathbb{E}[1 \wedge \|C_k - G_{k,M}\|_\infty] \leq \limsup_{k\to\infty} \mathbb{E}\min\left\{1, \sum_{n=M}^{N_k} |h_{k,n}|\right\} \xrightarrow[M\to\infty]{} 0.$$

(III). For all $j \in \overline{\mathbb{N}}$, the faces of $C_{j,M}$ corresponding to $n \in [\![N_j]\!] \setminus [\![M-1]\!]$ are horizontal. By convexity, we may assume they lie next to each other in the graph of $C_{j,M}$. Merging all the lengths $l_{j,n}$, $n \in [\![N_j]\!] \setminus [\![M-1]\!]$, yields a representation of $C_{j,M}$ with $N_j \wedge M$ faces. Fix $M \in \mathbb{N}$. Lemma 2.3 yields $\|C_{\infty,M} - C_{k,M}\|_\infty \to 0$ a.s. and thus $\mathbb{E}A_{(\mathbf{III})} = \mathbb{E}[1 \wedge \|C_{\infty,M} - C_{k,M}\|_\infty] \to 0$ as $k \to \infty$.

(II)&(IV). The idea is to apply (2.6) in Lemma 2.2(b) to bound $\|C_{j,M} - G_{j,M}\|_\infty$, with $F_{l,h}$, $G_{l,h,h'}$ and $F_{l,h'}$ in Lemma 2.2(b) given by $C_{j,M}$, $G_{j,M}$ and $F_{j,M}$, respectively. The piecewise linear convex function $F_{j,M}$, which shares the intervals of linearity with those of $G_{j,M}$, is yet to be defined.

Note that $G_{j,M}$ possesses a piecewise linear representation with at most $2M$ faces. Indeed, $G_{j,M}$ is linear on $(a_{j,n}, a_{j,n} + l_n)$, $n \in [\![N_j \wedge (M-1)]\!]$, and the complement $(0,T) \setminus \bigcup_{n=1}^{N_j\wedge(M-1)} [a_{j,n}, a_{j,n} + l_{j,n}]$ is a disjoint union of $M_j \leq M+1$ open intervals, say $(a'_{j,n}, a'_{j,n} + l'_{j,n})$, $n \in [\![M_j]\!]$. For each $n \in [\![M_j]\!]$, define the height $h'_{j,n} = \sum_{m\in S_{j,n}} h_{j,m}$, where $S_{j,n} = \{m \in [\![N_j]\!] \setminus [\![M-1]\!] : a_{j,m} \in (a'_{j,n}, a'_{j,n} + l'_{j,n})\}$. Put differently, the height $h'_{j,n}$ equals the sum of all the heights of the faces of C_j that lie above the interval $[a'_{j,n}, a'_{j,n} + l'_{j,n}]$. For any $j \in \overline{\mathbb{N}}$ and $t \in [0,T]$, define

$$F_{j,M}(t) = \sum_{n=1}^{N_j\wedge(M-1)} h_{j,n} \min\{(t - a_{j,n})^+/l_{j,n}, 1\} + \sum_{n=1}^{M_j} h'_{j,n} \min\{(t - a'_{j,n})^+/l'_{j,n}, 1\}. \tag{2.9}$$

We will show that $F_{j,M}$ is convex. It suffices to prove that the consecutive slopes of $F_{j,M}$ on adjacent intervals of linearity increase. If the consecutive intervals are $(a_{j,m}, a_{j,m} + l_{j,m})$ and $(a_{j,n}, a_{j,n} + l_{j,n})$ (i.e. they come from the first sum in (2.9)), then by construction the intervals must be adjacent with the same slopes in the convex function C_j, implying the corresponding slopes satisfy the correct ordering. Assume the consecutive intervals are $(a_{j,m}, a_{j,m} + l_{j,m})$ and $(a'_{j,n}, a'_{j,n} + l'_{j,n})$ (i.e. the first interval comes from first sum and the second interval comes from the second sum in (2.9)). Suppose $a_{j,m} = a'_{j,n} + l'_{j,n}$ and note that, for $a, c \in \mathbb{R}$ and $b, d > 0$ with $a/b \leq c/d$ we have $a/b \leq (a+c)/(b+d) \leq c/d$. Thus, by definition of $h'_{j,n}$, we have $h'_{j,n}/l'_{j,n} \le \sup_{i\in S_{j,n}} h_{j,i}/l_{j,i} \le h_{j,m}/l_{j,m}$, where the last inequality holds because $a_{j,m} = a'_{j,n} + l'_{j,n}$ and C_j is convex. The case $a'_{j,n} = a_{j,m} + l_{j,n}$ is analogous since the slope $h'_{j,n}/l'_{j,n}$ is a mean of slopes at least as large as $h_{j,m}/l_{j,m}$, implying the convexity of $F_{j,M}$.

Define the vectors

$$\begin{aligned} l &= (l_{j,1}, \dots, l_{j,N_j\wedge(M-1)}, l'_{j,1}, \dots, l'_{j,M_j}), \\ h &= (h_{j,1}, \dots, h_{j,N_j\wedge(M-1)}, h'_{j,1}, \dots, h'_{j,M_j}) \quad \text{and} \\ h' &= (h_{j,1}, \dots, h_{j,N_j\wedge(M-1)}, 0, \dots, 0). \end{aligned}$$

Note that the corresponding functions $F_{l,h}$, $F_{l,h'}$ and $G_{l,h,h'}$ in (2.5) equal $F_{j,M}$, $C_{j,M}$ and $G_{j,M}$, so (2.6) implies the inequality

$$\|G_{j,M} - C_{j,M}\|_\infty \le \|G_{j,M} - F_{j,M}\|_\infty + \|F_{j,M} - C_{j,M}\|_\infty \le 2\sum_{m=1}^{M_j} |h'_{j,m}| \le 2\sum_{n=M}^{N_j} |h_{j,n}|.$$

Thus, $\|G_{\infty,M} - C_{\infty,M}\|_\infty \to 0$ a.s. and hence $\mathbb{E}A_{\mathbf{(II)}} = \mathbb{E}[1 \wedge \|C_\infty - G_{\infty,M}\|_\infty] \to 0$ as $M \to \infty$. Moreover, by assumption in (2.7), we have

$$\begin{aligned} \limsup_{k\to\infty} \mathbb{E}A_{\mathbf{(IV)}} &= \limsup_{k\to\infty} \mathbb{E}[1 \wedge \|G_{k,M} - C_{k,M}\|_\infty] \\ &\le 2\limsup_{k\to\infty} \mathbb{E}\min\left\{1, \sum_{n=M}^{N_k} |h_{k,n}|\right\} \xrightarrow[M\to\infty]{} 0. \quad \square \end{aligned}$$

§2.4 The convex minorant of random walks

Let a function $f : [0,T] \to \mathbb{R}$ satisfy $f(0) = 0$. Given parameters $0 \le g \le u \le d \le T$, the *3214 transformation*, introduced in [1], is defined by

$$\Theta_{g,u,d} f(t) = \begin{cases} f(u+t) - f(u), & 0 \le t \le d-u, \\ f(d) - f(u) + f(g+t-(d-u)) - f(g), & d-u < t \le d-g, \\ f(d) - f(t-(d-g)), & d-g < t \le d, \\ f(t), & d < t. \end{cases}$$

The 3214 transformation reorders the segments of the graph of f as follows: the segments (I) $[0,g]$, (II) $[g,u]$, (III) $[u,d]$ and (IV) $[d,T]$ are moved to (III) $[0,d-u]$, (II) $[d-u,d-g]$, (I) $[d-g,d]$ and (IV) $[d,T]$, respectively (see also Figure 2.2 below). This transformation possesses the following remarkable property when applied to continuous piecewise linear functions with a given set of increments.

Proposition 2.5 ([1, Thm 1])**.** *Fix $n \in \mathbb{N}$ and let $x_1, \ldots, x_n$ be real numbers, such that no two subsets have the same mean. Let $\lfloor y \rfloor = \max\{m \in \mathbb{Z} : m \le y\}$, $y \in \mathbb{R}$, and $\pi : [\![n]\!] \to [\![n]\!]$ be a uniform random permutation. Define the polygonal random walk $R = (R(t))_{t \in [0,T]}$ by $R(T) = \sum_{k=1}^n x_k$ and*

$$R(t) = \sum_{k=1}^{\lfloor nt/T \rfloor} x_{\pi(k)} + (nt/T - \lfloor nt/T \rfloor)\, x_{\pi(\lfloor nt/T \rfloor + 1)}, \qquad t \in [0, T). \tag{2.10}$$

Let C_T^R denote the convex minorant of R and let $W \sim \mathrm{U}(0,T)$ be independent of R. Let $0 = V_0 < \cdots < V_N = T$ be the sequence of contact points between the piecewise linear functions R and C_T^R and $j \in [\![N]\!]$ the unique index such that $W \in (V_{j-1}, V_j]$. Define $U = \lceil Wn/T \rceil T/n$, $G = V_{j-1}$ and $D = V_j$. Then the 3214 transform with parameters (G, U, D) satisfies the identity in law

$$(U, R) \overset{d}{=} (D - G, \Theta_{G,U,D} R).$$

For completeness, we recall below a proof of Proposition 2.5 using a simple argument from [2].

Theorem 2.6. *Let $x_1, \ldots, x_n$ be arbitrary real numbers and $\pi : [\![n]\!] \to [\![n]\!]$ a uniform random permutation. Define R by (2.10) and let $(V_k)_{k \in \mathbb{N}}$ be an iid sequence of $\mathrm{U}(0,1)$ random variables independent of π. Define recursively $L_{n,0} = T$, $L_{n,k} = \lfloor L_{n,k-1} V_k n/T \rfloor T/n$, $\ell_{n,k} = L_{n,k-1} - L_{n,k}$ for $k \in \mathbb{N}$ and let $N \le n$ be the largest integer for which $\ell_{n,N} > 0$. Then the convex minorant C_T^R has the same law as the piecewise linear convex function defined in (2.5) with sequences of lengths $(\ell_{n,k})_{k=1}^N$ and heights $(R(L_{n,k-1}) - R(L_{n,k}))_{k=1}^N$.*

We stress that in Theorem 2.6, the reals $x_1, \ldots, x_n$ may have multiple subsets with the same mean. Our proof approximates a general sequence by one satisfying the "no ties" assumption of Proposition 2.5 and applies a convergence result for piecewise linear convex functions from §2.3. The proof of Theorem 2.6 in [1] subsamples the ties, resulting in a more involved statement of the theorem.

Proof of Theorem 2.6. First assume that no two subsets of the numbers $x_1, \ldots, x_n$ have the same mean. Let π and (G, U, D) be as in Proposition 2.5. By Proposition 2.5, the face decomposition of C_T^R contains the face with length-height pair equal to $(D - G, C_T^R(D) - C_T^R(G))$, which has the same law as $(U, R(U))$, and the faces of a copy of $C_{T-(D-G)}^R$ independent of the first face. Indeed, this copy is in fact the convex minorant of $\Theta_{G,U,D}R$ on $[T-(D-G),T]$ and we may apply the same procedure to this copy. Iterating this procedure, we obtain a (finite) sequence of lengths of the faces of C_T^R, which has the same law as the sequence $(\ell_{k,n})_{k=1}^N$, and the corresponding heights, which have the same law as $(R(L_{k-1,n})-R(L_{k,n}))_{k=1}^N$, completing the proof in this case.

To prove the general case, we recall that $\|C_T^x - C_T^y\|_\infty \le \|x-y\|_\infty$ for any bounded functions $x,y:[0,T]\to\mathbb{R}$. Indeed, this follows from the fact that the function $C_T^x - \|x-y\|_\infty$ is convex and $C_T^x - \|x-y\|_\infty \le x - \|x-y\|_\infty \le y$ pointwise. For any $\varepsilon>0$ consider real numbers $x_{1,\varepsilon},\ldots,x_{n,\varepsilon}$ such that no two subsets have the same mean and $\sum_{k=1}^n |x_k - x_{k,\varepsilon}| \le \varepsilon$. Let R_ε be the corresponding random walk in (2.10) (with the same permutation π). Note that $\|C_T^R - C_T^{R_\varepsilon}\|_\infty \le \|R-R_\varepsilon\|_\infty \le \varepsilon \to 0$ as $\varepsilon\to 0$. Moreover, by the argument in the previous paragraph, $C_T^{R_\varepsilon}$ has the same law as the piecewise linear convex function C_ε given by (2.5) with lengths $(\ell_{n,k})_{k=1}^N$ and heights $(R_\varepsilon(L_{n,k-1})-R_\varepsilon(L_{n,k}))_{k=1}^N$. Let C be the piecewise linear convex function given by (2.5) with lengths $(\ell_{n,k})_{k=1}^N$ and heights $(R(L_{n,k-1}) - R(L_{n,k}))_{k=1}^N$. Lemma 2.3 yields $\|C-C_\varepsilon\|_\infty \to 0$ a.s. as $\varepsilon\to 0$, implying $C \overset{d}{=} C_T^R$ and completing the proof. □

The proof of Proposition 2.5 requires the following lemma.

Lemma 2.7. *Let $x_1,\ldots,x_n$ be real numbers such that no two subsets have the same mean. Then there is a unique $k^*\in[\![n]\!]$ such that $\sum_{i=1}^k x_{(k^*+i) \bmod n} \ge \frac{k}{n}\sum_{i=1}^n x_i$ for all $k\in[\![n]\!]$, i.e. the walk with increments $x_{(k^*+1) \bmod n},\ldots,x_{(k^*+n) \bmod n}$ is above the line connecting zero with the endpoint $\sum_{i=1}^n x_i$.*

Proof. Define $s=\sum_{i=1}^n x_i/n$. If the walk $k\mapsto \sum_{i=1}^k (x_i - s)$, $k\in[\![n]\!]$, attained its minimum at two times $k_1<k_2$, then $\sum_{i=k_1+1}^{k_2} x_i/(k_2-k_1) = s$, contradicting the assumption. It is easily seen that the k^* in the statement of the lemma is the time at which this walk attains its minimum on $[\![n]\!]$. □

Proof of Proposition 2.5 ([2]). If a random element ζ is uniformly distributed in some finite set $\mathcal{Z}$ and if the map $\varphi:\mathcal{Z}\to\mathcal{Z}$ is injective (and thus bijective), then $\varphi(\zeta)$ is also uniformly distributed on $\mathcal{Z}$. Thus, since π and U are uniform and independent, it is sufficient to show that the transformation $(u,f)\mapsto(d-g,\Theta_{g,u,d}f)$ is injective.

Assume without loss of generality that $T=n$. To prove the injectivity, it suffices to describe the inverse transformation. Given $d-g$, and $\tilde f = \Theta_{g,u,d}f$, note that $d-u$ is the unique time in Lemma 2.7 for the increments of $\tilde f$ over the set $[\![d-g]\!]$, see Figure 2.2. Consider the convex minorant of $\tilde f$ on the interval $[d-g,T]$ and

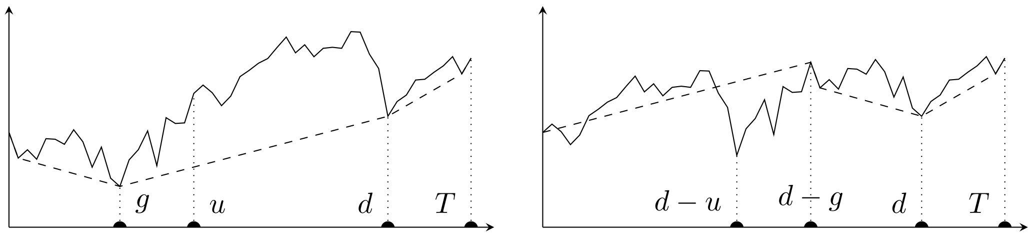

Figure 2.2: The pictures show a path of a random walk R (*solid*) and its convex minorant C_T^R (*dashed*) on $[0,T]$ on the left and their 3214 transforms on the right. The transform is associated to some $u \in (0,T)$ and the endpoints $\{g,d\}$ of the maximal face of C_T^R containing u.

note that d is the right end of the last face whose slope is less than $\tilde{f}(d-g)/(d-g)$. Thus we may identify d, u and g and then invert the 3214 transform to recover f. This shows that $(u,f) \mapsto (d-g, \Theta_{g,u,d} f)$ is injective, completing the proof. □

§2.5 Proof of the stick-breaking representation

The proof of Theorem 2.1 is split into 3 steps.

Step 1. Let $\widetilde{C}_k$ be the largest convex function on $[0,T]$ that is smaller than X pointwise on the set $D_k = \{Tn/2^k : n \in \{0,1,\ldots,2^k\}\}$. Since $D_k \subset D_{k+1}$, we have $\widetilde{C}_k(t) \geq \widetilde{C}_{k+1}(t)$ for all $t \in [0,T]$. Moreover, the limit $\widetilde{C}_\infty = \lim_{k\to\infty} \widetilde{C}_k$ is clearly convex and smaller than X pointwise on the dense set $\bigcup_{k\in\mathbb{N}} D_k$ in $[0,T]$. As X is càdlàg, $\widetilde{C}_\infty$ is pointwise smaller than X on $[0,T]$, implying $\widetilde{C}_\infty \leq C_T^X$. Since C_T^X is convex and smaller than X on D_k, the maximality of $\widetilde{C}_k$ yields $\widetilde{C}_k \geq C_T^X$ for all $k \in \mathbb{N}$, implying $\widetilde{C}_\infty \geq C_T^X$ and thus $\widetilde{C}_\infty = C_T^X$.

Step 2. Let $U_1, U_2, \ldots$ be iid $\mathrm{U}(0,1)$ random variables independent of X. Let $L_0 = T$, $L_n = U_n L_{n-1}$, $\ell_n = L_{n-1} - L_n$ and $\xi_n = X_{L_{n-1}} - X_{L_n}$ for $n \in \mathbb{N}$. For each $k \in \mathbb{N}$, define $L_{k,0} = T$, $L_{k,n} = \lfloor L_{k,n-1} U_n 2^k / T \rfloor T/2^k$, $\ell_{k,n} = L_{k,n-1} - L_{k,n}$ and $\xi_{k,n} = X_{L_{k,n-1}} - X_{L_{k,n}}$ for $n \in \mathbb{N}$. Let N_k be the largest natural number for which $\ell_{k,N_k} > 0$, so that $\ell_{k,n}$, $L_{k,n}$ and $\xi_{k,n}$ are all zero for all $n > N_k$. For each $k \in \mathbb{N}$, let C_k (resp. C_∞) be the piecewise linear convex function given in (2.5) with lengths $(\ell_{k,n})_{n=1}^{N_k}$ (resp. $(\ell_n)_{n=1}^\infty$) and heights $(\xi_{k,n})_{n=1}^{N_k}$ (resp. $(\xi_n)_{n=1}^\infty$). Next we show that $\|C_k - C_\infty\|_\infty \xrightarrow{\mathbb{P}} 0$ as $k \to \infty$. Since X has càdlàg paths with countably many jumps, L_n has a density for every $n \in \mathbb{N}$ and $L_{k,n} \to L_n$ a.s. as $k \to \infty$, we have $\xi_{k,n} \to \xi_n$ a.s. as $k \to \infty$ for all $n \in \mathbb{N}$. Thus, by Proposition 2.4, it suffices to prove that $\lim_{M\to\infty} \limsup_{k\to\infty} \mathbb{E}[1 \wedge P_{k,M}] = 0$, where $P_{k,M} = \sum_{n=M}^{N_k} |\xi_{k_m,n}|$.

Theorem 2.6 implies that $C_k \overset{d}{=} \widetilde{C}_k$. Let R_k be the continuous piecewise linear function connecting the skeleton of X on D_k with line segments. Since the minimum and the final value of the convex minorant $C_T^{R_k} = \widetilde{C}_k$ agree with the corresponding functionals of R_k, the total variation $\sum_{n=1}^{N_k} |\xi_{k,n}|$ of C_k has the same distribution as $X_T - 2\min_{t\in D_k} X_t$. Moreover, by the independence and the definition of $(L_{k,n})_{n\in\mathbb{N}}$, it is easily seen that $P_{k,M} = \sum_{n=M+1}^{N_k} |\xi_{k,n}| \overset{d}{=} X_{L_{k,M}} - 2\min_{t\in D_k\cap[0,L_{k,M}]} X_t$. By the inequality $L_{k,M} \le L_M$, we have

$$X_{L_{k,M}} - 2 \min_{t\in D_k\cap[0,L_{k,M}]} X_t \le X_{L_{k,M}} - 2\underline{X}_{L_{k,M}} \le \overline{X}_{L_M} - 2\underline{X}_{L_M}.$$

Since $L_M \to 0$ a.s. as $M \to \infty$ and $\overline{X}_t - 2\underline{X}_t \to 0$ a.s. as $t \to 0$, we have $\overline{X}_{L_M} - 2\underline{X}_{L_M} \to 0$ a.s. as $M \to \infty$, implying

$$\limsup_{k\to\infty} \mathbb{E}[1 \wedge P_{k,M}] \le \mathbb{E}[1 \wedge (\overline{X}_{L_M} - 2\underline{X}_{L_M})] \xrightarrow[M\to\infty]{} 0.$$

Step 3. Recall that, by Theorem 2.6, we have $C_k \overset{d}{=} \widetilde{C}_k$. Since $\|C_k - C_\infty\|_\infty \overset{\mathbb{P}}{\to} 0$ and $\|\widetilde{C}_k - C_T^X\|_\infty \to 0$ a.s. as $k \to \infty$, we conclude that $C_\infty \overset{d}{=} C_T^X$, implying Theorem 2.1. □

§2.6 The fluctuation theory of Lévy processes

The power of the stick-breaking representation in Theorem 2.1 (or its corollary in (2.2) for the vector $\overline{\chi}_T(X)$), for any fixed time horizon T lies in the fact that (2.2) essentially reduce the properties of the path functional $\overline{\chi}_T(X)$ to the properties of the marginals of X. We now illustrate this by deriving many of the classical highlights of the fluctuation theory of Lévy processes from Theorem 2.1. Note first that, since $-\log(\ell_n/T)$ is gamma distributed with density $s \mapsto s^{n-1}\mathrm{e}^{-s}/(n-1)!$ for $s > 0$, for a measurable $f : [0,T] \to \mathbb{R}_+$ we have

$$\mathbb{E}\sum_{n\in\mathbb{N}} f(\ell_n) = \int_0^T s^{-1} f(s)\mathrm{d}s. \tag{2.11}$$

In the definition of $\overline{\tau}_t(X)$ (resp. $\underline{\tau}_t(X)$), we take the first rather than last time the maximum (resp. minimum) is attained. Our first corollary shows that this choice makes little difference.

Corollary 2.8. *A Lévy process X attains its maximum at a unique time a.s. if and only if X is not a driftless compound Poisson process; then we have $\overline{\chi}_t(X) \overset{d}{=} (X_t, X_t - \underline{X}_t, t - \underline{\tau}_t(X))$ for all $t > 0$.*

Proof. If X is a driftless compound Poisson process, it has piecewise constant paths, making the time of the maximum not unique. Assume X is not a driftless compound Poisson, then $\mathbb{P}(X_t = 0) > 0$ for at most countably many $t > 0$. Indeed, either the law of X_t is diffuse or, by Doeblin's lemma [62, Lem. 15.22], X is compound Poisson with drift $\mu \neq 0$. In the latter case, $\mathbb{P}(X_t = 0) > 0$ if and only if $-\mu t$ is in a countable set generated by the atoms of the Lévy measure of X, implying the claim.

The time-reversal process $X' = (X'_s)_{s\in[0,t]}$, defined as $X'_s := X_{(t-s)-} - X_t$, $s \in [0, t]$, has the same law as $(-X_s)_{s\in[0,t]}$, implying $\overline{\tau}_t(X') \overset{d}{=} \underline{\tau}_t(X)$. The gap $(t - \overline{\tau}_t(X')) - \overline{\tau}_t(X) \geq 0$ between the time of the first and last maximum of X has expectation equal to zero and is hence zero a.s. Indeed, since $\mathbb{P}(X_t = 0) > 0$ for at most countably many $t > 0$, Theorem 2.1 and (2.11) yield

$$\begin{aligned} t - \mathbb{E}[\underline{\tau}_t(X)] - \mathbb{E}[\overline{\tau}_t(X)] &= \mathbb{E}\sum_{n=1}^{\infty} \ell_n \mathbb{1}_{\{X_{L_{n-1}}=X_{L_n}\}} \\ &= \mathbb{E}\sum_{n=1}^{\infty} \ell_n \mathbb{P}(X_{\ell_n} = 0|\ell_n) = \int_0^t \mathbb{P}(X_s = 0)\mathrm{d}s = 0. \end{aligned}$$

The identity in law follows from $\overline{\chi}_t(X') \overset{d}{=} \overline{\chi}_t(-X)$. $\square$

Corollary 2.9. *For any Lévy process X, the following formulae hold for any $t > 0$:*

$$\mathbb{E}[\overline{\tau}_t(X)] = \int_0^t \mathbb{P}(X_s > 0)\mathrm{d}s \qquad \text{and} \qquad \mathbb{E}[\overline{X}_t] = \int_0^t (\mathbb{E}\max\{X_s, 0\}/s)\mathrm{d}s. \quad (2.12)$$

Proof. Denote $\rho(s) := \mathbb{P}(X_s > 0)$ and take expectations in the third coordinate of the SB-representation in (2.3). Fubini's theorem and the formula in (2.11) imply

$$\mathbb{E}\overline{\tau}_t(X) = \sum_{n=1}^{\infty} \mathbb{E}[\ell_n \rho(\ell_n)] = \int_0^t \rho(s)\mathrm{d}s \quad \text{for any } t > 0.$$

The proof of the formula for the supremum is analogous. $\square$

Consider an exponential time horizon $T_\theta \sim \mathrm{Exp}(\theta)$ with parameter $\theta \in (0, \infty)$ (i.e. $\mathbb{E}T_\theta = 1/\theta$), independent of the Lévy process X. Let $\ell^{(\theta)} = (\ell_n^{(\theta)})_{n\in\mathbb{N}}$ be a stick-breaking process with a random time horizon T_θ. The random measure $\sum_{n=1}^{\infty} \delta_{\ell_n^{(\theta)}}$ on $(0, \infty)$ is easily seen to be a Poisson point process (see Subsection §2.6.1 below). By (2.11) its mean measure satisfies $\mathbb{E}\sum_{n\in\mathbb{N}} \delta_{\ell_n^{(\theta)}}(A) = \int_A t^{-1}\mathrm{e}^{-\theta t}\mathrm{d}t$ for a measurable set A (δ_z denotes the Dirac delta at the point z). Let $F(t, \mathrm{d}x) := \mathbb{P}(X_t \in \mathrm{d}x)$ denote the law of X_t for any $t > 0$. Marking each point $\ell_n^{(\theta)}$ by a random real number sampled from the law $F(\ell_n^{(\theta)}, \cdot)$, by the Marking Theorem [64, p. 55], produces a Poisson point process on $(0,\infty)\times\mathbb{R}$.

Proposition 2.10. *Let the time horizon $T_\theta \sim \mathrm{Exp}(\theta)$ and the stick-breaking process $\ell^{(\theta)}$ be independent of the Lévy process X. Define $\xi_n^{(\theta)} := X_{L_{n-1}^{(\theta)}} - X_{L_n^{(\theta)}}$, where $L^{(\theta)} = (L_k^{(\theta)})_{k\in\mathbb{N}\cup\{0\}}$ is the remainder process associated to $\ell^{(\theta)}$. Then $\Xi_\theta := \sum_{n=1}^\infty \delta_{(\ell_n^{(\theta)},\xi_n^{(\theta)})}$ is a Poisson point process with mean measure*

$$\mu_\theta(\mathrm{d}t,\mathrm{d}x) := t^{-1}\mathrm{e}^{-\theta t}\mathbb{P}(X_t\in\mathrm{d}x)\mathrm{d}t, \qquad (t,x)\in(0,\infty)\times\mathbb{R}.$$

An immediate corollary of Theorem 2.1 and Proposition 2.10 characterises the laws of the supremum and infimum of X on the exponential time horizon T_θ.

Corollary 2.11. *Let $T_\theta \sim \mathrm{Exp}(\theta)$ be independent of the Lévy process X. Then the moment generating functions of $\overline{X}_{T_\theta}$ and $-\underline{X}_{T_\theta}$ are given by the following formulae for any $u\geq 0$:*

$$\mathbb{E}\big[\mathrm{e}^{-u\overline{X}_{T_\theta}}\big] = \exp\bigg(\int_0^\infty\int_{(0,\infty)}(\mathrm{e}^{-ux}-1)\mathrm{e}^{-\theta t}t^{-1}\mathbb{P}(X_t\in\mathrm{d}x)\mathrm{d}t\bigg), \tag{2.13}$$

$$\mathbb{E}\big[\mathrm{e}^{u\underline{X}_{T_\theta}}\big] = \exp\bigg(\int_0^\infty\int_{(-\infty,0)}(\mathrm{e}^{ux}-1)\mathrm{e}^{-\theta t}t^{-1}\mathbb{P}(X_t\in\mathrm{d}x)\mathrm{d}t\bigg). \tag{2.14}$$

Proof. By Theorem 2.1, $\overline{X}_{T_\theta} \overset{d}{=} \int_{(0,\infty)^2} x\Xi_\theta(\mathrm{d}t,\mathrm{d}x)$, where Ξ_θ is a Poisson point process with mean measure μ_θ. Campbell's formula [64, p. 28] implies (2.13). Applying (2.13) to $-X$ yields (2.14). □

Recall that 0 is *regular* for the half-line $(0,\infty)$ if X visits $(0,\infty)$ almost surely immediately after time 0, i.e. $\mathbb{P}(\bigcap_{t>0}\bigcup_{s\leq t}\{X_s>0\})=1$.

Theorem 2.12 (Rogozin's criterion)**.** *The starting point 0 of X is regular for $(0,\infty)$ if and only if*

$$\int_0^1 t^{-1}\mathbb{P}(X_t>0)\mathrm{d}t=\infty. \tag{2.15}$$

Proof. Let the time horizon $T_\theta\sim\mathrm{Exp}(\theta)$ and random sequences $\ell^{(\theta)}$ and $\xi^{(\theta)}$ be as in Proposition 2.10 above. As $t\mapsto\overline{X}_t$ is non-decreasing a.s., 0 is not regular for $(0,\infty)$ if and only if $\mathbb{P}(\overline{X}_{T_\theta}=0)>0$. Since $\overline{X}_{T_\theta}\overset{d}{=}\int_A x\Xi_\theta(\mathrm{d}t,\mathrm{d}x)$, where $A:=(0,\infty)\times(0,\infty)$, the event $\{\overline{X}_{T_\theta}=0\}$ is equal to the event $\{\Xi_\theta(A)=0\}$ that the Poisson point process $\Xi_\theta=\sum_{n=1}^\infty\delta_{(\ell_n^{(\theta)},\xi_n^{(\theta)})}$ has no points in A. Thus, 0 is not regular for $(0,\infty)$ if and only if

$$\mathbb{P}\left(\overline{X}_{T_\theta}=0\right)=\mathbb{P}\big(\Xi_\theta\,(A)=0\big)=\exp\big(-\mathbb{E}\Xi_\theta\,(A)\,\big)$$
$$=\exp\left(-\int_0^\infty t^{-1}\mathrm{e}^{-\theta t}\mathbb{P}(X_t>0)\mathrm{d}t\right)>0$$

for some positive θ, which is equivalent to (2.15). □

We can now characterise the behaviour of X as $t\to\infty$.

Theorem 2.13 (Rogozin)**.** *Possibly degenerate variables* $\overline{X}_\infty := \sup_{t\geq 0} X_t$ *and* $\underline{X}_\infty := \inf_{t\geq 0} X_t$ *satisfy*

$$\mathbb{E}\big[\mathrm{e}^{-u\overline{X}_\infty}\big]=\exp\left(\int_0^\infty\int_{(0,\infty)}\big(\mathrm{e}^{-ux}-1\big)t^{-1}\mathbb{P}(X_t\in\mathrm{d}x)\mathrm{d}t\right),\tag{2.16}$$
$$\mathbb{E}\big[\mathrm{e}^{u\underline{X}_\infty}\big]=\exp\left(\int_0^\infty\int_{(-\infty,0)}\big(\mathrm{e}^{ux}-1\big)t^{-1}\mathbb{P}(X_t\in\mathrm{d}x)\mathrm{d}t\right),\tag{2.17}$$

for any $u\geq 0$*. Define the integrals*

$$I_+ := \int_1^\infty t^{-1}\mathbb{P}(X_t>0)\mathrm{d}t \qquad \& \qquad I_- := \int_1^\infty t^{-1}\mathbb{P}(X_t<0)\mathrm{d}t.$$

Then the following statements hold for any non-constant Lévy process X*:*

(a) *if* $I_+<\infty$*, then* $\overline{X}_\infty$ *is non-degenerate (*$\overline{X}_\infty<\infty$ *a.s.) infinitely divisible and* $\lim_{t\to\infty} X_t=-\infty$*;*

(b) *if* $I_-<\infty$*, then* $\underline{X}_\infty$ *is non-degenerate (*$\underline{X}_\infty>-\infty$ *a.s.) infinitely divisible and* $\lim_{t\to\infty} X_t=\infty$*;*

(c) *if* $I_+=I_-=\infty$*, then* $\limsup_{t\to\infty} X_t=-\liminf_{t\to\infty} X_t=\infty$*.*

Proof. Let $T_1\sim\mathrm{Exp}(1)$ be independent of X and note $T_1/\theta\sim\mathrm{Exp}(\theta)$ for any $\theta>0$. Since $\overline{X}_{T_1/\theta}\to\overline{X}_\infty$ as $\theta\to 0$ a.s., the corresponding Laplace transforms converge pointwise. Thus the monotone convergence theorem applied to the right-hand sides of (2.13)–(2.14) implies (2.16)–(2.17). Identity (2.16) (resp. (2.17)) implies that $I_+<\infty$ (resp. $I_-<\infty$) if and only if $\mathbb{E}\exp(-u\overline{X}_\infty)>0$ (resp. $\mathbb{E}\exp(u\underline{X}_\infty)>0$) for all $u\geq 0$. This implies part (c) and all but the limits in parts (a) & (b). The limits in (a) & (b) follow from identities (2.16)–(2.17), the strong Markov property, the Borel–Cantelli lemma and simple manipulations, see [91, p. 365] for details. □

Another easy corollary of Theorem 2.1 is the Wiener-Hopf factorisation.

Theorem 2.14 (Wiener-Hopf factorisation)**.** *Let the time horizon $T_\theta \sim \mathrm{Exp}(\theta)$ be independent of X. The random vectors $(\overline{\tau}_{T_\theta}(X), \overline{X}_{T_\theta})$ and $(T_\theta - \overline{\tau}_{T_\theta}(X), X_{T_\theta} - \overline{X}_{T_\theta})$ are independent, infinitely divisible with Fourier-Laplace transforms given by*

$$\Psi_\theta^+(u,v) := \mathbb{E}\big[\mathrm{e}^{u\overline{\tau}_{T_\theta}(X)+v\overline{X}_{T_\theta}}\big] = \frac{\varphi_+(-\theta,0)}{\varphi_+(u-\theta,v)}, \tag{2.18}$$

$$\Psi_\theta^-(u,-v) := \mathbb{E}\big[\mathrm{e}^{u(T_\theta-\overline{\tau}_{T_\theta}(X))-v(X_{T_\theta}-\overline{X}_{T_\theta})}\big] = \frac{\varphi_-(-\theta,0)}{\varphi_-(u-\theta,v)}, \tag{2.19}$$

for any $u,v \in \mathbb{C}$ with $\Re u, \Re v \le 0$. Here $\varphi_\pm$ is defined as follows: set $A_+ := (0,\infty)$, $A_- := (-\infty, 0]$,

$$\varphi_\pm(a,b) := \exp\left(\int_0^\infty \int_{A_\pm} \big(\mathrm{e}^{-t} - \mathrm{e}^{at+b|x|}\big)t^{-1}\mathbb{P}(X_t \in \mathrm{d}x)\mathrm{d}t\right), \tag{2.20}$$

for any $a,b \in \mathbb{C}$ such that the integrals in (2.20) *exist, including the cases where $\Re a < 0$, $\Re b \le 0$. The characteristic exponent Ψ of X_1 (i.e. $\mathbb{E}\exp(vX_1) = \exp\Psi(v)$ for $v \in \mathbb{C}$ with $\Re v = 0$) satisfies*

$$\theta/(\theta - u - \Psi(v)) = \Psi_\theta^+(u,v)\Psi_\theta^-(u,v), \qquad u,v \in \mathbb{C} \text{ with } \Re v = \Re u = 0. \tag{2.21}$$

Proof. Let $\ell^{(\theta)}$, $\xi^{(\theta)}$ and $\Xi_\theta = \sum_{n=1}^\infty \delta_{(\ell_n^{(\theta)},\xi_n^{(\theta)})}$ be as in Proposition 2.10. Applying Theorem 2.1 gives

$$(\overline{\tau}_{T_\theta}(X), \overline{X}_{T_\theta}) \overset{d}{=} \int_{B_+} (t,x)\Xi_\theta(\mathrm{d}t,\mathrm{d}x) \quad \text{and}$$
$$(T_\theta - \overline{\tau}_{T_\theta}(X), X_{T_\theta} - \overline{X}_{T_\theta}) \overset{d}{=} \int_{B_-} (t,x)\Xi_\theta(\mathrm{d}t,\mathrm{d}x),$$

where $B_\pm := (0,\infty) \times A_\pm$. Moreover, since the joint law of $(\overline{\tau}_{T_\theta}(X), \overline{X}_{T_\theta})$ and $(T_\theta - \overline{\tau}_{T_\theta}(X), X_{T_\theta} - \overline{X}_{T_\theta})$ equals that of the two integrals in the display above, the vectors are independent because $B_+ \cap B_- = \emptyset$. By Proposition 2.10, the mean measure of Ξ_θ equals $\mu_\theta(\mathrm{d}t,\mathrm{d}x) = t^{-1}\mathrm{e}^{-\theta t}\mathbb{P}(X_t \in \mathrm{d}x)\mathrm{d}t$. Hence

$$\Psi_\theta^\pm(u,v) = \exp\left(\int_{(0,\infty)\times A_\pm} \big(\mathrm{e}^{ut+vx} - 1\big)\frac{\mathrm{e}^{-\theta t}}{t}\mathbb{P}(X_t \in \mathrm{d}x)\mathrm{d}t\right)$$

for all $u,v \in \mathbb{C}$ with $\Re u, \Re v \le 0$, by Campbell's Theorem [64, p. 28]. This representation of $\Psi_\theta^\pm(u,v)$ and (2.20) imply (2.18)–(2.19). The independence and the formula $\mathbb{E}\exp(uT_\theta + vX_{T_\theta}) = \theta/(\theta - u - \Psi(v))$ imply identity (2.21). □

The question of the absolute continuity of the law of $\overline{\chi}_t(X) = (X_t, \overline{X}_t, \overline{\tau}_t(X))$

was the main topic in [28], investigated using excursion theory. Again, Theorem 2.1 provides an easy approach.

Theorem 2.15. *If* 0 *is regular for both half-lines* $(0,\infty)$ *and* $(-\infty,0)$ *and* X_t *is absolutely continuous for each* $t>0$*, then the law of* $\overline{X}_t$ *is absolutely continuous.*

Proof. The assumption on the regularity of zero for both half-lines implies that, with probability one, two increments in the first coordinate in the series in (2.3) have opposite signs, implying $\mathbb{P}(X_t<\overline{X}_t)=1$ and $\mathbb{P}(0<\overline{\tau}_t(X)<t)=1$. Condition on such an event occurring at indices $m\neq n$ and note that the law of (ℓ_n,ℓ_m) is absolutely continuous for any $n\neq m$. Recall that, given the stick-breaking process ℓ, the increment of X is by assumption absolutely continuous. A random vector A is absolutely continuous and independent of B, then $A+B$ has a density, implying the result. □

The proof of Theorem 2.15 is based on analogous arguments in [87], where the authors used a version of Theorem 2.1 for diffuse Lévy processes (their [87, Thm 1]) to establish a result similar to Theorem 2.15. The key difference lies in the fact that the proof of [87, Thm 1] uses in an essential way the following result of Rogozin, whose proof requires fluctuation theory.

Theorem 2.16 (Rogozin)**.** *If a Lévy process* X *has paths of infinite variation, then*

$$\limsup_{t\downarrow 0} X_t/t = -\liminf_{t\downarrow 0} X_t/t = \infty \quad a.s.$$

A circular argument would arise if one attempted to developing fluctuation theory for Lévy processes with diffuse transition laws using [87, Thm 1], because of its reliance on Theorem 2.16. In contrast, Theorem 2.1, applicable to all Lévy processes, has an elementary proof that does not use fluctuation theory. In fact, Theorem 2.1 implies Theorem 2.16. Indeed, $\limsup_{t\downarrow 0} X_t/t=\infty$ if and only if the right-derivative at 0 of the concave majorant of X over $[0,T_\theta]$ is infinite. By Theorem 2.1 and Proposition 2.10, this is equivalent to $\Xi_\theta(\{(t,x)\in(0,\infty)\times\mathbb{R}: x/t>b\})=\infty$ a.s. for all $b\in\mathbb{R}$, which holds if and only if $\int_0^1 t^{-1}\mathbb{P}(X_t>bt)\mathrm{d}t=\infty$ for all $b\in\mathbb{R}$.

We conclude this section with an application of Theorem 2.1, implying a novel factorisation identity.

Corollary 2.17. *Given* $b\in\mathbb{R}$*, define* $X_t^{(b)}:=X_t-bt$*,* $\overline{\tau}_t^{(b)}:=\overline{\tau}_t(X^{(b)})$ *and* $Z_t^{(b)}:=\overline{X}_t^{(b)}+b\overline{\tau}_t^{(b)}$ *and let* $T_\theta\sim\mathrm{Exp}(\theta)$ *be independent of* X*. Then, for real numbers* $a_1>\cdots>a_n$*, the random vectors* $(\overline{\tau}_{T_\theta}^{(a_1)},Z_{T_\theta}^{(a_1)})$*,* $(\overline{\tau}_{T_\theta}^{(a_{i+1})}-\overline{\tau}_{T_\theta}^{(a_i)},Z_{T_\theta}^{(a_{i+1})}-Z_{T_\theta}^{(a_i)})$*,* $(T_\theta-\overline{\tau}_{T_\theta}^{(a_n)},Z_{T_\theta}^{(a_n)}-X_{T_\theta})$*,* $i=1,\ldots,n-1$*, are independent.*

Proof. Since X does not jump at time T_θ a.s., the Lévy process $X^{(a)}$ and its concave majorant on $[0,T_\theta]$, given by $t \mapsto -C_{T_\theta}^{-X^{(a)}}(t) = -C_{T_\theta}^{-X}(t)+at$, share their final value, suprema, and time of suprema. Let $\Xi_\theta \coloneqq \sum_{n=1}^\infty \delta_{(\ell_n^{(\theta)},\xi_n^{(\theta)})}$ be a Poisson process on $A \coloneqq (0,\infty)\times\mathbb{R}$ with mean measure μ_θ. Applying Theorem 2.1 and Proposition 2.10 to the concave majorant of X on $[0,T_\theta]$, we obtain

$$\big(T_\theta, X_{T_\theta}, (\overline{\tau}_t^{(a_i)})_{i\in[\![n]\!]}, (Z_t^{(a_i)} - \overline{\tau}_t^{(a_i)})_{i\in[\![n]\!]}\big) \\ \stackrel{d}{=} \int_A \big(t, x, (\mathbb{1}_{\{x/t>a_i\}}t)_{i\in[\![n]\!]}, (\mathbb{1}_{\{x/t>a_i\}}x)_{i\in[\![n]\!]}\big)\Xi_\theta(\mathrm{d}t,\mathrm{d}x).$$

Thus, by writing the vectors in the statement of the corollary as integrals with respect to Ξ_θ on the disjoint sets $\{(t,x)\in A : x/t > a_1\}$, $\{(t,x)\in A : a_1 \ge x/t > a_2\}$, …, $\{(t,x)\in A : a_{n-1} \ge x/t > a_n\}$ and $\{(t,x)\in A : x/t \le a_n\}$, we obtain their independence. □

§2.6.1 Sticks on exponential intervals are a Poisson point processes

For $n \ge 2$, the Dirichlet law on the simplex $\{(x_1,\dots,x_n)\in(0,1]^n : \sum_{i=1}^n x_i = 1\}$ with parameters $\theta_i > 0$ has a density proportional to $(x_1,\dots,x_n)\mapsto \prod_{i=1}^n x_i^{\theta_i-1}$. D is a Dirichlet random measure on $(0,1]$ if for any $0 = t_0 < t_1 < \ldots < t_n = 1$, the random vector $(D((t_0,t_1]),\dots,D((t_{n-1},t_n]))$ follows the Dirichlet law with parameters $(t_i - t_{i-1})$. Let $(U_n)_{n\in\mathbb{N}}$ and $(V_n)_{n\in\mathbb{N}}$ be independent iid $\mathrm{U}(0,1)$ sequences, independent of a Dirichlet random measure D_0 on $(0,1]$. Elementary calculations imply that $D_1 = (1-V_1)\delta_{U_1} + V_1 D_0 \stackrel{d}{=} D_0$ and hence $D_n = (1-V_n)\delta_{U_n} + V_n D_{n-1} \stackrel{d}{=} D_0$ for all $n\in\mathbb{N}$. Since D_n converges to $D_\infty = \sum_{n\in\mathbb{N}} \ell_n \delta_{U_n}$ in total variation, where $\ell_n = (1-V_n)\prod_{k=1}^{n-1} V_k$ is a uniform stick-breaking process on $[0,1]$, we have $D_0 \stackrel{d}{=} D_\infty$. Moreover, by construction we have $\sum_{n\in\mathbb{N}}(\ell_n^{(\theta)}/T_\theta)\delta_{U_n} \stackrel{d}{=} D_\infty$, where $(\ell_n^{(\theta)})_{n\in\mathbb{N}}$ is a stick-breaking process with an independent time horizon $T_\theta \sim \mathrm{Exp}(\theta)$.

Let G be a gamma subordinator (i.e. G_t has density proportional to $s \mapsto s^{t-1}\mathrm{e}^{-\theta s}$). The jump of G at $t > 0$, $\Delta G_t = G_t - \lim_{s\uparrow t} G_s$, is zero for all but countably many t, making $D' = \sum_{t\in(0,1]}((\Delta G_t)/G_1)\,\delta_t$ a Dirichlet random measure on $(0,1]$, independent of $G_1 \sim \mathrm{Exp}(\theta)$. Indeed, for any $0 = t_0 < t_1 < \ldots < t_n = 1$, the Jacobian change-of-variable formula shows that the vector

$$\big(D'((t_0,t_1]),\dots,D'((t_{n-1},t_n]),G_1\big) = ((G_{t_1}-G_{t_0})/G_1,\dots,(G_{t_n}-G_{t_{n-1}})/G_1, G_1)$$

has the desired law. Thus $(D',G_1) \stackrel{d}{=} (\sum_{n\in\mathbb{N}}(\ell_n^{(\theta)}/T_\theta)\delta_{U_n}, T_\theta)$, implying that the law of the Poisson point process $\sum_{t\in(0,1]}\mathbb{1}_{\Delta G_t>0}\delta_{\Delta G_t}$ coincides with that of the random measure $\sum_{n\in\mathbb{N}}\delta_{\ell_n^{(\theta)}}$.

Chapter 3

The stick-breaking approximation

§3.1 Construction of the stick-breaking approximation

Based on the stick-breaking representation of the vector $\overline{\chi}$ in (2.3), we define the SBA $\overline{\chi}_n^{\mathrm{SB}}$ as follows:

$$\overline{\chi}_n^{\mathrm{SB}} = \sum_{k=1}^{n} (\xi_k, \xi_k^+, \ell_k \mathbb{1}_{\{\xi_k > 0\}}) + (Y_{L_n}, Y_{L_n}^+, L_n \mathbb{1}_{\{Y_{L_n} > 0\}}). \tag{3.1}$$

Since the residual sum $\sum_{k=n+1}^{\infty} \xi_k$ equals Y_{L_n} for any $n \in \mathbb{N}$, the first component of $\overline{\chi}_n^{\mathrm{SB}}$ coincides with that of $\overline{\chi}$, while, as we shall see in Theorem 3.1 below, $Y_{L_n}^+$ and $L_n \mathbb{1}_{\{Y_{L_n} > 0\}}$ reduce the errors of the corresponding partial sums in (3.1). The coupling (X, Y, ℓ) makes it possible to compare $\overline{\chi}$ and $\overline{\chi}_n^{\mathrm{SB}}$ on the same probability space and analyse the strong error $\overline{\chi} - \overline{\chi}_n^{\mathrm{SB}}$.

Denote the distribution of X_t by $F(t, x) = \mathbb{P}(X_t \le x)$, $x \in \mathbb{R}$, for $t > 0$. The following algorithm simulates exactly from the law of the SBA $\overline{\chi}_n^{\mathrm{SB}}$:

Algorithm 1. SB-Alg

Require: $n \in \mathbb{N}$, fixed time horizon $T > 0$

1: Set $L_0 = T$, $\mathcal{X}_0 = (0, 0, 0)$
2: **for** $k = 1, \ldots, n$ **do**
3: Sample $U_k \sim \mathrm{U}(0,1)$ and put $\ell_k = U_k L_{k-1}$ and $L_k = L_{k-1} - \ell_k$
4: Sample $\xi_k \sim F(\ell_k, \cdot)$ and put $\mathcal{X}_k = \mathcal{X}_{k-1} + (\xi_k, \xi_k^+, \ell_k \mathbb{1}_{\{\xi_k > 0\}})$
5: **end for**
6: Sample $\varsigma_n \sim F(L_n, \cdot)$ and **return** $\mathcal{X}_n + (\varsigma_n, \varsigma_n^+, L_n \mathbb{1}_{\{\varsigma_n > 0\}})$

SB-Alg clearly outputs a random vector with the same law as $\overline{\chi}_n^{\mathrm{SB}}$ in (3.1), using a total of $n+1$ sampling steps. Theorem 3.1 and §3.2 below show that $\overline{\chi}_n^{\mathrm{SB}}$ in (3.1) is an increasingly accurate approximation of $\overline{\chi}$ as n grows. Intuitively this is because the sum in the definition of $\overline{\chi}_n^{\mathrm{SB}}$ consists of the first n terms in (2.3) taken in a size-biased order with respect to $(\ell_n)_{n\in\mathbb{N}}$, making the remainder very small. It will become clear from Theorem 3.1 that the last step in SB-Alg reduces the error further. The computational cost of the algorithm is proportional to n if we can sample any increment of X in constant time. We stress that SB-Alg is *not* a version of the random walk approximation on a randomised grid as it does not require the computation of either max or arg max of a discretisation of X. Instead, the approximation for the supremum and its time are obtained by summing non-negative numbers, making SB-Alg numerically very stable. The convergence analysis of SB-Alg relies on the following result, which describes explicitly the law of its error.

Theorem 3.1. *Assume the Lévy process X is not compound Poisson with drift and let (X,Y,ℓ) be the coupling satisfying (2.3). For any $n\in\mathbb{N}$, define the vector of errors of the SBA by*

$$\begin{gathered}\overline{\chi}-\overline{\chi}_n^{\mathrm{SB}}=\big(0,\Delta_n^{\mathrm{SB}},\delta_n^{\mathrm{SB}}\big)=\big(0,\Delta_n-Y_{L_n}^+,\delta_n-L_n\mathbb{1}_{\{Y_{L_n}>0\}}\big),\quad \textit{where}\\ \Delta_n=\overline{X}_T-\sum_{k=1}^n\xi_k^+\qquad\textit{and}\qquad \delta_n=\tau_T-\sum_{k=1}^n\ell_k\mathbb{1}_{\{\xi_k>0\}}.\end{gathered}\tag{3.2}$$

Then, conditionally on L_n,

$$\begin{gathered}(Y_{L_n},\Delta_n,\delta_n)\overset{d}{=}\big(Y_{L_n},\overline{Y}_{L_n},\tau_{L_n}(Y)\big),\quad \textit{and hence}\\ \big(\Delta_n^{\mathrm{SB}},\delta_n^{\mathrm{SB}}\big)\overset{d}{=}\big(\overline{Y}_{L_n}-Y_{L_n}^+,\tau_{L_n}(Y)-L_n\mathbb{1}_{\{Y_{L_n}>0\}}\big).\end{gathered}\tag{3.3}$$

Moreover, the inequalities $0\le\Delta_{n+1}^{\mathrm{SB}}\le\Delta_n^{\mathrm{SB}}\le\Delta_n$, $0\le\delta_n\le L_n$ and $|\delta_n^{\mathrm{SB}}|\le L_n$ hold a.s.

Non-asymptotic (i.e. for fixed n) explicit descriptions of the law of the error, such as (3.3) in Theorem 3.1, are not common among the simulation algorithms for the supremum and related functionals of the path. Since L_n and Y are independent, the representation in (3.3) is easy to work with and provides a cornerstone for the results of §3.2. Note that, by Theorem 3.1, the sequences $(\Delta_n^{\mathrm{SB}})_{n\in\mathbb{N}}$, $(\Delta_n)_{n\in\mathbb{N}}$ and $(\delta_n)_{n\in\mathbb{N}}$ are nonincreasing almost surely and converge to 0. Furthermore, the following observations based on Theorem 3.1 motivate the final step in SB-Alg (i.e. the inclusion of the last summand in the definition in (3.1)): (I) the tail of the error Δ_n^{SB} may be strictly lighter than that of Δ_n (as $\overline{X}_t-X_t^+=\min\{\overline{X}_t,\overline{X}_t-X_t\}$ and

$\overline{X}_t - X_t \overset{d}{=} \sup_{s\in[0,t]}(-X_s)$ for all $t > 0$ [13, Prop. VI.3]); (II) for a large class of Lévy processes, δ_n^{SB} is *asymptotically centred* at 0, i.e. $\mathbb{E}[\delta_n^{\text{SB}}/L_n] \to 0$ as $n \to \infty$, while $\mathbb{E}[\delta_n/L_n]$ converges to a strictly positive constant (see Proposition 3.9 below for details). Theorem 3.1 is proved in §3.4.1.

Since $\mathbb{E}L_n = T2^{-n}$ and L_n is independent of Y, the convergence of SB-Alg is geometric (see also §3.2). Moreover, the error $(\Delta_n^{\text{SB}}, \delta_n^{\text{SB}})$ satisfies the following weak limit.

Corollary 3.2. *If the weak limit $X_t/a(t) \overset{d}{\to} Z_1$ (as $t \to 0$) holds for some (necessarily) α-stable process Z and a positive function a, which is necessarily $1/\alpha$-regularly varying at zero, then*

$$\left(\frac{Y_{L_n}}{a(L_n)}, \frac{\Delta_n}{a(L_n)}, \frac{\delta_n}{L_n}\right) \overset{d}{\to} \left(Z_1, \overline{Z}_1, \tau_1(Z)\right) \quad \textit{as } n \to \infty. \tag{3.4}$$

The assumption in Corollary 3.2 essentially amounts to both tails of the Lévy measure of X being regularly varying at zero with index $-1/\alpha$ (see [60, Thm 2]). This is a rather weak requirement, typically satisfied by Lévy based models in applied probability, which allows an arbitrary modification of the Lévy measure away from zero (see discussion in [60, Sec. 4]). Moreover, the index α is given by (3.30) and the function $a(t)$ is typically of the form $a(t) = C_0 t^{1/\alpha}$ for some constant $C_0 > 0$. The scaling in the limit (3.4) is stochastic; however, since $\mathbb{E}L_n = T2^{-n}$, the rate of decay of the error is clearly geometric. Corollary 3.2 is proved in §3.4.1 by applying Theorem 3.1 to the small-time weak limit of X.

Connections with the literature

In contrast with Theorem 3.1 for the SBA, the laws of the errors of all the other algorithms discussed in Chapter 2 are intractable. The error of the SBA $\overline{\chi}_n^{\text{SB}}$ in (3.1) decays geometrically in law in Corollary 3.2. Analogous weak limits have not been studied for other approximations except for the RWA, where the convergence is polynomial, see [17, 60]. Similarly, the error of the SBA $\overline{\chi}_n^{\text{SB}}$ in (3.1) also converges geometrically in L^p (see Theorem 3.3 below) as opposed to the polynomial decay for the other algorithms (see §3.2.1 below). The error in L^p of the SBA applied to locally Lipschitz and barrier-type functions arising in applications also decays geometrically (see Propositions 3.6 & 3.7 below). Such errors have not been analysed for algorithms other than the RWA, which has polynomial decay (see §3.2.2 for details).

The rate of the decay of the error in L^1 for these functions is crucial, since this error bounds the bias and is directly linked to the computational complexity of MC and MLMC estimates. Indeed, if the mean squared error is to be at most $\epsilon > 0$, the MC algorithm based on the SBA has (near optimal) complexity of order $\mathcal{O}(\epsilon^{-2}\log\epsilon)$. The MLMC scheme based on SB-Alg has (optimal) complexity of order $\mathcal{O}(\epsilon^{-2})$, which is in general neither the case for the RWA [49] nor the WHA [45] (see details in §3.2.4).

The remainder of this chapter is organised as follows. We develop the theory for the SBA as a Monte Carlo algorithm in §3.2. Each result is compared with its analogue (if it exists) for the approximations mentioned in Chapter 2 above. In §3.3 we provide numerical examples illustrating the performance of SB-Alg. The proofs of the results in §3.2 are presented in §3.4.

§3.2 SBA Monte Carlo: theory and applications

The present section describes the geometric convergence of SB-Alg and analyses the Monte Carlo estimation of the functions of interest in applied probability. In §3.2.1 we establish the geometric decay of the error in L^p. In §3.2.2 we show that the error in L^p (and hence the bias) of SB-Alg applied to the aforementioned functions also decays geometrically. In §3.2.3 we study the error of the MC estimator based on SB-Alg for the expected value of those functions via a central limit theorem and provide the corresponding asymptotic and non-asymptotic confidence intervals. §3.2.4 gives the computational complexity of the MC and MLMC estimators based on SB-Alg.

§3.2.1 Geometric decay in L^p of the error of the SBA

In the present subsection we study the decay in L^p of the error $(\Delta_n^{\mathrm{SB}}, \delta_n^{\mathrm{SB}})$ of the SBA $\overline{\chi}_n^{\mathrm{SB}}$ given in (3.2). Let (σ^2, ν, b) be the generating triplet of X associated with the cutoff function $x \mapsto \mathbb{1}_{\{|x|<1\}}$ (see [91, Ch. 2, Def. 8.2]). The existence of the moments of X_T and $\overline{X}_T$, necessary for the following result, can be characterised [91, Thm 25.3] in terms of the integrals

$$I_+^p = \int_{[1,\infty)} x^p \nu(\mathrm{d}x), \qquad I_-^p = \int_{(-\infty,-1]} |x|^p \nu(\mathrm{d}x), \quad p \geq 0. \tag{3.5}$$

Theorem 3.3. *The following results hold for any $p \geq 1$.*

(a) *The inequality $\max\{\mathbb{E}\big[|\delta_n^{\mathrm{SB}}|^p\big], \mathbb{E}[\delta_n^p]\} \leq T^p(1+p)^{-n}$ holds for any $n \in \mathbb{N}$.*

(b) *If $\min\{I_+^p, I_-^p\} < \infty$ (resp. $I_+^p < \infty$), then $\mathbb{E}[(\Delta_n^{\mathrm{SB}})^p]$ (resp. $\mathbb{E}[\Delta_n^p]$) is bounded* *above by $\mathcal{O}(\eta_p^{-n})$ as $n \to \infty$, where η_p lies in the interval $[3/2, 2]$ for any Lévy process X. Both η_p, defined in* (3.31)*, and the constants in $\mathcal{O}(\eta_p^{-n})$ are explicit in the characteristics (σ^2, ν, b) of X (see* (3.33)*).*

By Theorem 3.1, the error Δ_n^{SB} is bounded above by the supremum of the Lévy process over the stochastic interval $[0, L_n]$ with average length equal to $\mathbb{E}L_n = T2^{-n}$. The key step in the proof of Theorem 3.3, given in Lemma 3.12 below, consists of controlling the expectation of the supremum of X over short time intervals (see §3.4.2 below for details).

Since $\eta_2 = 2$ (see definition in (3.31) below), an application of Theorem 3.3(b) for $p \in \{1, 2\}$ yields $\mathbb{E}\Delta_n^{\mathrm{SB}} = \mathcal{O}((3/2)^{-n})$ and $\mathbb{E}\big[\big(\Delta_n^{\mathrm{SB}}\big)^2\big] = \mathcal{O}(2^{-n})$. These two moments are used in the analysis of the MLMC estimator based on SB-Alg (see §3.2.4 below). A further application of Theorem 3.3 yields a geometric bound on the L^p-Wasserstein distance $\mathcal{W}_p(\mathcal{L}(\overline{\chi}), \mathcal{L}(\overline{\chi}_n^{\mathrm{SB}}))$ between the laws $\mathcal{L}(\overline{\chi})$ and $\mathcal{L}(\overline{\chi}_n^{\mathrm{SB}})$ of the corresponding random vectors (see (3.34) below for the definition of the Wasserstein distance and §3.4.2 for the proof of Corollary 3.4).

Corollary 3.4. *If $I_+^p \wedge I_-^p < \infty$ for some $p \geq 1$, then $\mathcal{W}_p(\mathcal{L}(\overline{\chi}), \mathcal{L}(\overline{\chi}_n^{\mathrm{SB}})) = \mathcal{O}(\eta_p^{-n/p})$ as $n \to \infty$. As in Theorem 3.3(b) above, η_p lies in the interval $[3/2, 2]$ and the constant in $\mathcal{O}(\eta_p^{-n/p})$, given in Equation* (3.35)*, is explicit.*

Comparison

The algorithm based on the RWA with time-step T/n outputs a vector $\overline{\chi}_n^{\mathrm{RW}}$, which is comprised of the final state X_T, maximum $\max_{t\in\{0,T/n,\ldots,T\}} X_t$ of X on the grid and first time $\arg\max_{t\in\{0,T/n,\ldots,T\}} X_t$ this maximum is attained on the grid. The L^1 bounds on the error $\Delta_n^{\mathrm{RW}} = \overline{X}_T - \max_{t\in\{0,T/n,\ldots,T\}} X_t$ of the RWA have a long history. Using the weak limit of the error of the RWA, the L^1 bound $\mathbb{E}\Delta_n^{\mathrm{RW}} = \mathcal{O}(n^{-1/2})$ is established for the Brownian motion with drift in [7, 22]. The same bound holds when the jumps of X have finite activity (i.e. $\nu(\mathbb{R}) < \infty$ and $\sigma \neq 0$) [42]. The approach of [42], based on Spitzer's identity, was extended in [30, Thm 5.2.1] to the case without a Brownian component. If X has paths of finite variation, these bounds were further improved via a different methodology in [17]. In particular, by [17, Thm 4.1], we have: $\mathbb{E}\Delta_n^{\mathrm{RW}} = \mathcal{O}(n^{-1/2})$ if X has a Brownian component (i.e. $\sigma \neq 0$), $\mathbb{E}\Delta_n^{\mathrm{RW}} = \mathcal{O}(n^{-1})$ if X has paths of finite variation (i.e. $\int_{(-1,1)} |x|\nu(\mathrm{d}x) < \infty$ and $\sigma = 0$) and $\mathbb{E}\Delta_n^{\mathrm{RW}} = \mathcal{O}(n^{\delta-1/\beta})$ otherwise, for any small $\delta > 0$ and $\beta \in [1, 2]$ defined in (3.14) below.

Bounds for $\mathbb{E}\big[\big(\Delta_n^{\mathrm{RW}}\big)^p\big]$, $p > 0$, analysed in [17, 42], are as follows. By [17, Thm 4.1], for $\alpha \in [0, 2]$ given in (3.30) below, the decay is $\mathcal{O}(n^{-1})$ for $p > \alpha$ and

$\mathcal{O}(n^{\delta-p/\alpha})$ for $0 < p \le \alpha$ and any small $\delta > 0$ (we may take $\delta = 0$ if either $\alpha = 1$ and X is of finite variation or $\alpha = 2$). If X is spectrally negative (i.e. $\nu((0,\infty)) = 0$) and has jumps of finite variation (i.e. $\int_{(-1,0)} |x|\nu(\mathrm{d}x) < \infty$), then for $p > 1$ the decay is of order $\mathcal{O}(n^{-p})$ (resp. $\mathcal{O}(n^{-p/2}\log^p n)$) if $\sigma = 0$ (resp. $\sigma \neq 0$) [42, Lem. 6.5]. Interestingly, as noted in [17, Rem. 4.4], if X has jumps of both signs, then for any $p > 0$, the error of the RWA satisfies $\liminf_{n\to\infty} n\mathbb{E}\big[(\Delta_n^{\mathrm{RW}})^p\big] > 0$. Put differently, the error cannot be of order $\mathrm{o}(n^{-1})$.

Intuitively, the error committed by the RWA is due to the skeleton missing the fluctuations of the process over the interval of length $1/n$ where the process attained its supremum. Since these fluctuations can be substantial in the presence of high jump activity and heavy tails, the decay of the resulting error is polynomial in n. In contrast, the error of the SBA is by Theorem 3.3(b) bounded by $\mathcal{O}(\eta_p^{-n})$ with $\eta_p \in [3/2, 2]$, as it commits the same error as the RWA but over the interval $[0, L_n]$ with average length of $T/2^n$. Numerical results show that the biases of the RWA and the SBA over 2^n and n steps, respectively, are comparable (Figure 3.1 below).

Recall that the WHA, applicable to a parametric class of Lévy processes [70], is given by $(X_{G_n}, \overline{X}_{G_n})$, where G_n is an independent gamma random variable with mean $\mathbb{E}G_n = T$ and variance T^2/n. Since $\overline{X}_{s+t} - \overline{X}_s$ is stochastically dominated by $\overline{X}_t$ and $X_{t+s} - X_s \overset{d}{=} X_t$, the L^p norm of the error is linked to both, the small time behaviour of $t \mapsto (X_t, \overline{X}_t)$ and the deviations of G_n from T. Therefore, the moments of the errors depend on those of $|G_n - T|$ and satisfy $\mathbb{E}[|X_T - X_{G_n}|^p] = \mathcal{O}(n^{-1/q})$ and $\mathbb{E}[|\overline{X}_T - \overline{X}_{G_n}|^p] = \mathcal{O}(n^{-1/q})$ for $p \in \{1, 2\}$, where $q = 4$ if $p = 1$ and X is of infinite variation and $q = 2$ otherwise [45, Prop. 4.5]. These bounds are based on a martingale decomposition of the Lévy process X (see [45, Lem. 4.4]), while our analogous results use the Lévy-Itô decomposition, see Lemma 3.12.

Intuitively, the error in the WHA is due to the censored fluctuations of X over a stochastic interval of length $|G_n - T|$. This is analogous to the error of the SBA over a stochastic interval of length L_n. However, since $\mathbb{E}[|G_n - T|]$ is asymptotically equal to $T\sqrt{2/(n\pi)}$ (by the central limit theorem and [14, Thm 5.4]) and $\mathbb{E}[L_n] = T2^{-n}$, the speed of convergence is polynomial in the WHA and geometric in the SBA.

The first two moments of the error of the JAGA with cost n were analysed in [38, 39], resulting in the bound $\mathcal{O}(n^{-\min\{1,1/\beta_+\}} + n^{1/4-1/\beta_+}\sqrt{\log n})$ if X has no Brownian component (i.e. $\sigma = 0$) and $\mathcal{O}(n^{1/4-\min\{3/4,1/\beta_+\}}\sqrt{\log n})$ otherwise, where β_+, given in (3.15), is *slightly* larger than the Blumenthal-Getoor index $\beta \in [0, 2]$ in (3.14). Intuitively, this error is the result of missing the fluctuations of X between consecutive points on the random grid and the error incurred from approximating the small-jump component with an additional Brownian motion.

§3.2.2 SBA for certain functions: geometric decay of the strong error

Throughout the chapter we consider a measurable function $g : \mathbb{R} \times \mathbb{R}_+ \times [0, T] \to \mathbb{R}$ satisfying $\mathbb{E}|g(\overline{\chi})| < \infty$, where $\mathbb{R}_+ = [0, \infty)$. We focus our attention on the classes of functions that arise in application areas such as financial mathematics [35, 92], risk theory [5, 93] and insurance [37]. More specifically, we study the following three classes of functions: (I) Lipschitz in Proposition 3.5, (II) locally Lipschitz in Proposition 3.6 and (III) barrier-type in Proposition 3.7. These results are a consequence of the representation of the law of the error in Theorem 3.1, bounds from Theorem 3.3 and a tail estimate (without integrability assumptions) for the error Δ_n in Lemma 3.18.

Lipschitz functions of $\overline{\chi}$ arise in applications, for example, in the pricing of hindsight [22, 42, 49, 95] and perpetual American [82] puts under exponential Lévy models. Indeed, for fixed $S_0, K_0 > 0$, these two examples require computing the expectations of $(K_0 - S_0 \mathrm{e}^{X_T - \overline{X}_T})^+$ and $\mathrm{e}^{X_T - \overline{X}_T}$, both of which are bounded and Lipschitz in $(X_T, \overline{X}_T)$ since $\overline{X}_T \geq X_T$. The next result, proved in §3.4.3 below, shows that the convergence of SB-Alg is also geometric for these functions.

Proposition 3.5. *Assume* $|g(x, y, t) - g(x, y', t')| \leq K(|y - y'| + |t - t'|)$ *for some* $K > 0$ *and all* $x \in \mathbb{R}$, $y, y' \in \mathbb{R}_+$, $t, t' \in [0, T]$. *Suppose* $p \geq 1$ *satisfies* $\|g\|_\infty \wedge I_+^p \wedge I_-^p < \infty$, *where* $\|g\|_\infty = \sup\{|g(x, y, t)| : (x, y, t) \in \mathbb{R} \times \mathbb{R}_+ \times [0, T]\}$, *and let* $\eta_p \in [3/2, 2]$ *be as in* (3.31). *Then we have*

$$\mathbb{E}[|g(\overline{\chi}) - g(\overline{\chi}_n^{\mathrm{SB}})|^p] = \mathcal{O}(\eta_p^{-n}) \quad \text{as } n \to \infty.$$

Moreover, the constant in $\mathcal{O}(\eta_p^{-n})$, *given in Equation* (3.38) *below, is explicit in* K, $\|g\|_\infty$ *and the characteristics* (σ^2, ν, b) *of the Lévy process* X.

The pricing of lookback puts, hindsight calls [22, 42, 49] and perpetual American calls [82] involve expectations of continuous functions of $\overline{\chi}$, such as $(S_0 \mathrm{e}^{\overline{X}_T} - K_0)^+$ and $\mathrm{e}^{\overline{X}_T}$, which are only locally Lipschitz. By Proposition 3.6, under appropriate assumptions on large positive jumps, the error of SB-Alg decays geometrically for such functions.

Proposition 3.6. *Assume that* $|g(x, y, t) - g(x, y', t')| \leq K(|y - y'| + |t - t'|)\mathrm{e}^{\max\{y, y'\}}$ *for some* $K > 0$ *and all* $(x, y, y', t, t') \in \mathbb{R} \times \mathbb{R}_+^2 \times [0, T]^2$. *Let* $p \geq 1$ *and* $q > 1$ *satisfy* $\int_{[1,\infty)} \mathrm{e}^{pqx} \nu(\mathrm{d}x) < \infty$ *and let* $\eta_{pq'} \in [3/2, 2]$ *be as in* (3.31), *where* $q' = (1 - 1/q)^{-1}$. *Then we have*

$$\mathbb{E}[|g(\overline{\chi}) - g(\overline{\chi}_n^{\mathrm{SB}})|^p] = \mathcal{O}\big(\eta_{pq'}^{-n/q'}\big) \quad \text{as } n \to \infty.$$

Moreover, the constant in $\mathcal{O}\big(\eta_{pq'}^{-n/q'}\big)$*, given in Equation* (3.40) *below, is explicit in* p, q, K *and the characteristics* (σ^2, ν, b) *of the Lévy process* X*.*

In order to obtain the smallest value $\eta_{pq'}^{-1/q'}$ in Proposition 3.6, one needs to take the largest possible q allowed by the assumptions (see Remark 3.19 below for details). Hence, the rate of decay is determined by the exponential moments of the Lévy measure $\nu|_{[1,\infty)}$. In the context of financial mathematics, it is natural to assume that the returns in the exponential Lévy model have finite variance, i.e. $\mathbb{E}\mathrm{e}^{2X_t} < \infty$. This is equivalent to $\int_{[1,\infty)} \mathrm{e}^{2x}\nu(\mathrm{d}x) < \infty$ [91, Thm 25.3], implying for example $q = 2$ (for $p = 1$) with the bound $\mathcal{O}(2^{-n/2})$.The proof of Proposition 3.6 is in §3.4.3. A numerical example is in §3.3.1.

Barrier-type functions of $\overline{\chi}$, which are discontinuous in the trajectory of the Lévy process, arise in the pricing of contingent convertibles [37], the evaluation of ruin probabilities [65] and as payoffs of barrier options [21, 22, 95]. By Theorem 3.1, the error Δ_n^{SB} in (3.2) of the second coordinate $\overline{X}_T - \Delta_n^{\mathrm{SB}}$ of the SBA $\overline{\chi}_n^{\mathrm{SB}}$ satisfies $0 \le \Delta_n^{\mathrm{SB}} \searrow 0$ a.s. as $n \to \infty$. Hence, the limit $\mathbb{P}(\overline{X}_T - \Delta_n^{\mathrm{SB}} \le x) \searrow \mathbb{P}(\overline{X}_T \le x)$ as $n \to \infty$ holds for any fixed $x > 0$. The rate of convergence in this limit is both crucial for the control of the bias of barrier-type functions and intimately linked to the quality of the right-continuity of the distribution function $x \mapsto \mathbb{P}(\overline{X}_T \le x)$ of $\overline{X}_T$. We will thus need the following assumption.

Assumption (H)**.** *Given* $M, K, \gamma > 0$*, the inequality* $\mathbb{P}(M < \overline{X}_T \le M + x) \le Kx^\gamma$ *holds for all* $x \ge 0$*.*

Proposition 3.7. *Define* $g(\overline{\chi}) = h(X_T)\mathbb{1}_{\{\overline{X}_T \le M\}}$*, where* $h : \mathbb{R} \to \mathbb{R}$ *is bounded and measurable and* $M > 0$*. Let Assumption (H) hold for* M *and some* $K, \gamma > 0$*. Fix any* $p, q \ge 1$ *and let* $\eta_q \in [3/2, 2]$ *be as in* (3.31)*. Then we have*

$$\mathbb{E}[|g(\overline{\chi}) - g(\overline{\chi}_n^{\mathrm{SB}})|^p] = \mathcal{O}\big(\eta_q^{-n\gamma/(\gamma+q)}\big), \quad \text{as } n \to \infty.$$

Moreover, the constant in $\mathcal{O}\big(\eta_q^{-n\gamma/(\gamma+q)}\big)$*, given in Equation* (3.41) *below, is explicit in* K*,* γ*,* p*,* q*,* $\|h\|_\infty$ *and the characteristics* (σ^2, ν, b) *of the Lévy process* X*.*

The proof of Proposition 3.7 is in §3.4.3 below. Minimising $\eta_q^{-\gamma/(\gamma+q)}$ as a function of q is not trivial (see Remark 3.20 below for the optimal choice of q). In the special case when $\gamma = 1$ (i.e. the distribution function of $\overline{X}_T$ is Lipschitz from the right at M) we have: (a) if X has paths of finite variation, then the optimal choice $q = 1$ gives $\eta_1 = 2$ and the bound $\mathcal{O}(2^{-n/2})$; (b) if $\sigma \neq 0$, then the optimal choice $q = 2$ yields the bound $\mathcal{O}(2^{-n/3})$.

The rate of decay in Proposition 3.7 is essentially controlled by the rate of convergence in the Kolmogorov distance of $\overline{X}_T - \Delta_n^{\mathrm{SB}}$ to $\overline{X}_T$. In general, as mentioned above, $\overline{X}_T - \Delta_n^{\mathrm{SB}}$ is known to converge to $\overline{X}_T$ weakly. As the Kolmogorov distance does not metrise the topology of weak convergence (cf. [85, Ex. 1.8.32, p.43]), we require an additional assumption, such as (H), to obtain a rate in Proposition 3.7.

Assumption (H) holds for a wide class of Lévy processes. By the Lebesgue differentiation theorem [33, Thm 6.3.3], the function $x \mapsto \mathbb{P}(\overline{X}_T \le x)$ is differentiable a.e. and Assumption (H) holds for $\gamma = 1$ and Lebesgue almost every M. If the density of $\overline{X}_T$ exists and is bounded around M, then $x \mapsto \mathbb{P}(\overline{X}_T \le x)$ is locally Lipschitz at M, again satisfying Assumption (H) with $\gamma = 1$. This is the case if the density of $\overline{X}_T$ is continuous at M, which holds for stable processes or if $\sigma \neq 0$ [29], and, more generally, if X converges weakly under the zooming-in procedure and $\alpha > 1$ in (3.30), see [17, Lem. 5.7]. Moreover, by [29, Prop. 2] and [13, Sec. VI.4, Thm 19], the density of $\overline{X}_T$ is continuous at M if the ascending ladder height process of X has positive drift (e.g. if X is spectrally negative of infinite variation) or if X is in a certain class of subordinated Brownian motions [72, Prop. 4.5]. However, the continuity of the density of $\overline{X}_T$ is known to fail if X is of bounded variation with no negative jumps and has a Lévy measure with atoms [71, Lem. 2.4]. Furthermore, for any $\gamma \in (0,1)$, the function $x \mapsto \mathbb{P}(\overline{X}_T \le x)$ may be continuous at M but not locally γ-Hölder continuous even if the Lévy measure has no atoms, demonstrating again the necessity of an condition such as Assumption (H) in Proposition 3.7.

We stress that, even if the density is locally bounded at M, it appears to be very difficult to give bounds (based on the Lévy characteristics) on the value it takes at M. This means that, unlike in the case of a (locally)-Lipschitz function $g(\overline{\chi})$, in the context of barrier options we cannot provide non-asymptotic confidence intervals based on Proposition 3.7, cf. §3.2.3 below.

Comparison

The results in [17, 38, 39, 42, 45], discussed in §3.2.1 above, yield bounds in L^p on the error of a Lipschitz function of $(X_T, \overline{X}_T)$. The orders of decay are the same as those reported in §3.2.1 above for the respective approximations. The error of the time of the supremum τ_T, geometrically convergent for the SBA by Theorem 3.3(a) and Proposition 3.5, appears not to have been studied for the other algorithms.

In the case of locally Lipschitz functions, only the decay of the error in L^1 for the RWA seems to have been analysed. Define for any $q > 0$ the integral

$$E_+^q = \int_{[1,\infty)} \mathrm{e}^{qx}\nu(\mathrm{d}x). \tag{3.6}$$

If X has finite activity (i.e. $\nu(\mathbb{R}) < \infty$), then the bias equals $\mathcal{O}(n^{-1/2})$ if $\sigma \neq 0$ and $E_+^q < \infty$ for some $q > 2$ [42, Prop. 5.1] and $\mathrm{o}(n^{-(q-1)/q})$ if $\sigma = 0$ and $E_+^q < \infty$ for some $q > 1$ [42, Rem. 5.3]. In the case $\sigma = 0$ and $\nu(\mathbb{R}) = \infty$, for any $q > 1$ satisfying $E_+^q < \infty$ and any arbitrarily small $\delta > 0$, the bias decays as follows: $\mathcal{O}((n/\log n)^{\delta-(q-1)/q})$ if the process is of finite variation (i.e. $I_0^1 < \infty$), $\mathcal{O}(n^{\delta-(q-1)/q})$ if $\int_{(-1,1)} |x| \log |x| \nu(\mathrm{d}x) < \infty$ and $\mathcal{O}(n^{\delta-(q-1)/(2q)})$ otherwise [42, Thm 6.2]. If the Lévy process X is spectrally negative with jumps of finite variation (i.e. $\nu(\mathbb{R}_+) = 0$ and $\int_{(-1,0)} |x|\nu(\mathrm{d}x) < \infty$) and if $E_+^q < \infty$ for some $q > 1$, the error decays as $\mathcal{O}(n^{-1})$ (resp. $\mathcal{O}(n^{-1/2} \log n)$) if $\sigma = 0$ (resp. $\sigma \neq 0$) [42, Prop. 6.4].

Discontinuous payoffs under variance gamma (VG), normal inverse Gaussian (NIG) and spectrally negative α-stable (with $\alpha > 1$) processes are considered in [49]. Under the assumption that the density of the supremum is bounded around the barrier level in all three models, the errors in L^p of the RWA decay as $\mathcal{O}(n^{\delta-1})$, $\mathcal{O}(n^{\delta-1/2})$ and $\mathcal{O}(n^{\delta-1/\alpha})$ for arbitrarily small $\delta > 0$, respectively [49, Prop. 5.5]. In the case $\nu(\mathbb{R}) < \infty$ and $\sigma \neq 0$, the error decays as $\mathcal{O}(1/\sqrt{n})$, see [43, Prop. 2.2 & Rem. 2.3]. This result was first established in [21] for Brownian motion with drift.

As noted in [17, Sec. 5.3], if X has a jointly continuous density $(t, x) \mapsto \frac{\partial}{\partial x}\mathbb{P}(X_t \leq x)$ bounded for (t, x) away from the origin $(0, 0)$ (e.g. if Orey's condition holds for $\gamma > 1$ [91, Prop. 28.3] or $\sigma > 0$, see also the paragraphs following Proposition 3.7), $\nu(\mathbb{R}) = \infty$ and $\alpha \geq 1$ (defined in (3.30)), then the error in L^p of the RWA for a barrier option decays as $\mathcal{O}(n^{\delta-1/\alpha})$ for any small $\delta > 0$. Moreover, by [17, Lem. 5.8], $\liminf_{n\to\infty} n\mathbb{P}(\overline{X}_T > x \geq \max_{k\in\{1,\dots,n\}} X_{kT/n}) > 0$ if X has jumps of both signs. Put differently, the error in L^p of the RWA for a general barrier option cannot be of order $\mathrm{o}(n^{-1})$. As far as the author knows, such results for the WHA [70] are currently unavailable.

§3.2.3 The central limit theorem (CLT) and the confidence intervals (CIs)

Let $(\overline{\chi}_{n,i}^{\mathrm{SB}})_{i\in\{1,\dots,N\}}$ be the output produced by $N \in \mathbb{N}$ independent runs of SB-Alg using n steps. The *Monte Carlo estimator* $\sum_{i=1}^{N} g(\overline{\chi}_{n,i}^{\mathrm{SB}})/N$ of $\mathbb{E}g(\overline{\chi})$, where $g : \mathbb{R} \times \mathbb{R}_+ \times [0, T] \to \mathbb{R}$ is a measurable function of interest in applied probability (e.g. in one of the classes from §3.2.2 above), has an error

$$\Delta^g_{n,N} = \frac{1}{N}\sum_{i=1}^{N} g(\overline{\chi}^{\mathrm{SB}}_{n,i}) - \mathbb{E}g(\overline{\chi}). \tag{3.7}$$

The aim is to understand the rate of convergence of the error in (3.7) as the number of samples N tends to infinity.

Theorem 3.8 (CLT)**.** *If $\mathbb{P}(\overline{\chi} \in D_g) = 0$, where D_g is the discontinuity set of g, and*

(a) there is a measurable function $G : \mathbb{R} \times \mathbb{R}_+ \times [0, T] \to \mathbb{R}_+$ such that:

- (i) *$|g(x, y, t)| \le G(x, y, t)$ for all $(x, y, t) \in \mathbb{R} \times \mathbb{R}_+ \times [0, T]$,*
- (ii) *for all $x \in \mathbb{R}$, $(y, t) \mapsto G(x, y, t)$ is non-decreasing in both coordinates,*
- (iii) *$\mathbb{E}[G(X_T, \overline{X}_T, T)^2] < \infty$,*

(b) $\mathbb{E}g(\overline{\chi}) = \mathbb{E}g(\overline{\chi}^{\mathrm{SB}}_n) + \mathrm{o}(\eta_g^{-n})$ for some $\eta_g > 1$.

Denote $\mathbb{V}[g(\overline{\chi})] = \mathbb{E}[(g(\overline{\chi}) - \mathbb{E}[g(\overline{\chi})])^2]$ and set $n_N = \lceil \log N / \log(\eta_g^2) \rceil$ for every $N \in \mathbb{N}$, where we denote $\lceil x \rceil = \inf\{n \in \mathbb{N} : n \ge x\}$ for $x \in \mathbb{R}$. Then the following weak convergence holds

$$\sqrt{N}\Delta^g_{n_N,N} \xrightarrow{d} N(0, \mathbb{V}[g(\overline{\chi})]), \qquad \text{as } N \to \infty. \tag{3.8}$$

Theorem 3.8 is not an iid CLT since the bias of the MC estimator forces the increase in the number of steps taken by SB-Alg as the number of samples $N \to \infty$. Its proof (see §3.4.4 below) establishes Lindeberg's condition and then applies the CLT for triangular arrays. The condition $\mathbb{P}(\overline{\chi} \in D_g) = 0$ is satisfied if e.g. the Lebesgue measure of D_g is zero and 0 is regular for X for both half-lines [28, Thm 3]. This assumption is important as it allows us to construct asymptotic confidence intervals for barrier options using the limit in (3.8). Assumption (a) ensures the convergence of $\mathbb{V}[g(\overline{\chi}^{\mathrm{SB}}_n)]$ to $\mathbb{V}[g(\overline{\chi})]$ and might seem restrictive at first sight. However, the function G is very easy to identify (see Remark 3.21 below) in the contexts of Propositions 3.5, 3.6 and 3.7, where Assumption (b) also clearly holds.

Since $|\Delta^g_{n,N}| \le |\mathbb{E}g(\overline{\chi}) - \mathbb{E}g(\overline{\chi}^{\mathrm{SB}}_n)| + |\Delta^g_{n,N} - \mathbb{E}\Delta^g_{n,N}|$, we may construct a confidence interval for the MC estimator $\sum_{i=1}^{N} g(\overline{\chi}^{\mathrm{SB}}_{n,i})/N$ at level $1 - \epsilon \in (0, 1)$ using the implication:

$$\left.\begin{aligned} &|\mathbb{E}g(\overline{\chi}) - \mathbb{E}g(\overline{\chi}^{\mathrm{SB}}_n)| < r_1, \\ &\mathbb{P}(|\Delta^g_{n,N} - \mathbb{E}\Delta^g_{n,N}| < r_2) \ge 1 - \epsilon, \end{aligned}\right\} \implies \mathbb{P}(|\Delta^g_{n,N}| < r_1 + r_2) \ge 1 - \epsilon. \tag{3.9}$$

In (3.9), r_1 may be chosen as a function of the number n of steps in SB-Alg in various ways depending on the properties of g (see Propositions 3.5 and 3.6 of §3.2.2). Note that this requires the explicit dependence of the constant on the model characteristics.

Having fixed n, pick r_2 in (3.9) as a function of ϵ either via concentration inequalities (not relying on Theorem 3.8) or the CLT in Theorem 3.8:
(i) Non-asymptotic CI: since we have $\mathbb{P}\big(|\Delta^g_{n,N} - \mathbb{E}\Delta^g_{n,N}| > r\big) \leq \mathbb{V}[g(\overline{\chi}^{\text{SB}}_n)]/(r^2 N)$ by Chebyshev's inequality, we only need to bound the variance $\mathbb{V}[g(\overline{\chi}^{\text{SB}}_n)]$ (e.g. by the function G in Remark 3.21).
(ii) Asymptotic CI: since $\Delta^g_{n,N} - \mathbb{E}\Delta^g_{n,N} = N^{-1}\sum_{i=1}^N g(\overline{\chi}^i_n) - \mathbb{E}g(\overline{\chi}^{\text{SB}}_n)$, we may use the CLT for fixed n in Remark 3.22 (as in (i), we bound $\mathbb{V}[g(\overline{\chi}^{\text{SB}}_n)]$ by elementary methods).

In the case we do not have access to the constants in the bound on the bias in (3.9) in terms of the model parameters (e.g. barrier options in Proposition 3.7), we apply the CLT result in Theorem 3.8 to the estimator $\Delta^g_{n_N,N}$ directly, to obtain an asymptotic CI. See §3.3.2 below for the numerical examples of asymptotic and non-asymptotic CIs.

§3.2.4 Computational complexity of SB-Alg and the MLMC

Suppose the expected computational cost of drawing a sample from the distribution $F(t,\cdot)$ in SB-Alg is bounded above by a constant that does not depend on $t \in [0,T]$. Then the expected computational cost of a single draw from the law of $\overline{\chi}^{\text{SB}}_n$ via SB-Alg is bounded by $\mathcal{O}(n)$. The CLT in Theorem 3.8 (applicable to (locally) Lipschitz and barrier-type functions, cf. §3.2.3 above) implies that the L^2-norm of the error in (3.7) of the MC estimator can be made smaller than ϵ, i.e. $\mathbb{E}[(\Delta^g_{n,N})^2] \leq \epsilon^2$, at a computational cost of $\mathcal{O}(\epsilon^{-2}\log\epsilon)$ as $\epsilon \to 0$. The cost of the Monte Carlo estimator based on SB-Alg is thus only a log-factor away from the optimal Monte Carlo cost $\mathcal{O}(\epsilon^{-2})$, arising when exact simulation with finite expected running time is possible.

The main aim of MLMC, introduced in [48, 59], is to reduce the computational cost of an MC algorithm for a given level of accuracy. We will apply a general MLMC result [32, Thm 1], stated in Theorem 3.23 for ease of reference. Let $P = g(\overline{\chi})$ and $P_n = g(\overline{\chi}^{\text{SB}}_n)$, $n \in \mathbb{N}$, for any function g that satisfies the assumptions of Theorem 3.8 (see also Remark 3.21 below). Note that the expected computational cost of a single draw in Theorem 3.23 is allowed to grow geometrically in n. Since in the context of the present section sampling P_n has a cost of $\mathcal{O}(n)$, we may choose an arbitrarily small rate $q_3 > 0$ in Theorem 3.23.

A key component of any MLMC scheme is the coupling (P_n, P_{n+1}). In the case of SB-Alg (and the notation therein), this consists of using the same sequence of sticks $(\lambda_k)_{k\in\{1,\dots,n\}}$ and increments $(\xi_k)_{k\in\{1,\dots,n\}}$ in the consecutive levels and setting $\varsigma_n = \xi_{n+1} + \varsigma_{n+1}$. Since

$$\mathbb{V}[P_{n+1} - P_n] \le \mathbb{E}[(P_{n+1} - P_n)^2] \le 2(\mathbb{E}[(P_{n+1} - P)^2] + \mathbb{E}[(P - P_n)^2]), \qquad (3.10)$$

Assumption (b) in Theorem 3.23 follows easily from the bound $\mathbb{E}[(P - P_n)^2] = \mathcal{O}(2^{-nq_2})$ for all functions g of interest (see Propositions 3.5, 3.6 and 3.7 above for the corresponding $q_2 > 0$). These observations imply that the computational complexity of the MLMC estimator in (3.47) is bounded above by $\mathcal{O}(\epsilon^{-2})$ (take $q_3 = q_2/2$ for all choices of g in the propositions above). The implementation of the MLMC estimator based on SB-Alg for a barrier-type function g under the NIG model numerically confirms this bound, see §3.3.3 below.

Comparison

The computational complexity of MC and MLMC procedures based on the SB-Alg is given by $\mathcal{O}(\epsilon^{-2}|\log\epsilon|)$ and $\mathcal{O}(\epsilon^{-2})$, respectively, for a function $g(\overline{\chi})$, which is Lipschitz, locally Lipschitz or barrier-type. This makes SB-Alg robust, as its performance does not depend on the structure of the problem. In particular, minor changes in model parameters will not result in major differences in the computational complexity. We compare this to the extant MC and MLMC algorithms in the literature.

Lipschitz function g. We first review results for Lipschitz functions of $(X_T, \overline{X}_T)$. For the RWA, α as in (3.30) below and a small $\delta > 0$ ($\delta = 0$ if $\alpha \in \{1,2\}$), [17, Thm 4.1] implies that the cost of an MC estimator is $\mathcal{O}(\epsilon^{-2-\max\{1,\alpha+\delta\}})$. In particular, if $\sigma \neq 0$, the complexity of the RWA is $\mathcal{O}(\epsilon^{-4})$ (see also [30, 42, 49]). Their MLMC counterparts, derived following the procedure of [49], together with the bounds in [17, Thm 4.1] and (3.10), have a complexity of $\mathcal{O}(\epsilon^{-2}\log^2(\epsilon))$. Moreover, if the process is spectrally negative without a Brownian component and either an infinite variation stable process [49, Prop. 5.5] or of finite variation [42, Lem. 6.5], then the MLMC estimator for a Lipschitz function of $(X_T, \overline{X}_T)$ has optimal cost $\mathcal{O}(\epsilon^{-2})$. For the WHA, the MC (resp. MLMC) estimator for a Lipschitz function of $(X_T, \overline{X}_T)$ has a complexity of $\mathcal{O}(\epsilon^{-4})$ (resp. $\mathcal{O}(\epsilon^{-3})$) if the process is of finite variation and of $\mathcal{O}(\epsilon^{-6})$ (resp. $\mathcal{O}(\epsilon^{-4})$) otherwise [45, Thm 4.6]. For the JAGA, the complexity of the MC estimator is $\mathcal{O}(\epsilon^{-2-\max\{2,4\beta_+/(4-\beta_+)\}})$ if $\sigma \neq 0$ and $\mathcal{O}(\epsilon^{-2}\max\{\epsilon^{-\max\{1,\beta_+\}}, \epsilon^{-4\beta_+/(4-\beta_+)}\log(1/\epsilon)^{2\beta_+/(4-\beta_+)}\})$ otherwise (see (3.15) for the definition of $\beta_+ \in (0,2]$). The complexity of the corresponding

MLMC estimator is $\mathcal{O}(\epsilon^{-2}\log(1/\epsilon)^{3\cdot\mathbb{1}\{\sigma\neq 0\}})$ if $\beta_+ < 1$, $\mathcal{O}(\epsilon^{-2}|\log\epsilon|^{2+\mathbb{1}\{\sigma\neq 0\}})$ if $\beta_+ = 1$, $\mathcal{O}(\epsilon^{-2-4(1-1/\beta_+)}|\log\epsilon|^{2-2/\beta_+})$ if $\beta_+ \in (1, 4/3]$ and $\sigma \neq 0$, and of order $\mathcal{O}(\epsilon^{-2-8(\beta_+-1)/(4-\beta_+)}|\log\epsilon|^{4(\beta_+-1)/(4-\beta_+)})$ otherwise. In the worst case $\beta_+ = 2$, the MLMC complexity for the JAGA is $\mathcal{O}(\epsilon^{-6})$.

Locally Lipschitz function g. In the case of locally Lipschitz functions, only the MC analysis of the RWA appears to be available in the literature. The error in this case is at best $\mathcal{O}(\epsilon^{-3})$, attained only when the Lévy process is spectrally negative with jumps of finite variation and no Brownian component (i.e. $\nu(\mathbb{R}_+) = 0$, $\int_{(-1,0)}|x|\nu(\mathrm{d}x) < \infty$ and $\sigma = 0$) and the inequality $E_+^q < \infty$ holds for some $q > 1$, see [42, Prop. 6.4] (recall the definition of E_+^q in (3.6) above). If X has a Brownian component (i.e. $\sigma \neq 0$), then the cost is either $\mathcal{O}(\epsilon^{-4})$ if $\nu(\mathbb{R}) < \infty$ and $E_+^q < \infty$ for some $q > 2$ [42, Prop. 6.4] or $\mathcal{O}(\epsilon^{-4}\log^2(\epsilon))$ if X is spectrally negative with jumps of finite variation and $E_+^q < \infty$ for some $q > 1$ [42, Prop. 5.1]. If $\sigma = 0$ and X has infinite activity, then for any arbitrarily small $\delta > 0$, the condition $E_+^q < \infty$ (for some $q > 1$) implies an MC complexity of $\mathcal{O}(\epsilon^{-2-2q/(q-1)-\delta})$. In the last case, the decay may be improved to $\mathcal{O}(\epsilon^{-2-q/(q-1)-\delta}|\log(\epsilon)|)$ (resp. $\mathcal{O}(\epsilon^{-2-q/(q-1)-\delta})$) if $\int_{(-1,1)}|x|\nu(\mathrm{d}x) < \infty$ (resp. $\int_{(-1,1)}|x|\log|x|\nu(\mathrm{d}x) < \infty$) [42, Thm 6.2].

Barrier-type function g. To the best of the author's knowledge, there are no non-parametric MLMC results in the literature for barrier options under the RWA. Recently the MLMC for the RWA under VG, NIG and spectrally negative α-stable (with $\alpha > 1$) processes has been shown in [49] to have the computational cost of $\mathcal{O}(\epsilon^{-2-\delta})$, $\mathcal{O}(\epsilon^{-3-\delta})$ and $\mathcal{O}(\epsilon^{-1-\alpha-\delta})$ for small $\delta > 0$, respectively. We are not aware of any results for WHA, introduced in [70], for barrier options.

§3.3 Numerical examples

The implementation of SB-Alg above can be found in the repository [50] together with a simple algorithm for the simulation of the increments of the VG, NIG and weakly stable processes. This implementation was used in §3.3.1 below.

§3.3.1 Numerical comparison: SBA and RWA

Let $X = (X_t)_{t\geq 0}$ be given by $X_t = B_{Z_t} + bt$, where Z is a subordinator with Lévy measure $\nu_Z(\mathrm{d}x) = \mathbb{1}_{\{x>0\}}\gamma x^{-\alpha-1}\mathrm{e}^{-\lambda x}\mathrm{d}x$ ($\alpha \in [0,1)$, $\gamma, \lambda > 0$) and drift $\sigma_Z \geq 0$, B is a standard Brownian motion and $b \in \mathbb{R}$. The Lévy measure of X equals $\nu(\mathrm{d}x)/\mathrm{d}x = \frac{\gamma}{\sqrt{2\pi}}|x|^{-2\alpha-1}\int_0^\infty s^{-\alpha-3/2}\mathrm{e}^{-\lambda s x^2 - s^{-1}/2}ds$ by [91, Thm 30.1], implying that the Blumenthal-Getoor index of X is $\beta = 2\alpha \in [0,2)$, and its Brownian

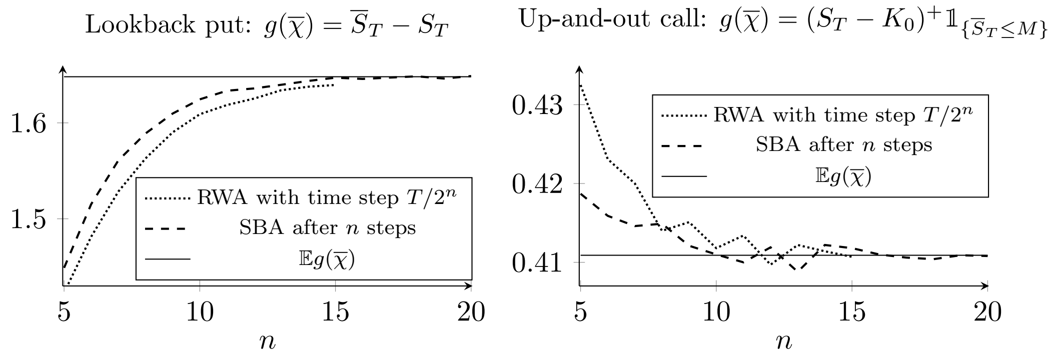

Figure 3.1: We take $\alpha = 0.75$, $\gamma = 0.1$, $\lambda = 4$, $\sigma_Z = 0.05$, $b = -0.05$ and $S_0 = 2$, $K_0 = 3$, $M = 5$, $T = 1$ and $N = 10^7$. The value $\mathbb{E}g(\overline{\chi})$ is obtained by running SB-Alg for $n = 100$ steps and using $N = 10^8$ samples. The RWA is approximately $(2^n/n)$-times slower than the SBA for the same amount of bias, making it infeasible for $n > 15$ as at least $60000 < 2^n$ steps are needed in the time interval $[0, 1]$.

component equals $\sigma^2 = \sigma_Z^2$. Moreover, the increment X_t can be simulated in constant expected computational time for any $t > 0$.

We consider the estimator $\sum_{i=1}^N g(\overline{\chi}_n^i)/N$, where $(\overline{\chi}_n^i)_{i\in\{1,\ldots,N\}}$ are N iid samples produced by running the SB-Alg over n steps. We compare the results with the output of the RWA, based on a time step of size $T/2^n$ and the same number N of iid samples. The function $g(\overline{\chi})$ corresponds to either a lookback put or an up-and-out call under the exponential Lévy model $S = S_0 \exp(X)$. Figure 3.1 shows that the accuracy of the two algorithms is comparable as suggested by Propositions 3.6 and 3.7 above (note $E_\pm^q < \infty$ if and only if $q^2 < 2$, since $\mathbb{E}\big[e^{qX_t}\big] = e^{bt}\mathbb{E}\big[e^{q^2 Z_t/2}\big]$).

§3.3.2 Asymptotic and non-asymptotic CIs

Let X be a Normal Inverse Gaussian process (NIG) with parameters $(b, \kappa, \sigma, \theta)$, i.e. with characteristic function $\mathbb{E}\big[e^{iuX_t}\big] = \exp(t(b + 1/\kappa) - (t/\kappa)\sqrt{1 - 2iu\theta\kappa + \kappa\sigma^2u^2})$, whose Lévy measure is given by

$$\frac{\nu(\mathrm{d}x)}{\mathrm{d}x} = \frac{C}{|x|}e^{Ax}K_1(B|x|), \quad \text{with} \quad A = \frac{\theta}{\sigma^2},\ B = \frac{\sqrt{\theta^2 + \sigma^2/\kappa}}{\sigma^2},\ C = \frac{\sqrt{\theta^2 + 2\sigma^2/\kappa}}{2\pi\sigma\kappa^{3/2}},$$

where K_1 is the modified Bessel function of the second kind, which satisfies

$$K_1(x) = \frac{1}{x} + \mathcal{O}(1), \text{ as } x \to 0, \quad \text{and} \quad K_1(x) = e^{-x}\sqrt{\frac{\pi}{2|x|}}(1 + \mathcal{O}(1/|x|)), \text{ as } x \to \infty.$$

We simulate the increments of the NIG process by [35, Alg. 6.12]. Figure 3.2 presents confidence intervals at level $1 - \epsilon = 99\%$ for the prices of hindsight put and barrier

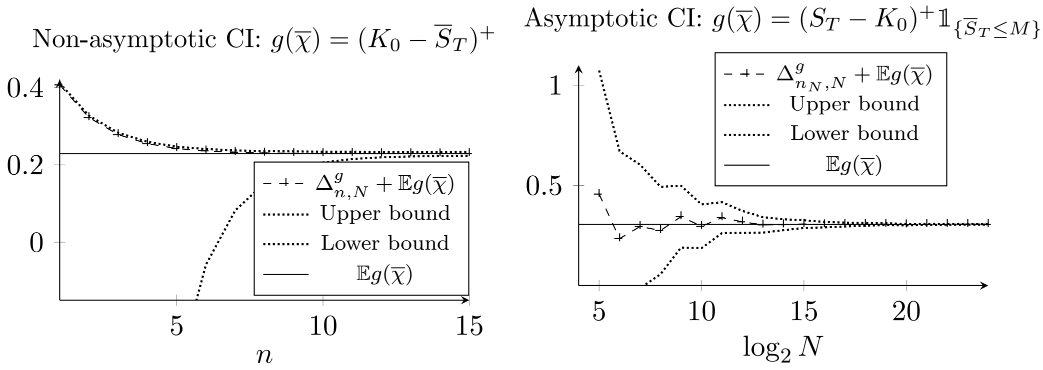

Figure 3.2: The pictures show the point estimation and CIs for the hindsight put (left) and the up-and-out call (right) under the NIG model. NIG parameters: $\sigma = 1$, $\theta = 0.1$, $\kappa = 0.1$ and $b = -0.05$. Option parameters: $S_0 = 2$, $K_0 = 3$, $M = 8$ and $T = 1$. The number of samples in the plot on the left equals $N = 10^7$. The confidence level of $1 - \epsilon = 99\%$ applies to both plots.

up-and-out call under the NIG model $S = S_0 \exp(X)$.

The non-asymptotic CI for the hindsight put is constructed via Chebyshev's inequality as discussed in §3.2.3 above. In particular, note that the payoff of the hindsight put $g : (x, y, t) \mapsto (K_0 - S_0 \mathrm{e}^y)^+$ is non-increasing in y and does not depend on x and t. Since $\overline{X}_T$ dominates the second coordinate $\overline{X}_T - \Delta_n^{\mathrm{SB}}$ of the SBA $\overline{\chi}_n^{\mathrm{SB}}$ in (3.1), we apply $\mathbb{E}g(\overline{\chi}_n^{\mathrm{SB}}) \ge \mathbb{E}g(\overline{\chi})$ and find

$$\left.\begin{array}{l} 0 \le \mathbb{E}g(\overline{\chi}_n^{\mathrm{SB}}) - \mathbb{E}g(\overline{\chi}) < r_1, \\ \mathbb{P}(|\Delta^g_{n,N} - \mathbb{E}\Delta^g_{n,N}| < r_2) \ge 1 - \epsilon \end{array}\right\} \implies \mathbb{P}(-r_1 - r_2 < \Delta^g_{n,N} < r_2) \ge 1 - \epsilon,$$

where $\Delta^g_{n,N}$ is defined in (3.7), reducing the upper bound of the CI to the error r_2, which depends on the bound on g and the number of samples N but not on n.

As explained in §3.2.3 above, if explicit constants in the bounds on the bias are not available in terms of the model parameters, as is the case with an up-and-out call option (see Proposition 3.7 above and remarks following it), we resort to the CLT in Theorem 3.8 above. The plot on the right in Figure 3.2 depicts the asymptotic CI for an up-and-out call as a function of $\log_2 N$, where N is the number of samples used to estimate $\mathbb{E}g(\overline{\chi})$ and the asymptotic variance in (3.8) of Theorem 3.8 is estimated using the sample.

§3.3.3 MLMC for a barrier payoff under NIG

We apply the MLMC algorithm for the SBA to the up-and-out call option in [49, §6.3] (with payoff $g(\overline{\chi}) = (S_T - K_0)^+ \mathbb{1}_{\{\overline{S}_T \le M\}}$, where $S_T = S_0 \exp(X_T)$) under the

NIG model. The top left (resp. right) plot in Figure 3.3 graphs the estimated and theoretically predicted mean (resp. variance) of the difference of two consecutive levels (as a function of n).

It is common practice in MLMC to estimate the bias and level variances (rather than use the theoretical bounds such as those in Theorem 3.23) first and then compute the numbers of samples $(N_k)_{k\in\{1,\ldots,n\}}$ at each level by solving a simple optimisation problem. This often improves the overall performance of the algorithm but requires an initial computational investment. The fact that $(N_k)_{k\in\{1,\ldots,n\}}$ are based on estimates gives rise to some oscillation in their behaviour and, consequently, in that of the computational cost. However, as expected from (3.45), the bottom left plot in Figure 3.3 shows that $(N_k)_{k\in\{1,\ldots,n\}}$ constitute approximately straight lines for various levels of accuracy. The bottom right plot in Figure 3.3 shows that the computational complexity is approximately constant, as expected from the analysis in §3.2.4 above. Moreover, the difference in the complexity between the MC and MLMC is numerically seen to be small. This is not surprising since, as explained in §3.2.4 above, the two differ by a log-factor. The analogous figure for the MLMC based on the RWA for identical model parameters and option is given in [49, Fig. 7].

The computational complexity of MLMC in Figure 3.3 is greater than that of the MC (for $\epsilon > 1/8000$) due to the size of the leading constant. Overall, the performance of both MC and MLMC in this examples is good, with the actual decay rates of the bias and level variances being better than the theoretical bounds by a factor of 2.

§3.4 Proofs and technical results of Chapter 3

§3.4.1 The law of the error and the proof of Theorem 3.1

In the present subsection we will prove Theorem 3.1. We also state and prove Proposition 3.9, which explains why the error δ_n^{SB} of the SBA $\overline{\chi}_n^{\mathrm{SB}}$ is typically smaller than δ_n.

Proof of Theorem 3.1. By the a.s. equality in (2.3), and the definition in (3.2), we obtain

$$(Y_{L_n}, \Delta_n, \delta_n) = \sum_{k=n+1}^{\infty} \big(\xi_k, \xi_k^+, \ell_k \mathbb{1}_{\{\xi_k>0\}}\big).$$

In particular, we have $\delta_n \le \sum_{k=n+1}^{\infty} \ell_k = L_n$ and thus $|\delta_n^{\mathrm{SB}}| \le L_n$.

We next apply (2.2) to conclude that the tail sum in the display above has the required law. Note first that, given L_n, $(\ell_{n+k})_{k\in\mathbb{N}}$ is a stick-breaking process on

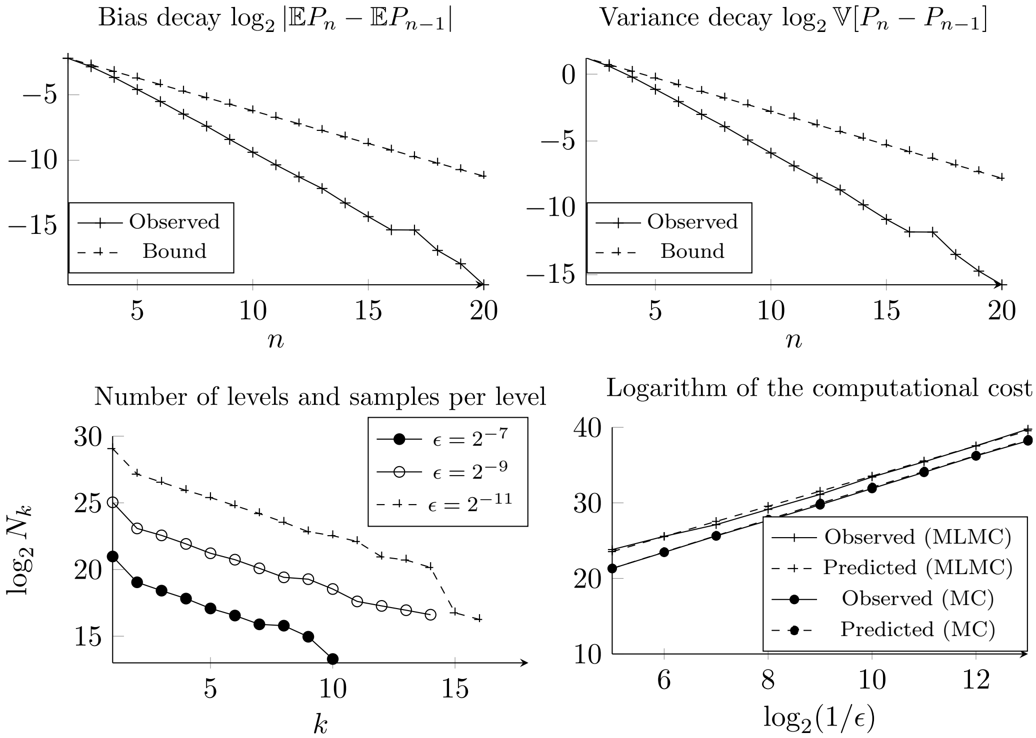

Figure 3.3: The pictures show the level bias decay, level variance decay, samples per level and complexities of MC and MLMC implementations for the up-and-out call $g(\overline{\chi}) = \mathrm{e}^{-rT}(S_T - K)^+ \mathbb{1}_{\{\overline{S}_T < M\}}$ and the NIG process. NIG parameters: $\sigma = 0.1836$, $\theta = -0.1313$, $\kappa = 1.2819$ and $b = 0.1571$ (see [49, Sec. 3] and the reference therein). Option parameters: $S_0 = 100$, $K_0 = 100$, $M = 115$, $T = 1$ and $r = 0.05$. The bounds in the top two graphs are based on Proposition 3.7 (with $\gamma = q = 1$) and synchronous coupling. See §3.2.4 for the computational complexity of MC and MLMC in the bottom right.

the interval $[0, L_n]$. Thus, since Y and ℓ are independent, the law of the sequence $((\ell_{n+k}, Y_{L_{k+n-1}} - Y_{L_{k+n}}))_{k\in\mathbb{N}}$, given L_n, is the same law as that of the right-hand side of (2.2) applied to the interval $[0, L_n]$. Put differently, by (2.2), this sequence has the same law as the sequence of the faces of the concave majorant of the Lévy process Y over the interval $[0, L_n]$ in size-biased order. Hence, identity (4.2) applied to the interval $[0, L_n]$ (instead of $[0, T]$), together with the independence of Y and ℓ, yields the first equality in law in (3.3):

$$(Y_{L_n}, \overline{Y}_{L_n}, \tau_{L_n}(Y)) \overset{d}{=} \sum_{k=n+1}^{\infty} \big(\xi_k, \xi_k^+, \ell_k \mathbb{1}_{\{\xi_k>0\}}\big).$$

The second distributional identity in (3.3) follows from the definition of $(\Delta_n^{\mathrm{SB}}, \delta_n^{\mathrm{SB}})$ as a measurable transformation of $(Y_{L_n}, \Delta_n, \delta_n)$.

For any $n \in \mathbb{N}$, the second identity in (3.3) implies $0 \le \Delta_n^{\mathrm{SB}}$. The definition of Δ_n in (3.2) and the inequality $Y_{L_n}^+ \le \xi_{n+1}^+ + Y_{L_{n+1}}^+$ yield:

$$\Delta_{n+1}^{\mathrm{SB}} = \Delta_{n+1} - Y_{L_{n+1}}^+ = \Delta_n - \xi_{n+1}^+ - Y_{L_{n+1}}^+ \le \Delta_n - Y_{L_n}^+ = \Delta_n^{\mathrm{SB}} \le \Delta_n,$$

completing the proof. □

Proposition 3.9. *Denote $\rho_t = \mathbb{P}(X_t > 0)$ for $t > 0$. The following statements hold.*
(a) *For any $t > 0$, we have $\mathbb{E}\tau_t(X) = \int_0^t \rho_s \mathrm{d}s$.*
(b) *If $t^{-1}\int_0^t \rho_s \mathrm{d}s - \rho_t \to 0$ as $t \searrow 0$, then $\mathbb{E}[\delta_n^{\mathrm{SB}}/L_n] \to 0$ as $n \to \infty$.*
(c) *If $\rho_t \to \rho_0 \in [0, 1]$ as $t \searrow 0$, then* (b) *holds and $\mathbb{E}[\delta_n/L_n] \to \rho_0$ as $n \to \infty$.*
(d) *If $\rho_t = \rho_0 \in [0, 1]$ for all $t \in (0, T]$, then $\mathbb{E}[\delta_n^{\mathrm{SB}}|L_n] = \mathbb{E}[\delta_n|L_n] - L_n\rho_0 = 0$ a.s.*

Remark 3.10. (i) *Note that $\tau_T \in [\tau_T - \delta_n, \tau_T - \delta_n + L_n]$ and, given L_n, SBA $\overline{\chi}_n^{\mathrm{SB}}$ chooses randomly the endpoints of the interval via a Bernoulli random variable with mean $\mathbb{P}(Y_{L_n} > 0|L_n)$.*
(ii) *The assumption in* (d) *holds if e.g. X is a subordinated stable or a symmetric Lévy process. Moreover, it implies that the third coordinate in $\overline{\chi}_n^{\mathrm{SB}}$ is unbiased, since the expectation of its error vanishes: $\mathbb{E}[\delta_n^{\mathrm{SB}}] = 0$. In contrast, $\mathbb{E}[\delta_n] = \rho_0 T/2^n$.*
(iii) *The bias of the third coordinate of $\overline{\chi}_n^{\mathrm{SB}}$, conditional on $L_n = t$, is equal to $\int_0^t \rho_s \mathrm{d}s - t\rho_t$ by* (3.12) *below. This quantity is generally well behaved as $t \to 0$. More specifically, we have $t^{-1}\int_0^t \rho_s \mathrm{d}s - \rho_t \to 0$ as $t \searrow 0$ (thus satisfying the assumption in (b)) if $t \mapsto \rho_t$ is slowly varying at 0 [16, Prop. 1.5.8].*
(iv) *Note that the assumption in* (c) *implies that of* (b). *This assumption, known as Spitzer's condition [13, Thm VI.3.14], is satisfied if for example X converges weakly under the zooming-in procedure [17, Sec. 2.2].*

The result is mostly a consequence of the following formula, valid for any uniform stick-breaking ℓ on $[0, T]$ and positive, measurable function f:

$$\mathbb{E}\sum_{n=1}^{\infty} f(\ell_n) = \mathbb{E}\sum_{n=1}^{\infty} f(L_n) = \int_0^T \frac{f(t)}{t}\mathrm{d}t. \tag{3.11}$$

The formula in (3.11) follows from the fact that $\ell_n \overset{d}{=} L_n$ have the same law as $\exp(-G_n)$ for a gamma random variable G_n with unit scale and shape n.

Proof. Denote $\rho(t) = \rho_t$ for all $t > 0$.
(a) Apply (3.3) to the interval $[0, t]$ and use (3.11) to obtain part (a):

$$\mathbb{E}\tau_t(X) = \sum_{n=1}^{\infty} \mathbb{E}\big[\ell_n \mathbb{1}_{\{\xi_n > 0\}}\big] = \sum_{n=1}^{\infty} \mathbb{E}\big[\ell_n \rho_{\ell_n}\big] = \int_0^T \rho_t \mathrm{d}t.$$

(b) By Theorem 3.1, conditional on L_n, we have $\delta_n^{\mathrm{SB}} \overset{d}{=} \tau_{L_n}(Y) - L_n \mathbb{1}_{\{Y_{L_n} > 0\}}$. Hence, by (a),

$$\mathbb{E}[\delta_n^{\mathrm{SB}} | L_n] = \int_0^{L_n} \rho_s \mathrm{d}s - L_n \rho_{L_n}. \tag{3.12}$$

Since $L_n \to 0$ as $n \to \infty$, the assumption in (b) and (3.12) give $\mathbb{E}[\delta_n^{\mathrm{SB}} | L_n]/L_n \to 0$ a.s. as $n \to \infty$. Using Jensen's inequality and the inequality $|\delta_n^{\mathrm{SB}}/L_n| \le 1$ we deduce that $|\mathbb{E}[\delta_n^{\mathrm{SB}} | L_n]/L_n| \le \mathbb{E}[|\delta_n^{\mathrm{SB}}|/L_n | L_n] \le 1$. Hence, the dominated convergence theorem [62, Thm 1.21] gives $\mathbb{E}[\delta_n^{\mathrm{SB}}/L_n] = \mathbb{E}[\mathbb{E}[\delta_n^{\mathrm{SB}} | L_n]/L_n] \to 0$ as $n \to \infty$.

(c) Since the assumption implies that of (b), the conclusion of (b) holds. Moreover, by (b),

$$\lim_{n\to\infty} \mathbb{E}[\delta_n/L_n | L_n] = \lim_{n\to\infty} \mathbb{E}\big[\delta_n^{\mathrm{SB}}/L_n + \mathbb{1}_{\{Y_{L_n} > 0\}} \big| L_n\big] = \lim_{n\to\infty} \rho_{L_n} = \rho_0 \quad \text{a.s.}$$

The dominated convergence theorem, applied as in the proof of (b), gives the result.

(d) Since $\rho(t) = \rho_0$ for all $t \in [0, T]$, the right-hand side in (3.12) equals 0 a.s., as claimed. Similarly, we have $\mathbb{E}[\delta_n | L_n] = \mathbb{E}[\delta_n^{\mathrm{SB}} + L_n \mathbb{1}_{\{Y_{L_n} > 0\}} | L_n] = L_n \rho_0$ a.s. □

Corollary 3.2. We assume the existence of a function a on the positive reals, such that $(X_{t\delta}/a(\delta))_{t\ge 0}$ converges weakly to some process $(Z_t)_{t\ge 0}$ as $\delta \searrow 0$ in the sense of finite-dimensional distributions. It is known that the limiting process is then self-similar [16, Thm 8.5.2] and thus α-stable and the function a is regularly varying with index $1/\alpha \in [2, \infty)$. Moreover, the convergence extends to the Skorokhod space $\mathcal{D}[0, \infty)$ [61, Cor. VII.3.6]. (For a detailed description of a and the limit criteria see [60, Thm 2].)

Note that $Z^\delta = (Y_{t\delta}/a(\delta))_{t\in[0,1]}$ converges to $Z = (Z_t)_{t\in[0,1]}$ in $\mathcal{D}[0,1]$ and that $\tau_1(Z^\delta) = \tau_\delta(Z)/\delta$. It is well known that the supremum mapping $x \mapsto \sup_{t\in[0,1]} x_t$ and the projection $x \mapsto x_1$ are continuous a.s. with respect to the law of Y. Next, since the time of the maximum of a stable process $(Z_t \vee Z_{t-})_{t\in[0,1]}$ is a.s. unique, then τ_1 is a.s. continuous with respect to the law of Z (see e.g. [62, Lem. 14.12]). Thus, as $\delta \searrow 0$, this yields

$$\chi^\delta = (Y_\delta/a(\delta), \overline{Y}_\delta/a(\delta), \tau_\delta(Y)/\delta) = (Z_1^\delta, \overline{Z}_1^\delta, \tau_1(Z^\delta)) \xrightarrow{d} (Z_1, \overline{Z}_1, \tau_1(Z)) = \chi^0.$$

By the equality in law given in (3.3), we obtain

$$(Y_{L_n}/a(L_n), \Delta_n/a(L_n), \delta_n/L_n) \overset{d}{=} (Y_{L_n}/a(L_n), \overline{Y}_{L_n}/a(L_n), \tau_{L_n}(Y)/L_n). \quad (3.13)$$

Hence, the result will follow if we prove that $\chi^{L_n} \xrightarrow{d} \chi^0$. Recall that the weak convergence is equivalent to $\mathbb{E}f(\chi^\delta) \to \mathbb{E}f(\chi^0)$ as $\delta \searrow 0$ for every bounded and continuous f. Since ℓ and Y are independent and $L_n \to 0$ a.s., conditional on the sequence $(L_n)_{n\in\mathbb{N}}$ we get $\mathbb{E}[f(\chi^{L_n})|L_n] \to \mathbb{E}f(\chi^0)$. The sequence of random variables $(\mathbb{E}[f(\chi^{L_n})|L_n])_{n\in\mathbb{N}}$ is bounded (since f is) and converges to $\mathbb{E}f(\chi^0)$ a.s. Hence, by the dominated convergence theorem, it converges in L^1, implying $\chi^{L_n} \xrightarrow{d} \chi^0$. Hence, the weak limit holds for the left-hand side of (3.13), which yields Corollary 3.2. $\square$

§3.4.2 Convergence in L^p and the proof of Theorem 3.3

Recall that (σ^2, ν, b) is the generating triplet of X associated with the cutoff function $x \mapsto \mathbb{1}_{|x|<1}$ (see [91, Ch. 2, Def. 8.2]). The moments of the Lévy measure ν at infinity are linked with the moments of X_t^+ and $\overline{X}_t$ for any $t > 0$ as follows. By dominating X path-wise with a Lévy process Z equal to X with its jumps in $(-\infty, -1]$ removed and applying [91, Thm 25.3] to Z, we find that, for any $p > 0$, the conditions $I_+^p < \infty$ and $\mathrm{e}_+^p < \infty$ (see (3.5) and (3.6) for definition) imply $\mathbb{E}\left[(X_t^+)^p\right] < \infty$ and $\mathbb{E}\exp(pX_t^+) < \infty$, respectively, for all $t > 0$. Similarly, by applying [91, Thm 25.18] to Z we obtain that $I_+^p < \infty$ and $\mathrm{e}_+^p < \infty$ imply $\mathbb{E}[\overline{X}_t^p] < \infty$ and $\mathbb{E}\exp(p\overline{X}_t) < \infty$, respectively.

Let β be the *Blumenthal-Getoor index* [18], defined as

$$\beta = \inf\{p > 0 : I_0^p < \infty\}, \quad \text{where} \quad I_0^p = \int_{(-1,1)} |x|^p \nu(\mathrm{d}x), \quad \text{for any } p \geq 0, \quad (3.14)$$

and note that $\beta \in [0,2]$ since $I_0^2 < \infty$. Moreover, $I_0^1 < \infty$ if and only if the jumps of X have finite variation, in which case we may define the natural drift

$b_0 = b - \int_{(-1,1)} x\nu(\mathrm{d}x)$. Note that $I_0^p < \infty$ for any $p > \beta$ but I_0^β can be either finite or infinite. If $I_0^\beta = \infty$ we must have $\beta < 2$ and can thus pick $\delta \in (0, 2-\beta)$, satisfying $\beta + \delta < 1$ whenever $\beta < 1$, and define

$$\beta_+ = \beta + \delta \cdot \mathbb{1}_{\{I_0^\beta = \infty\} \cup \{\beta = 0\}} \in [\beta, 2]. \tag{3.15}$$

Note that β_+ is either equal to β or arbitrarily close to it. In either case we have $\beta_+ > 0$ and $I_0^{\beta_+} < \infty$.

The main aim of the present subsection is to prove Propositions 3.5, 3.6 & 3.7 and Theorem 3.3. With this in mind, we first establish three lemmas and a corollary.

Lemma 3.11. *The Lévy measure ν of X satisfies the following for all $\kappa \in (0, 1]$:*

$$\begin{aligned} \overline{\nu}(\kappa) &= \nu(\mathbb{R} \setminus (-\kappa, \kappa)) \le \kappa^{-\beta_+} I_0^{\beta_+} + \overline{\nu}(1), \\ \overline{\sigma}_\kappa^2 &= \int_{(-\kappa,\kappa)} x^2 \nu(\mathrm{d}x) \le \kappa^{2-\beta_+} I_0^{\beta_+}. \end{aligned} \tag{3.16}$$

Moreover the following inequalities hold:

$$\int_{(-1,-\kappa]\cup[\kappa,1)} |x|^p \nu(\mathrm{d}x) \le \kappa^{-(\beta_+-p)^+} I_0^{\beta_+}, \quad \text{for } p \in \mathbb{R}, \tag{3.17}$$

$$\int_{(-\kappa,\kappa)} |x|^p \nu(\mathrm{d}x) \le \kappa^{p-\beta_+} I_0^{\beta_+}, \quad \text{for } p \ge \beta_+. \tag{3.18}$$

Proof. Multiplying the integrands by (I) $(|x|/\kappa)^{\beta_+}$, (II) $(\kappa/|x|)^{2-\beta_+}$, (III) $(|x|/\kappa)^{\beta_+-p}$ if $p \le \beta_+$ or $|x|^{\beta_+-p}$ otherwise and (IV) $(\kappa/|x|)^{p-\beta_+}$, respectively, and extending the integration set to $(-1, 1)$ yields the bounds. □

Recall the definition in (3.5) of I_+^p and I_-^p for $p \ge 0$. Denote $\lceil x \rceil = \inf\{m \in \mathbb{Z} : m \ge x\}$ for any $x \in \mathbb{R}$. Recall that the Stirling numbers of the second kind $\left\{ m \atop k \right\}$ arise in the formula for the moments of a Poisson random variable H with mean $\mu \ge 0$: for any $m \in \mathbb{N}$ we have

$$\mathbb{E}[H^m] = \sum_{k=1}^m \left\{ m \atop k \right\} \mu^k, \qquad \text{where} \qquad \left\{ m \atop k \right\} = \frac{1}{k!} \sum_{i=0}^k (-1)^i \binom{k}{i} (k-i)^m. \tag{3.19}$$

In particular, we have $\left\{ m \atop 0 \right\} = 0$ for all $m \in \mathbb{N}$. Throughout, we will use the following inequality

$$\left(\sum_{k=1}^m x_i \right)^p \le m^{(p-1)^+} \sum_{k=1}^m x_i^p, \quad \text{where } m \in \mathbb{N},\ x_1, \dots, x_m \ge 0 \text{ and } p \ge 0. \tag{3.20}$$

This inequality follows easily from the subadditivity $x \mapsto x^p$ when $p < 1$ and Jensen's inequality when $p \geq 1$.

Lemma 3.12. *For all $t \in [0,T]$ and $p > 0$, the condition $I_+^p < \infty$ implies*

$$\mathbb{E}[\overline{X}_t^p] \leq m_X^p(t) = 4^{(p-1)^+}\big(C_{p,1}t^{p/\beta_+} + C_{p,2}t^{p/2} + C_{p,3}t^p + C_{p,4}t^{\min\{1,p/\beta_+\}}\big), \quad (3.21)$$

where the constants $\{C_{p,i}\}_{i=1}^4$ are given by

$$\begin{aligned} C_{p,1} &= 2^{(p-1)^+}T^{p-p/\beta_+}\big(I_0^{\beta_+}\big)^p + T^{-p/\beta_+}\Big(2^pT^{p/2}\big(I_0^{\beta_+}\big)^{p/2}\cdot \mathbb{1}_{\{p\leq 2\}} \\ &\qquad + 2(p^2/(p-1))^p \exp\big(TI_0^{\beta_+} - p\big)\cdot \mathbb{1}_{\{p>2\}}\Big), \\ C_{p,2} &= |\sigma|^p\Gamma\Big(\frac{p+1}{2}\Big)\frac{2^{p/2}}{\sqrt{\pi}}, \qquad C_{p,3} = 2^{(p-1)^+}\big(b^+\mathbb{1}_{\{I_0^1=\infty\}} + b_0^+\mathbb{1}_{\{I_0^1<\infty\}}\big)^p, \\ C_{p,4} &= T^{(1-p/\beta_+)^+}\big(I_+^p + I'\big)\sum_{k=1}^{\lceil p\rceil}\left\{\begin{matrix}\lceil p\rceil \\ k\end{matrix}\right\}T^{k-1}\big(I' + \nu([1,\infty))\big)^{k-1}, \end{aligned} \quad (3.22)$$

where $I' = \int_{(0,1)} x^{\beta_+}\nu(\mathrm{d}x)$ and $\Gamma(\cdot)$ is the Gamma function. Moreover,

$$\mathbb{E}[\overline{X}_t] \leq |\sigma|\sqrt{\frac{2t}{\pi}} + \begin{cases} (b^+ + I_+^1)t + 2\sqrt{I_0^2 t}, & \beta_+ = 2, \\ (b^+ + I_+^1)t + 2\big(\sqrt{C'} + C'\big)(t/T)^{1/\beta_+}, & \beta_+ \in (1,2), \\ \Big(b_0^+ + \int_{(0,\infty)} x\nu(\mathrm{d}x)\Big)\,t, & \beta_+ \leq 1, \end{cases} \quad (3.23)$$

where $C' = TI_0^{\beta_+}$.

Remark 3.13. *(i) The formula in* (3.23) *essentially follows from [30, Lem. 5.2.2 & Eq. (5.2)] for $\beta_+ \in (1,2]$ and from [42, Prop. 3.4] for $\beta_+ \leq 1$. A new proof of* (3.23) *given below is based on the methodology used to establish a more general inequality in* (3.21)*. Moreover, the dominant powers of t in both bounds* (3.21) *and* (3.23) *coincide in the case $p = 1$ with slightly better constants in* (3.23)*. The estimate in* (3.21) *works for all $p > 0$ and is for the reasons of clarity applied in the proofs that follow even in the case $p = 1$.*
(ii) Note that $C_{p,2} = 0$ if $\sigma = 0$ and, if X is spectrally negative, we have $C_{p,4} = 0$.
(iii) The constants in (3.22) *are well defined even if the assumption $I_+^p < \infty$ fails. The inequality in* (3.21) *holds trivially in this case since $C_{p,4} = \infty$.*

Recall that the *Lévy-Itô decomposition* [91, Thms 19.2 & 19.3] of the Lévy process X with generating triplet (σ^2, ν, b) at a level $\kappa \in (0,1]$ is given by $X_t = b_\kappa t + \sigma B_t + J_t^{1,\kappa} + J_t^{2,\kappa}$ for all $t \geq 0$, where $b_\kappa = b - \int_{(-1,1)\setminus(-\kappa,\kappa)} x\nu(\mathrm{d}x)$ and

$J^{1,\kappa} = (J_t^{1,\kappa})_{t\geq 0}$ (resp. $J^{2,\kappa} = (J_t^{2,\kappa})_{t\geq 0}$) is Lévy with triplet $(0, \nu|_{(-\kappa,\kappa)}, 0)$ (resp. $(0, \nu|_{\mathbb{R}\setminus(-\kappa,\kappa)}, b - b_\kappa)$ - recall that we are using the cutoff function $x \mapsto \mathbb{1}_{|x|\leq 1}$) and $B = (B_t)_{t\geq 0}$ is a standard Brownian motion. Moreover, the processes $B, J^{1,\kappa}, J^{2,\kappa}$ are independent, $J^{1,\kappa}$ is an L^2-bounded martingale with the magnitude of jumps at most κ and $J^{2,\kappa}$ is a compound Poisson process with intensity $\overline{\nu}_\kappa$ (see (3.16) above) and no drift.

Proof. By the discussion above we have $\overline{X}_t \leq b_\kappa^+ t + |\sigma|\overline{B}_t + \overline{J}_t^{1,\kappa} + \overline{J}_t^{2,\kappa}$. Then (3.20) implies

$$\mathbb{E}\big[\overline{X}_t^p\big] \leq 4^{(p-1)^+}\big((b_\kappa^+)^p t^p + |\sigma|^p \mathbb{E}\big[\overline{B}_t^p\big] + \mathbb{E}\big[(\overline{J}_t^{1,\kappa})^p\big] + \mathbb{E}\big[(\overline{J}_t^{2,\kappa})^p\big]\big), \tag{3.24}$$

where $\overline{B}_t \overset{d}{=} |B_t|$ and so $\mathbb{E}\big[\overline{B}_t\big] = t^{p/2}\Gamma\big(\frac{p+1}{2}\big)2^{p/2}/\sqrt{\pi}$ [62, Prop. 13.13], which yields $C_{p,2}$ in all cases. By Lemma 3.11 we have

$$b_\kappa^+ \leq \begin{cases} b_0^+ + \int_{(-\kappa,\kappa)} |x|\nu(\mathrm{d}x) \leq b_0^+ + \kappa^{1-\beta_+} I_0^{\beta_+}, & I_0^1 < \infty \quad \text{(i.e. } \beta_+ \leq 1) \\ b^+ + \kappa^{1-\beta_+} I_0^{\beta_+}, & I_0^1 = \infty \quad \text{(i.e. } \beta_+ > 1). \end{cases}$$

Hence, by (3.20), we obtain

$$\begin{aligned}(b_\kappa^+)^p &\leq \big(\kappa^{1-\beta_+} I_0^{\beta_+} + \mathbb{1}_{\{I_0^1=\infty\}} b^+ + \mathbb{1}_{\{I_0^1<\infty\}} b_0^+\big)^p \\ &\leq 2^{(p-1)^+}\big(\kappa^{p-p\beta_+}\big(I_0^{\beta_+}\big)^p + \mathbb{1}_{\{I_0^1=\infty\}}(b^+)^p + \mathbb{1}_{\{I_0^1<\infty\}}(b_0^+)^p\big).\end{aligned} \tag{3.25}$$

$\overline{J}_t^{2,\kappa}$ is dominated by the sum of the positive jumps of $J^{2,\kappa}$ over the interval $[0,t]$. This sum has the same law as $\sum_{k=1}^{N_t} R_k$ for iid random variables $(R_k)_{k\in\mathbb{N}}$ with law $\nu|_{[\kappa,\infty)}/\nu([\kappa,\infty))$ and an independent Poisson random variable N_t with mean $t\nu([\kappa,\infty))$. Note that since N_t is a non-negative integer, then $N_t^{(p-1)^+ + 1} \leq N_t^{\lceil p\rceil}$. Hence, the independence between $(R_k)_{k\in\mathbb{N}}$ and N_t, the inequality $(\sum_{k=1}^{N_t} R_k)^p \leq N_t^{(p-1)^+}\sum_{k=1}^{N_t} R_k^p$ (which follows from (3.20)) and (3.19) yield

$$\begin{aligned}\mathbb{E}\big[(\overline{J}_t^{2,\kappa})^p\big] &\leq \mathbb{E}\left[\left(\sum_{k=1}^{N_t} R_k\right)^p\right] \leq \mathbb{E}\left[N_t^{(p-1)^+}\sum_{k=1}^{N_t} R_k^p\right] \leq \mathbb{E}[R_1^p]\mathbb{E}\big[N_t^{\lceil p\rceil}\big] \\ &= \left(\int_{[\kappa,\infty)} x^p \frac{\nu(\mathrm{d}x)}{\nu([\kappa,\infty))}\right)\left(\sum_{k=1}^{\lceil p\rceil}\left\{ {\lceil p\rceil \atop k} \right\}(t\nu([\kappa,\infty)))^k\right).\end{aligned}$$

Denote $I' = \int_{(0,1)} x^{\beta_+}\nu(\mathrm{d}x)$. The first inequality in (3.16) and the bound in (3.17) of Lemma 3.11 applied to $\nu|_{(0,\infty)}$ and the facts $\kappa \le 1$ and $t \le T$ yield

$$\begin{aligned}\mathbb{E}\big[(\overline{J}_t^{2,\kappa})^p\big] &\le t\bigg(I_+^p + \int_{[\kappa,1)} x^p \nu(\mathrm{d}x)\bigg) \sum_{k=1}^{\lceil p\rceil} \left\{ {\lceil p\rceil \atop k} \right\} \Big(t\kappa^{-\beta_+} I' + t\nu([1,\infty))\Big)^{k-1} \\ &\le t\kappa^{-(\beta_+-p)^+} \left(I_+^p + I'\right) \sum_{k=1}^{\lceil p\rceil} \left\{ {\lceil p\rceil \atop k} \right\} \Big(t\kappa^{-\beta_+} I' + T\nu([1,\infty))\Big)^{k-1}.\end{aligned} \tag{3.26}$$

Assume $p \le 2$. Jensen's inequality applied to the function $x \mapsto x^{2/p}$ and Doob's martingale inequality [62, Prop. 7.6] applied to $J^{1,\kappa}$ yield

$$\mathbb{E}\big[(\overline{J}_t^{1,\kappa})^p\big] \le \mathbb{E}\big[(\overline{J}_t^{1,\kappa})^2\big]^{p/2} \le 2^p \mathbb{E}\big[(J_t^{1,\kappa})^2\big]^{p/2} = 2^p \left(\overline{\sigma}(\kappa)\right)^p t^{p/2}, \tag{3.27}$$

where $\overline{\sigma}_\kappa$ denotes the positive square root of $\overline{\sigma}_\kappa^2$. Hence (3.24) for $p=1$, the first inequality in (3.26) and the estimate in (3.27) give

$$\mathbb{E}\overline{X}_t \le \bigg(b_\kappa^+ + \int_{[\kappa,1)} x\nu(\mathrm{d}x) + I_+^1\bigg)t + \bigg(|\sigma|\sqrt{\frac{2}{\pi}} + 2\overline{\sigma}(\kappa)\bigg)\sqrt{t}. \tag{3.28}$$

If $\beta_+ = 2$, then taking $\kappa = 1$ in (3.28) yields the first formula in (3.23). If $\beta_+ \le 1$ then $I_0^1 < \infty$. Letting $\kappa \to 0$ in (3.28) we obtain the third formula in (3.23). Set $\kappa = (t/T)^{1/\beta_+}$ and apply Lemma 3.11 to get $t\overline{\sigma}_\kappa^2 \le t^{2/\beta_+} T^{1-2/\beta_+} I_0^{\beta_+}$. Hence $t\int_{[\kappa,1)} x\nu(\mathrm{d}x) \le t^{1/\beta_+} T^{1-1/\beta_+} I_0^{\beta_+}$, and (3.25) & (3.28) yield the second formula in (3.23), completing the proof of (3.23). To prove (3.21) for general $p \in (0,2]$, we again set $\kappa = (t/T)^{1/\beta_+}$ and use the inequalities $t \le T$ and (3.25)–(3.27) as before. More specifically, (I) (3.25), (II) (3.26) and (III) (3.25) & (3.27) establish the values of (I) $C_{p,3}$, (II) $C_{p,4}$ and (III) $C_{p,1}$, respectively. This concludes the proof for the case $p \le 2$.

Assume $p > 2$. The only bound from the case $p \le 2$ above that does not apply in this case is the one on $\mathbb{E}[(\overline{J}_t^{1,\kappa})^p]$. Doob's martingale inequality and the bound $|x|^p \le (p/e)^p \mathrm{e}^{|x|}$ for all $x \in \mathbb{R}$ yield

$$\begin{aligned}\mathbb{E}\big[(\overline{J}_t^{1,\kappa})^p\big] &\le \Big(\frac{p}{p-1}\Big)^p \mathbb{E}\big[|J_t^{1,\kappa}|^p\big] \\ &= \Big(\frac{\kappa p}{p-1}\Big)^p \mathbb{E}\big[(\kappa^{-1}|J_t^{1,\kappa}|)^p\big] \le \Big(\frac{\kappa p^2/e}{p-1}\Big)^p \mathbb{E}\big[\mathrm{e}^{\kappa^{-1}|J_t^{1,\kappa}|}\big].\end{aligned}$$

Note $\mathbb{E}\big[\mathrm{e}^{\kappa^{-1}|J_t^{1,\kappa}|}\big] \le \mathbb{E}\big[\mathrm{e}^{\kappa^{-1}J_t^{1,\kappa}} + \mathrm{e}^{-\kappa^{-1}J_t^{1,\kappa}}\big] = \mathrm{e}^{t\psi_\kappa(\kappa^{-1})} + \mathrm{e}^{t\psi_\kappa(-\kappa^{-1})}$, where ψ_κ is the Lévy-Khintchine exponent of $J_1^{1,\kappa}$, i.e. $\psi_\kappa(u) = \int_{(-\kappa,\kappa)} (\mathrm{e}^{ux} - 1 - ux)\nu(\mathrm{d}x)$ for $u \in \mathbb{R}$. Using the elementary bound $\mathrm{e}^x - 1 - x \le x^2$ for all $|x| \le 1$ and (3.16), we find that $\psi_\kappa(u) \le u^2\overline{\sigma}^2(\kappa) \le u^2\kappa^{2-\beta_+}I_0^{\beta_+}$ for $|u| \le \kappa^{-1}$. By setting $\kappa = (t/T)^{1/\beta_+}$, we obtain

$$\mathbb{E}\big[(\overline{J}_t^{1,\kappa})^p\big] \le 2\Big(\frac{\kappa p^2/e}{p-1}\Big)^p \mathrm{e}^{t\kappa^{-\beta_+}I_0^{\beta_+}} = 2t^{p/\beta_+}T^{-p/\beta_+}\Big(\frac{p^2}{p-1}\Big)^p \mathrm{e}^{TI_0^{\beta_+}-p}. \tag{3.29}$$

As before we obtain (3.21) as follows: (I) (3.25), (II) (3.26) and (III) (3.25) & (3.29) establish the values of (I) $C_{p,3}$, (II) $C_{p,4}$ and (III) $C_{p,1}$, respectively, which completes the proof. □

Recall that β, I_0^1 and β_+ are defined in (3.14) and (3.15) above. To describe the dominant power (as $t \downarrow 0$) in the preceding results, define $\alpha \in [\beta, 2]$ and $\alpha_+ \in [\beta_+, 2]$ by

$$\begin{aligned} \alpha &= 2 \cdot \mathbb{1}_{\{\sigma\neq 0\}} + \mathbb{1}_{\{\sigma=0\}} \begin{cases} 1, & I_0^1 < \infty \text{ and } b_0 \neq 0 \\ \beta, & \text{otherwise,} \end{cases} \qquad \text{and} \\ \alpha_+ &= \alpha + (\beta_+ - \beta) \cdot \mathbb{1}_{\{\alpha=\beta\}}. \end{aligned} \tag{3.30}$$

Note that the index α agrees with the one in [17, Eq. (2.5)] and $\alpha_+ > 0$. Define

$$\eta_p = 1 + \mathbb{1}_{\{p>\alpha\}} + \frac{p}{\alpha_+} \cdot \mathbb{1}_{\{p\le\alpha\}} \in (1, 2], \quad \text{for any} \quad p > 0, \tag{3.31}$$

and note that $\eta_p \ge 3/2$ for $p \ge 1$.

Remark 3.14. *(i) In Theorem 3.3 and Propositions 3.5, 3.6 and 3.7 we assumed that $p \ge 1$ for reasons of clarity. This is not a necessary assumption and the proofs can be made to work with minor modifications for any $p > 0$. However, since $\eta_p \to 1$ as $p \to 0$, the convergence may become arbitrarily slow as $p \to 0$ (to be expected since $x^p \to 1$ as $p \to 0$ for any $x > 0$).*
(ii) The constants $C_{p,2}$ and $C_{p,3}$ in Lemma 3.12 above satisfy the following: (a) *if $\alpha < 2$, then $\sigma = 0$ and hence $C_{p,2} = 0$;* (b) *if $\alpha < 1$, then $I_0^1 < \infty$ and $b_0 = 0$ and hence $C_{p,3} = 0$.*

Corollary 3.15. *Pick $p > 0$, let $\{C_{p,i}\}_{i=1}^4$ be as in Lemma 3.12 and define the constants $C_p(X)$ and $C_p^*(X)$ as follows:*

$$\begin{aligned} C_p(X) &= 4^{(p-1)^+} \cdot \begin{cases} C_{p,1}T^{\frac{p}{\beta_+}-\frac{p}{\alpha_+}} + C_{p,2} + C_{p,3}T^{p-\frac{p}{\alpha_+}} \\ \qquad + C_{p,4}T^{\min\{1,\frac{p}{\beta_+}\}-\frac{p}{\alpha_+}}, & p \le \alpha, \\ C_{p,1}T^{\frac{p}{\beta_+}-1} + C_{p,2}T^{\frac{p}{2}-1} + C_{p,3}T^{p-1} + C_{p,4}, & p > \alpha, \end{cases} \\ C_p^*(X) &= C_p(X) \cdot \mathbb{1}_{\{I_+^p<\infty\}} + C_p(-X) \cdot \mathbb{1}_{\{I_+^p=\infty\}}. \end{aligned} \tag{3.32}$$

Then, if $I_+^p < \infty$ (resp. $\min\{I_+^p, I_-^p\} < \infty$), the inequality

$$\mathbb{E}[\overline{X}_t^p] \le C_p(X) t^{\eta_p - 1} \quad (\textit{resp. } \mathbb{E}[(\overline{X}_t - X_t^+)^p] \le C_p^*(X) t^{\eta_p - 1}).$$

holds for all $t \in [0, T]$.

Proof. Since $\overline{X}_t - X_t^+ = \min\{\overline{X}_t, \overline{X}_t - X_t\}$ is stochastically dominated by both $\overline{X}_t$ and $\overline{(-X)}_t$, then it suffices to prove the result for $\overline{X}_t$. (It is critical here, as seen in the definition of $C_p^*(X)$ in (3.32), that the definition of α is the same for X and $-X$.) Since $t^{q+r} \le t^q T^r$ for $t \in [0, T]$ and $r \ge 0$, then it suffices to show that the exponent of t in each term of (3.21) is at least $\eta_p - 1$. By Remark 3.14(ii), this is trivially the case when $p \le \alpha \le \alpha_+ \le 2$. Recall that α_+ is arbitrarily close (or equal) to α. Hence, in the case $p > \alpha$, we may assume that $p > \alpha_+ \ge \beta_+$ and use Remark 3.14(ii) to obtain the result and conclude the proof. □

Remark 3.16. *If X is spectrally negative (i.e. $\nu(\mathbb{R}_+) = 0$), then $C_{p,4} = 0$ and therefore $\mathbb{E}[\overline{X}_t^p] = \mathcal{O}(t^{p/\max\{1,\alpha_+\}})$ as $t \searrow 0$, implying the rate in [42, Lem. 6.5], which is the best in the literature to date for the spectrally negative case. In certain specific cases, Lemma 3.12 implies a rate better than the one stated in Corollary 3.15. For example, if $\beta < 1$ (thus $\beta_+ < 1$), $\sigma = 0$, $I_+^p < \infty$ and the natural drift satisfies $b_0 < 0$ (thus $\alpha = 1$), then by Lemma 3.12 we have $\mathbb{E}[\overline{X}_t^p] = \mathcal{O}(t^{p/\beta_+})$ if $p \le \beta$, which is sharper than the bound $\mathbb{E}[\overline{X}_t^p] = \mathcal{O}(t^p)$ implied by Corollary 3.15. Analogous improvements can be stated for $\overline{X}_t - X_t^+$, when either ($I_+^p < \infty$ & $b_0 < 0$) or ($I_-^p < \infty$ & $b_0 > 0$). For the sake of presentation, throughout the chapter we work with bounds in Corollary 3.15.*

Lemma 3.17. *Let X, Δ_n and Δ_n^{SB} be as in Theorem 3.1. If $\mathbb{E}[\overline{X}_t^p] \le Ct^q$ (resp. $\mathbb{E}[(\overline{X}_t - X_t^+)^p] \le Ct^q$) for some $C, q, p > 0$ and all $t \in [0, T]$, then*

$$\mathbb{E}\big[\Delta_n^p\big] \le CT^q(1+q)^{-n} \quad (\textit{resp. } \mathbb{E}\big[\big(\Delta_n^{\mathrm{SB}}\big)^p\big] \le CT^q(1+q)^{-n}) \textit{ for all } n \in \mathbb{N}.$$

Proof. By assumption and (3.3) in Theorem 3.1, we have $\mathbb{E}[\Delta_n^p | L_n] = \mathbb{E}[\overline{Y}_{L_n}^p | L_n] \le CL_n^q$ and thus $\mathbb{E}[\Delta_n^p] \le \mathbb{E}[CL_n^q] = CT^q(1+q)^{-n}$. The result for Δ_n^{SB} is analogously proven. □

Proof of Theorem 3.3. (a) By Theorem 3.1, the errors δ_n and $|\delta_n^{\mathrm{SB}}|$ are both bounded by L_n. Since $\mathbb{E}[L_n^p] = T^p(1+p)^n$, the claim follows.
(b) By Corollary 3.15, we may apply Lemma 3.17 to obtain part (b) of the theorem. Indeed,

$$\mathbb{E}[\Delta_n^p] \le C_p(X) T^{\eta_p - 1} \eta_p^{-n} \quad \big(\text{resp. } \mathbb{E}\big[\big(\Delta_n^{\mathrm{SB}}\big)^p\big] \le C_p^*(X) T^{\eta_p - 1} \eta_p^{-n}\big), \tag{3.33}$$

where $C_p(X)$ (resp. $C_p^*(X)$) is as in (3.32) in Corollary 3.15. $\square$

For $p \geq 1$, let $\|\cdot\|_p$ denote the p-norm on $\mathbb{R}^d$. The L^p-Wasserstein distance between distributions μ_x and μ_y on $\mathbb{R}^d$ is defined as

$$\mathcal{W}_p(\mu_x, \mu_y) = \inf_{X \sim \mu_x, Y \sim \mu_y} \mathbb{E}[\|\mathcal{X} - \mathcal{Y}\|_p^p]^{1/p}, \tag{3.34}$$

where the infimum is taken over all couplings of $(\mathcal{X}, \mathcal{Y})$, such that $\mathcal{X}$ and $\mathcal{Y}$ follow the laws μ_x and μ_y, respectively.

Proof of Corollary 3.4. Recall that $\overline{\chi} - \overline{\chi}_n^{\mathrm{SB}} = (0, \Delta_n^{\mathrm{SB}}, \delta_n^{\mathrm{SB}})$ (cf. Theorem 3.1 above). By Theorem 3.3(a), Equation (3.33) and the inequality $1 + p \geq 2 \geq \eta_p$ (since $p \geq 1$), we have

$$\begin{aligned}\mathbb{E}[\|\chi - \overline{\chi}_n^{\mathrm{SB}}\|_p^p] = \mathbb{E}[|\Delta_n^{\mathrm{SB}}|^p + |\delta_n^{\mathrm{SB}}|^p] &\leq C_p^*(X) T^{\eta_p - 1} \eta_p^{-n} + T^p (1+p)^{-n} \\ &\leq (C_p^*(X) T^{\eta_p - 1} + T^p) \eta_p^{-n}.\end{aligned}$$

Since for any coupling of $(\overline{\chi}, \overline{\chi}_n^{\mathrm{SB}})$ we have $\mathcal{W}_p(\mathcal{L}(\overline{\chi}), \mathcal{L}(\overline{\chi}_n^{\mathrm{SB}})) \leq \mathbb{E}[\|\chi - \overline{\chi}_n^{\mathrm{SB}}\|_p^p]^{1/p}$, the L^p-Wasserstein distance is bounded by $C' \eta_p^{-n/p}$, where the constant takes the form

$$C' = (C_p^*(X) T^{\eta_p - 1} + T^p)^{1/p}, \tag{3.35}$$

concluding the proof. $\square$

§3.4.3 Proofs of Propositions 3.5, 3.6 and 3.7

The following result about the tail probabilities of Δ_n (defined in Theorem 3.1) is key in the proofs below.

Lemma 3.18. *Fix $p > 0$ and $T > 0$. Let $C_p(Z)$ be the constant in* (3.32) *of Corollary 3.15 for the Lévy process $Z = X - J^{2,1}$, where $J^{2,1}$ is the compound Poisson process in the Lévy-Itô decomposition of X (see the paragraph preceding the proof of Lemma 3.12). Using the notation $\overline{\nu}(1) = \nu(\mathbb{R} \setminus (-1, 1))$, for any $r, p > 0$, we have*

$$\mathbb{P}\big(\Delta_n \geq r\big) \leq \overline{\nu}(1) T 2^{-n} + r^{-p} C_p(Z) T^{\eta_p - 1} \eta_p^{-n}, \tag{3.36}$$

$$\mathbb{E}\big[\min\{\Delta_n, r\}^p\big] \leq r^p \overline{\nu}(1) T 2^{-n} + C_p(Z) T^{\eta_p - 1} \eta_p^{-n}. \tag{3.37}$$

Proof. Since $\mathbb{P}\big(\Delta_n \geq r\big) = \mathbb{P}\big(\min\{\Delta_n, r\}^p \geq r^p\big) \leq \mathbb{E}\big[\min\{\Delta_n, r\}^p\big]/r^p$ by Markov's inequality, we only need to prove (3.37).

Let Y be as in Theorem 3.1. Pick any $t>0$. Let A be the event on which $J^{2,1}$ does not have a jump on the interval $[0,t]$. Then $\mathbb{P}(A)=\mathrm{e}^{-\overline{\nu}(1)t}\le 1-\overline{\nu}(1)t$ and thus $P(A^c)\le\overline{\nu}(1)t$. By Corollary 3.15 applied to Z we have $\mathbb{E}\big[\overline{Z}_t^p\big]\le C_p(Z)t^{\eta_p-1}$. Since $\overline{X}_t=\overline{Z}_t$ a.s. on the event A we get $\min\{\overline{X}_t,r\}^p\le r^p\cdot\mathbb{1}_{A^c}+\overline{Z}_t^p\cdot\mathbb{1}_A\le r^p\cdot\mathbb{1}_{A^c}+\overline{Z}_t^p$, implying

$$\mathbb{E}\left[\min\{\overline{X}_t,r\}^p\right]\le r^p\overline{\nu}(1)t+C_p(Z)t^{\eta_p-1}.$$

This inequality, Theorem 3.1, $\mathbb{E}[L_n]=T2^{-n}$ and the equality $X\overset{d}{=}Y$ imply (3.37): $\mathbb{E}[\min\{\Delta_n,r\}^p]=\mathbb{E}[\mathbb{E}\big[\min\{\overline{Y}_{L_n},r\}^p|L_n\big]]\le\mathbb{E}[r^p\overline{\nu}(1)L_n+C_p(Z)L_n^{\eta_p-1}]$. □

Proof of Proposition 3.5. Assume first $\|g\|_\infty<\infty$. Since $\min\{a+b,c\}\le\min\{a,c\}+b$ for all $a,b,c\ge 0$, we have

$$|g(x,y,t)-g(x,y',t')|\le\min\{K|y-y'|,\|2g\|_\infty\}+K|t-t'|.$$

Recall that the output of SB-Alg is a copy of $\overline{\chi}_n^{\mathrm{SB}}$. Since, by Theorem 3.1, we a.s. have $0\le\Delta_n^{\mathrm{SB}}\le\Delta_n$ and $|\delta_n^{\mathrm{SB}}|\le L_n$, by (3.20) and (3.37) we obtain

$$\begin{aligned}\mathbb{E}[|g(\overline{\chi})-g(\overline{\chi}_n^{\mathrm{SB}})|^p]&\le 2^{(p-1)^+}\left(\mathbb{E}[K^p\min\{\Delta_n,\|2g\|_\infty/K\}^p]+K^p\mathbb{E}[L_n^p]\right)\\&\le 2^{(p-1)^+}[\|g\|_\infty^p\overline{\nu}(1)T2^{1-n}+K^p(C_p(Z)T^{\eta_p-1}\eta_p^{-n}+T^p(1+p)^{-n})],\end{aligned}$$

where $Z=X-J^{2,1}$. Now assume that $\min\{I_+^p,I_-^p\}<\infty$. Then, again by Theorems 3.1 & 3.3 and Equation (3.33), we obtain

$$\begin{aligned}\mathbb{E}[|g(\overline{\chi})-g(\overline{\chi}_n^{\mathrm{SB}})|^p]&\le 2^{(p-1)^+}K^p(\mathbb{E}[\Delta_n^p]+\mathbb{E}[L_n^p])\\&\le 2^{(p-1)^+}K^p(C_p^*(X)T^{\eta_p-1}\eta_p^{-n}+T^p(1+p)^{-n}).\end{aligned}$$

Since $\eta_p\le 2\le 1+p$ for $p\ge 1$, this yields the result: $\mathbb{E}[|g(\overline{\chi})-g(\overline{\chi}_n^{\mathrm{SB}})|]\le C'\eta_p^{-n}$ for

$$C'=2^{(p-1)^+}\begin{cases}2\|g\|_\infty^p\overline{\nu}(1)T+K^p(C_p(Z)T^{\eta_p-1}+T^p), & \|g\|_\infty<\infty,\\ K^p(C_p^*(X)T^{\eta_p-1}+T^p), & \|g\|_\infty=\infty.\end{cases}\tag{3.38}$$

The proof is thus complete. □

Proposition 3.6. Recall that the second component of $\overline{\chi}_n^{\mathrm{SB}}$ (resp. $\overline{\chi}$) equals $\overline{X}_T-\Delta_n^{\mathrm{SB}}$ (resp. $\overline{X}_T$). Recall from Theorem 3.1 that $|\delta_n^{\mathrm{SB}}|\le L_n$. Since $0\le\Delta_n^{\mathrm{SB}}\le\Delta_n$, the locally Lipschitz property of g implies:

$$|g(\overline{\chi})-g(\overline{\chi}_n^{\mathrm{SB}})|\le K(\Delta_n+L_n)\mathrm{e}^{\overline{X}_T}.$$

From the definition of q' we get $1/q' + 1/q = 1$. Thus Hölder's inequality gives:

$$\mathbb{E}\big[|g(\overline{\chi}) - g(\overline{\chi}_n^{\mathrm{SB}})|^p\big] \le K^p \mathbb{E}\big[(\Delta_n + L_n)^{pq'}\big]^{\frac{1}{q'}} \mathbb{E}\left[\mathrm{e}^{pq\overline{X}_T}\right]^{\frac{1}{q}}, \tag{3.39}$$

where the second expectation on the right-hand side of (3.39) is finite by assumption $\mathrm{e}_+^{pq} < \infty$ and the argument in the first paragraph of §3.4.2 above.

We now estimate both expectations on the right-hand side of (3.39). Note that $I_+^r < \infty$ for all $r > 0$ as $\mathrm{e}_+^{pq} < \infty$. By (3.20), we have $\mathbb{E}\big[(\Delta_n + L_n)^{pq'}\big] \le 2^{(pq'-1)^+}\mathbb{E}\big[\Delta_n^{pq'} + L_n^{pq'}\big]$. Hence Theorem 3.3, (3.33) and the inequality $(x+y)^{1/q'} \le x^{1/q'} + y^{1/q'}$ for $x, y \ge 0$ imply

$$\begin{aligned}\mathbb{E}\big[(\Delta_n + L_n)^{pq'}\big]^{1/q'} &\le 2^{(p-1/q')^+}\big(C_{pq'}(X)T^{\eta_{pq'}-1}\eta_{pq'}^{-n} + T^{pq'}(1+pq')^{-n}\big)^{1/q'} \\ &\le 2^{(p-1/q')^+}\big(C_{pq'}(X)^{1/q'}T^{(\eta_{pq'}-1)/q'}\eta_{pq'}^{-n/q'} + T^p(1+pq')^{-n/q'}\big).\end{aligned}$$

It remains to obtain an explicit bound for the expectation $\mathbb{E}[\exp(pq\overline{X}_T)]$. Let $\Psi(u) = \log \mathbb{E}[\mathrm{e}^{uX_1}]$ for $u \ge 0$ and note that $(\exp(pX_t - t\Psi(p)))_{t\ge 0}$ is a positive martingale. Moreover, we have $\zeta = \sup_{t\in[0,T]}(pX_t - \Psi(p)t) \ge p\overline{X}_T - T\Psi(p)^+$. Thus, Doob's martingale inequality yields

$$\begin{aligned}\mathbb{E}\left[\mathrm{e}^{pq\overline{X}_T}\right] \le \mathrm{e}^{qT\Psi(p)^+}\mathbb{E}\big[\mathrm{e}^{q\zeta}\big] &\le \Big(\frac{q}{q-1}\Big)^q \mathrm{e}^{qT\Psi(p)^+}\mathbb{E}\big[\mathrm{e}^{pqX_T - qT\Psi(p)}\big] \\ &= \Big(\frac{q}{q-1}\Big)^q \mathrm{e}^{qT(-\Psi(p))^+ + T\Psi(pq)}.\end{aligned}$$

Therefore, using (3.39) and the inequalities $\eta_{pq'} \le 2 \le 1 + pq'$ (as $pq' \ge 1$), we obtain the bound $\mathbb{E}\big[|g(\overline{\chi}) - g(\overline{\chi}_n^{\mathrm{SB}})|^p\big] \le C'\eta_{pq'}^{-n/q'}$, where

$$C' = \big(C_{pq'}(X)^{1/q'}T^{(\eta_{pq'}-1)/q'} + T^p\big)\frac{2^{(p-1/q')^+}K^p q}{q-1}\mathrm{e}^{T(-\Psi(p))^+ + (T/q)\Psi(pq)}, \tag{3.40}$$

and the constant $C_{pq'}(X)$ is defined in (3.32) and $\Psi(u) = \log \mathbb{E}[\mathrm{e}^{uX_1}]$. □

Remark 3.19. *The rate $\eta_{pq'}^{-1/q'}$ in the bound of Proposition 3.6 is smallest (as a function of q) for the largest q satisfying the exponential moment condition in Proposition 3.6. Indeed, let $r = pq'$ and note that, since p is fixed, minimising $\eta_{pq'}^{-1/q'}$ in q is equivalent to maximising $\eta_r^{1/r}$ in r. By (3.31), the function $r \mapsto \eta_r^{1/r}$ is decreasing and hence takes its maximal value at the smallest possible r (i.e. largest possible q).*

Proposition 3.7. Recall from Theorem 3.1 that $0 \le \Delta_n^{\mathrm{SB}} \le \Delta_n$. Let $\epsilon_n = \eta_q^{-n/(\gamma+q)}$

and note

$$\begin{aligned}
\mathbb{E}\left[\frac{|h(X_t)|^p}{\|h\|_\infty^p}\left|\mathbb{1}_{\overline{X}_T-\Delta_n^{\mathrm{SB}}\le x}-\mathbb{1}_{\overline{X}_T\le x}\right|^p\right] &\le \mathbb{P}(\overline{X}_T-\Delta_n^{\mathrm{SB}}\le x<\overline{X}_T)\\
&\le \mathbb{P}(\overline{X}_T-\Delta_n\le x<\overline{X}_T)\\
&= \mathbb{P}(\overline{X}_T-\Delta_n\le x<\overline{X}_T-\epsilon_n)\\
&\qquad+\mathbb{P}(\overline{X}_T-\Delta_n\le x<\overline{X}_T\le x+\epsilon_n)\\
&\le \mathbb{P}(\epsilon_n<\Delta_n)+\mathbb{P}(x<\overline{X}_T\le x+\epsilon_n).
\end{aligned}$$

By (3.36) in Lemma 3.18 we have

$$\mathbb{P}(\epsilon_n<\Delta_n)\le\overline{\nu}(1)T2^{-n}+\epsilon_n^{-q}C_q(Z)T^{\eta_q-1}\eta_q^{-n}=\overline{\nu}(1)T2^{-n}+C_q(Z)T^{\eta_q-1}\eta_q^{-n\gamma/(\gamma+q)}.$$

The assumed Hölder continuity of the distribution function of $\overline{X}_T$ in Assumption (H) implies that $\mathbb{P}(x<\overline{X}_T\le x+\epsilon_n)\le K\epsilon_n^\gamma$. Given the formula for $C_q(Z)$ in (3.32), the explicit constant

$$C'=\|h\|_\infty^p(\overline{\nu}(1)T+C_q(Z)T^{\eta_q-1}+K), \tag{3.41}$$

satisfies $\mathbb{E}[|g(\overline{\chi})-g(\overline{\chi}_n^{\mathrm{SB}})|^p]\le C'\eta_q^{-n\gamma/(\gamma+q)}$. □

Remark 3.20. *Minimising the rate $\eta_q^{-\gamma/(\gamma+q)}$ as a function of q in Proposition 3.7 is somewhat involved. On the interval (α_+,∞), the rate $q\mapsto\eta_q^{-\gamma/(\gamma+q)}=2^{-\gamma/(\gamma+q)}$ is strictly increasing, so the optimal q always lies in $(0,\alpha_+]$. On the interval $(0,\alpha_+]$ the problem is equivalent to maximising the map $r\mapsto\mathrm{e}^{f(r)}=\eta_q^{\gamma/(\gamma+q)}$ on the interval $(0,1]$, where $r=\frac{q}{\alpha_+}\in(0,1]$ and $f:x\mapsto\log(1+x)/\big(1+\frac{\alpha_+}{\gamma}x\big)$. Since*

$$\frac{\gamma}{\alpha_+}\Big(1+\frac{\alpha_+}{\gamma}x\Big)^2\frac{d}{\mathrm{d}x}f(x)=\frac{\frac{\gamma}{\alpha_+}-1}{1+x}-(\log(1+x)-1),$$

the critical point of f, obtained by solving for $s=\log(1+x)-1$ in $s\mathrm{e}^s=\mathrm{e}^{-1}(\frac{\gamma}{\alpha_+}-1)$, is given by $r_0=\mathrm{e}^{W(\mathrm{e}^{-1}(\gamma/\alpha_+-1))+1}-1$, where W is the Lambert W function, defined as the inverse of $x\mapsto x\mathrm{e}^x$. Since f is increasing on $[0,r_0]$ and decreasing on (r_0,∞), then $r=\min\{r_0,1\}$ maximises $f|_{(0,1]}$, implying that the optimal q equals

$$q=\alpha_+\min\Big\{1,\mathrm{e}^{W(\mathrm{e}^{-1}(\gamma/\alpha_+-1))+1}-1\Big\}.$$

In particular, the choice $q=\alpha_+$ is optimal if and only if $\gamma/\alpha_+\ge2\log(2)-1=0.38629\ldots$, and leads to the bound $\mathcal{O}(2^{-n/(1+\alpha_+/\gamma)})$. Hence, if $\gamma=1$, the best bound in Proposition 3.7 is $\mathcal{O}(2^{-n/(1+\alpha_+)})$.

§3.4.4 The proof of the central limit theorem

Proof of Theorem 3.8. Recall $n_N = \lceil \log N / \log(\eta_g^2) \rceil$ and note that $1 \geq \sqrt{N}\eta_g^{-n_N} \geq \eta_g^{-1}$. Hence Assumption (b) yields

$$\sqrt{N}\mathbb{E}\Delta^g_{n_N,N} \to 0 \qquad \text{as } N \to \infty. \tag{3.42}$$

The coupling in (2.3), used in Theorem 3.1, implies that for all $n \in \mathbb{N}$ the following relations between the vectors $\overline{\chi}$ and the SBA $\overline{\chi}_n^{\mathrm{SB}}$ in (3.1) hold a.s.: $Y_T = X_T$, $\overline{X}_T - \Delta_n^{\mathrm{SB}} \leq \overline{X}_T$ and $\tau_T - \delta_n^{\mathrm{SB}} \leq T$. Hence parts (i) and (ii) of Assumption (a) imply that $g(\overline{\chi}_n^{\mathrm{SB}})$ and $g(\overline{\chi}_n^{\mathrm{SB}})^2$ are dominated by $\zeta = G(X_T, \overline{X}_T, T)$ and ζ^2, respectively. Since ζ and ζ^2 are integrable by assumption, the dominated convergence theorem yields, as $n \to \infty$,

$$\mathbb{V}[g(\overline{\chi}_n^{\mathrm{SB}})] = \mathbb{E}[g(\overline{\chi}_n^{\mathrm{SB}})^2] - [\mathbb{E}g(\overline{\chi}_n^{\mathrm{SB}})]^2 \to \mathbb{E}[g(\overline{\chi})^2] - [\mathbb{E}g(\overline{\chi})]^2 = \mathbb{V}[g(\overline{\chi})]. \tag{3.43}$$

Recall that $(\overline{\chi}_n^i)_{i\in\{1,\ldots,N\}}$ is the output produced by N independent runs of SB-Alg using n steps. Define the normalised centred random variables

$$\zeta_{i,N} = \left(g\left(\overline{\chi}^i_{n_N}\right) - \mathbb{E}g\left(\overline{\chi}^i_{n_N}\right)\right) / \sqrt{N\mathbb{V}[g(\overline{\chi})]}, \quad \text{where } i \in \{1, \ldots, N\}.$$

Hence (3.43) implies $\sum_{i=1}^N \mathbb{E}\zeta_{i,N}^2 = \mathbb{V}[g(\overline{\chi})]^{-1}(1/N)\sum_{i=1}^N \mathbb{V}[g(\overline{\chi}^i_{n_N})] \to 1$ as $N \to \infty$. Moreover, we have

$$\sum_{i=1}^N \zeta_{i,N} = \sqrt{N/\mathbb{V}[g(\overline{\chi})]}\Delta^g_{n_N,N} + \mathrm{o}(1) \qquad \text{as } N \to \infty,$$

where $\mathrm{o}(1)$ is a deterministic sequence, proportional to the one in (3.42). Hence, (3.8) holds if and only if $\sum_{i=1}^N \zeta_{i,N} \xrightarrow{d} N(0,1)$ as $N \to \infty$.

To conclude the proof, we shall use Lindeberg's CLT [62, Thm 5.12], for which it remains to prove that Lindeberg's condition holds, i.e. $\sum_{i=1}^N \mathbb{E}[\zeta_{i,N}^2 \mathbb{1}_{\{\zeta_{i,N}>r\}}] \to 0$ as $N \to \infty$ for all $r > 0$. By the coupling from the second paragraph of this proof, we find $|g(\overline{\chi}_{n,i}^{\mathrm{SB}})| \leq |\zeta_i|$ for all $i \in \{1, \ldots, N\}$ and $n \in \mathbb{N}$, where $(\zeta_i)_{i\in\{1,\ldots,N\}}$ are iid with the law equal to $G(X_T, \overline{X}_T, T)$. Crucially, ζ_i does not depend on the number of steps n_N in the SB-Alg. Moreover, note that iid random variables $\xi_i = (|\zeta_i| + \mathbb{E}|\zeta_i|)$

satisfy $\mathbb{E}\xi_i^2 < \infty$ and $|\zeta_{i,N}| \le \xi_i/\sqrt{N\mathbb{V}[g(\overline{\chi})]}$ for any $i \in \{1,\dots,N\}$. Hence we find

$$\begin{aligned}\mathbb{V}[g(\overline{\chi})]\sum_{i=1}^{N}\mathbb{E}[\zeta_{i,N}^2\mathbb{1}_{\{\zeta_{i,N}>r\}}] &\le \sum_{i=1}^{N}\frac{1}{N}\mathbb{E}\left[\xi_i^2\mathbb{1}_{\{\xi_i>rN\mathbb{V}(g(\overline{\chi}))\}}\right]\\ &= \mathbb{E}\left[\xi_1^2\mathbb{1}_{\{\xi_1>rN\mathbb{V}(g(\overline{\chi}))\}}\right] \to 0\end{aligned}$$

as $N \to \infty$, implying Lindeberg's condition and the theorem. □

Remark 3.21. *Identifying the appropriate G in Theorem 3.8 is usually simple. For instance, the following choices of G can be made in the contexts of interest.*

(a) Let g be Lipschitz (as in Proposition 3.5). Then we can take

  (i) $G(x,y,t) = \|g\|_\infty$, if $\|g\|_\infty < \infty$;

  (ii) $G(x,y,t) = |g(x,y,t)| + 2K(y+t)$, if $I_+^2 < \infty$.

(b) Let g be locally Lipschitz with the Lipschitz constant exponentially increasing as in Proposition 3.6. Then we can take

  (i) $G(x,y,t) = K\mathrm{e}^y$, if $g(x,y,t) \le K\mathrm{e}^y$ and $E_+^2 < \infty$ (lookback and hindsight options fall in this category);

  (ii) $G(x,y,t) = |g(x,y,t)| + 2K(y+t)\mathrm{e}^y$ if $E_+^{2q} < \infty$ for some $q > 1$.

(c) If g is a barrier option (as in Proposition 3.7), then take $G(x,y,t) = \|g\|_\infty$.

Remark 3.22. *If we are prepared to centre, it is possible to apply the standard iid CLT to the estimator based on SB-Alg. Indeed, for fixed n, assuming $\mathbb{V}[P_n] < \infty$ where $P_n = g(\overline{\chi}_n^{\mathrm{SB}})$, the classical CLT yields*

$$\frac{1}{\sqrt{N\mathbb{V}[P_n]}}\sum_{i=1}^{N}(P_n^i - \mathbb{E}P_n) \xrightarrow{d} N(0,1) \qquad \text{as } N \to \infty.$$

In contrast, the gist of Theorem 3.8 is that one need not centre the sample with a function of n, which itself depends on the sample.

§3.5 MC and MLMC estimators

§3.5.1 Monte Carlo estimator

Consider random variables $P, P_1, P_2, \dots$ in L^2. Let $\{P_k^i\}_{k,i\in\mathbb{N}}$ be independent with $P_k^i \overset{d}{=} P_k$ for $k, i \in \mathbb{N}$. Suppose $|\mathbb{E}P - \mathbb{E}P_k| \le B(k)$ for all $k \in \mathbb{N}$ and assume $C(n)$

bounds the expected computational cost of simulating a single value of P_n. Pick arbitrary $\epsilon > 0$ and define $m = \inf\{k \in \mathbb{N} : B(k) < \epsilon/\sqrt{2}\}$, $N = \lceil 2\mathbb{V}[P_m]/\epsilon^2 \rceil$. Then the Monte Carlo estimator

$$\hat{P} = \frac{1}{N}\sum_{i=1}^{N} P_m^i \quad \text{of } \mathbb{E}P \text{ is } L^2\text{-accurate at level } \epsilon, \quad \text{i.e. } \mathbb{E}\big[(\hat{P} - \mathbb{E}P)^2\big]^{1/2} < \epsilon,$$

since $\mathbb{E}[(\hat{P} - \mathbb{E}P)^2] = \mathbb{V}[\hat{P}] + (\mathbb{E}P_m - \mathbb{E}P)^2$ and $\mathbb{V}[\hat{P}] < \epsilon^2/2$ (by the definition of N), while $(\mathbb{E}P_m - \mathbb{E}P)^2 < \epsilon^2/2$ (by the definition of m). Thus, if the bound $B(m)$ on the bias is asymptotically sharp, the formulae for $m, N \in \mathbb{N}$ above result in the computational complexity given by $\mathcal{C}_{\mathrm{MC}}(\epsilon) = NC(m) = \lceil 2\mathbb{V}[P_m]/\epsilon^2 \rceil\, C(m)$. Although in practice one does not have access to the variance $\mathbb{V}[P_m]$, it is typically close to $\mathbb{V}[P]$ (which often has an *a priori* bound) or can be estimated via simulation.

§3.5.2 Multilevel Monte Carlo estimator

This section is based on [48, 59]. Let $P, P_1, P_2, \ldots$ be square integrable random variables and set $P_0 = 0$. Let $\{D_k^i\}_{k\in\mathbb{N}\cup\{0\}, i\in\mathbb{N}}$ be independent random variables satisfying $D_k^i \overset{d}{=} D_k^1$ and $\mathbb{E}[D_k^i] = \mathbb{E}[P_{k+1} - P_k]$ for any $k \in \mathbb{N}\cup\{0\}$ and $i \in \mathbb{N}$. For $k \in \mathbb{N}\cup\{0\}$, assume that the bias and level variance satisfy $B(k) \geq |\mathbb{E}P - \mathbb{E}P_k|$ and $V(k) \geq \mathbb{V}[D_k^1]$ for some functions $k \mapsto B(k)$ and $k \mapsto V(k)$, respectively, and let $C(k)$ bound the expected computational complexity of simulating a single value of D_k^1. For $m \in \mathbb{N}\cup\{0\}$ and any $N_0, \ldots, N_m \in \mathbb{N}$, the MLMC estimator

$$\hat{P} = \sum_{k=0}^{m} \frac{1}{N_k} \sum_{i=1}^{N_k} D_k^i$$

satisfies $\mathbb{E}[(\hat{P} - \mathbb{E}P)^2] = \mathbb{V}[\hat{P}] + (\mathbb{E}P_m - \mathbb{E}P)^2$, since $\mathbb{E}\hat{P} = \mathbb{E}P_m$. Thus, for any $\epsilon > 0$, the inequality $\mathbb{E}\big[(\hat{P} - \mathbb{E}P)^2\big] < \epsilon^2$ holds if the number of levels in $\hat{P}$ equals

$$m = \inf\{k \in \mathbb{N}\cup\{0\} : B(k) < \epsilon/\sqrt{2}\} \tag{3.44}$$

and the variance is bounded by $\mathbb{V}[\hat{P}] = \sum_{k=0}^{m} \mathbb{V}[D_k^1]/N_k \leq \sum_{k=0}^{m} V(k)/N_k \leq \epsilon^2/2$. Since the computational complexity of $\hat{P}$, $\mathcal{C}_{\mathrm{ML}}(\epsilon) = \sum_{k=0}^{m} C(k)N_k$, is linear in the number of samples N_k on each level k, we only require that the variance $\mathbb{V}[\hat{P}]$ be of the same order as $\epsilon^2/2 = \sum_{k=0}^{m} V(k)/N_k$. Then, by the Cauchy-Schwartz inequality, we have

$$\mathcal{C}_{\mathrm{ML}}(\epsilon)\epsilon^2/2 = \Big(\sum_{k=1}^{m} C(k)N_k\Big)\Big(\sum_{k=0}^{m}\frac{V(k)}{N_k}\Big) \geq \Big(\sum_{k=0}^{m}\sqrt{C(k)V(k)}\Big)^2.$$

The lower bound does not depend on $N_0, \dots, N_m$ and is attained if and only if

$$N_k = \left\lceil 2\epsilon^{-2}\sqrt{\frac{V(k)}{C(k)}}\sum_{j=0}^{m}\sqrt{C(j)V(j)}\right\rceil \quad \text{for } k \in \{0,\dots,n\}, \tag{3.45}$$

ensuring that the expected cost is a multiple of

$$\mathcal{C}_{\mathrm{ML}}(\epsilon) = 2\epsilon^{-2}\Big(\sum_{k=0}^{m}\sqrt{C(k)V(k)}\Big)^2. \tag{3.46}$$

Moreover, if B, V and C are asymptotically sharp, the formulae in (3.45), up to constants, minimise the expected computational complexity. Consequently, the computational complexity analysis of the MLMC estimator is reduced to the analysis of the behaviour of $\sum_{j=0}^{m}\sqrt{C(j)V(j)}$ as $\epsilon \downarrow 0$. This analysis yields the following result.

Theorem 3.23. *Assume that for some $q_1 \geq (q_2 \wedge q_3)/2 > 0$, $c_1, c_2, c_3 > 0$ and all $n \in \mathbb{N}$ we have (a) $|\mathbb{E}P - \mathbb{E}P_n| \leq c_1 2^{-nq_1}$, (b) $\mathbb{V}[P_{n+1} - P_n] \leq c_2 2^{-nq_2}$, (c) the expected computational cost $\mathcal{C}(n)$ of constructing a single sample of (P_n, P_{n-1}) is bounded by $c_3 2^{nq_3}$. Then for every $\epsilon > 0$ there exist $n, N_1, \dots, N_n \in \mathbb{N}$ such that the MLMC estimator*

$$\hat{P} = \sum_{k=1}^{n}\frac{1}{N_k}\sum_{i=1}^{N_k} D_k^i \quad \textit{is } L^2\textit{-accurate at level } \epsilon, \quad \mathbb{E}\big[(\hat{P} - \mathbb{E}P)^2\big] < \epsilon^2, \tag{3.47}$$

and the computational complexity is of order

$$\mathcal{C}_{\mathrm{ML}}(\epsilon) = \begin{cases} \mathcal{O}(\epsilon^{-2}) & \text{if } q_2 > q_3, \\ \mathcal{O}(\epsilon^{-2}\log^2\epsilon) & \text{if } q_2 = q_3, \\ \mathcal{O}(\epsilon^{-2-(q_3-q_2)/q_1}) & \text{if } q_2 < q_3. \end{cases}$$

The number of levels equals $n = \lceil \log_2(\sqrt{2}c_1\epsilon^{-1})/q_1\rceil$ and the number of samples at *level $k \in \{1,\dots,n\}$ is*

$$N_k = \begin{cases} \lceil 2c_2\epsilon^{-2}2^{-(q_2+q_3)k/2}/(1-2^{-(q_2-q_3)/2})\rceil & \textit{if } q_2 > q_3, \\ \lceil 2c_2\epsilon^{-2}n2^{-q_3k}\rceil & \textit{if } q_2 = q_3, \\ \lceil 2c_2\epsilon^{-2}2^{n(q_3-q_2)/2-(q_2+q_3)k/2}/(1-2^{-(q_3-q_2)/2})\rceil & \textit{if } q_2 < q_3. \end{cases}$$

§3.6 Regularity of the density of the supremum

Let us briefly discuss the necessity of Assumption (H) in Proposition 3.7. Recall $\overline{\sigma}_\kappa^2 = \int_{(-\kappa,\kappa)} x^2\nu(\mathrm{d}x)$ for $\kappa \in (0,1)$ and note that X in the example below has smooth transition densities by [91, Prop. 28.3].

Example. For any $\gamma \in (0,1)$ there exists a Lévy process X with an absolutely continuous Lévy measure ν such that $\liminf_{u\downarrow 0} u^{\alpha-2}\overline{\sigma}_u^2 > 0$ holds for some $\alpha \in (0,1)$ and Assumption (H) fails for γ at countably many $M > 0$.

Proof. The essence of the proof is to construct any such M as a singularity of the density of ν. For simplicity and to make things explicit, we shall prove it for a single and fixed $M > 0$. To that end, let S be an α-stable process with positivity parameter $\rho = \mathbb{P}(S_1 > 0) \in (0,1)$ satisfying $\alpha\rho + \alpha + \rho < \gamma$. Let Z be an independent Lévy process with finite Lévy measure ν_Z given by $\nu_Z((-\infty,x]\setminus\{0\}) = \min\{1, (\max\{x,M\} - M)^\rho\}$ and put $X = S + Z$. Hereafter consider only small enough $\epsilon > 0$, namely, $\epsilon < \min\{(T/2)^{1/\alpha}, \min\{M,1\}/2\}$. The goal is to bound from below the probability $\mathbb{P}(\overline{X}_T \in [M, M+3\epsilon))$. To do this, we consider the event where Z jumps exactly once, S is small, $\overline{S} \le M$ at the time of that jump and S does not increase too much after the jump.

Since the density of S_1 is positive, continuous and bounded, it follows from the scaling property that there is some constant $K_1 > 0$ (not depending on ϵ) such that for all $t \le \epsilon^\alpha$,

$$\mathbb{P}(S_t \in [0,\epsilon), \overline{S}_t \le M) = \mathbb{P}(S_1 \in [0, t^{-1/\alpha}\epsilon), \overline{S}_1 \le t^{-1/\alpha}M) \ge K_1.$$

From [15, Thm 4A], we also know that $\mathbb{P}(\overline{S}_t \le \epsilon) \ge K_2\epsilon^{\alpha\rho}$ for some constant $K_2 > 0$ and all $t > T - \epsilon^\alpha/2$. Now, $Z_T \in [M, M+\epsilon)$ has probability $\mathrm{e}^{-T}T\epsilon^\rho$ since it can only happen if Z had a single jump on $[0,T]$, whose time U is then conditionally distributed $\mathrm{U}(0,T)$. For fixed $t \in (0,T)$ let $\hat{S}_t = \sup_{s\in[0,T-t]} S_{t+s} - S_t$ and note that the Markov property gives

$$\mathbb{P}(\hat{S}_t \in A, (S_t, \overline{S}_t) \in B \times C) = \mathbb{P}(\overline{S}_{T-t} \in A)\mathbb{P}((S_t, \overline{S}_t) \in B \times C),$$

for all measurable $A, B, C \subset \mathbb{R}$. Hence, multiplying by the density of U at t, integrating and using the independence of (U, Z) and S, we obtain

$$
\begin{aligned}
&\mathbb{P}(\overline{X}_T \in [M, M+3\epsilon)) \\
&\quad \geq \mathbb{P}(Z_T \in [M, M+\epsilon), S_U \in [0,\epsilon), \overline{S}_U \leq M, \overline{X}_T \in [M, M+3\epsilon)) \\
&\quad \geq \mathrm{e}^{-T} T \epsilon^\rho \int_0^T \mathbb{P}\big(\hat{S}_t \leq \epsilon, S_t \in [0,\epsilon), \overline{S}_t \leq M \big| Z_T \in [M, M+\epsilon), U = t\big) \frac{dt}{T} \\
&\quad \geq \mathrm{e}^{-T} \epsilon^\rho \int_0^{\epsilon^\alpha} \mathbb{P}(\overline{S}_{T-t} \leq \epsilon) \mathbb{P}(S_t \in [0,\epsilon), \overline{S}_t \leq M) dt \geq \mathrm{e}^{-T} K_1 K_2 \epsilon^{\alpha\rho+\alpha+\rho}.
\end{aligned}
$$

This implies that $x \mapsto \mathbb{P}(\overline{X}_T \leq x)$ is not locally γ-Hölder continuous at M. $\square$

Chapter 4

The stick-breaking Gaussian approximation

§4.1 The Gaussian approximation of the extrema of a Lévy process

Lévy processes are increasingly popular for the modeling of the market prices of risky assets. They naturally address the shortcoming of the diffusion models by allowing large (often heavy-tailed) sudden movements of the asset price observed in the markets [36, 68, 92]. For risk management, it is therefore crucial to quantify the probabilities of rare and/or extreme events in Lévy models. Of particular interest in this context are the distributions of the drawdown (the current decline from a historical peak) and its duration (the elapsed time since the historical peak), see e.g. [10, 26, 74, 96, 98]. Together with the hedges for barrier options [8, 49, 69, 94] and ruin probabilities in insurance [65, 75, 81], the expected drawdown and its duration constitute risk measures dependent on the random vector X.

Among the approximate simulation algorithms of $\overline{\chi}_T = (X_T, \overline{X}_T, \overline{\tau}_T(X))$, the SBA presented in Chapter 3 is the fastest in terms of its computational complexity, as it samples from the law of $\overline{\chi}_T$ with a geometrically decaying bias. However, the drawback is that it is only valid for Lévy process whose increments can be sampled. Such a requirement does not hold for large classes of widely used Lévy processes, including the general CGMY (aka KoBoL) model [25]. Moreover, nonparametric estimation of Lévy processes typically yields Lévy measures whose transitions cannot be sampled [23, 31, 34, 83, 88], again making a direct application SBA infeasible.

If the increments of X cannot be sampled, a general approach is to use the Gaussian approximation [6], which substitutes the small-jump component of the

Lévy process by a Brownian motion. Thus, the Gaussian approximation process is a jump diffusion and the exact sample of the random vector (consisting of the state of the process, the supremum and the time the supremum is attained) can be obtained by applying [40, Alg. MAXLOCATION] between the consecutive jumps. However, little is known about how close these quantities are to the vector $\overline{\chi}_T$ that is being approximated in either Wasserstein or Kolmogorov distances. Indeed, bounds on the distances between the marginal of the Gaussian approximation and X_T have been considered in [41] and recently improved in [24, 76]. A Wasserstein bound on the supremum is given in [41] but so far no improvement analogous to the marginal case has been established. Moreover, to the best of our knowledge, there are no corresponding results either for the joint law of $(X_T, \overline{X}_T)$ or the time $\overline{\tau}_T(X)$. Furthermore, as explained in §4.4.1 below, the exact simulation algorithm for the supremum and the time of the supremum of a Gaussian approximation based on [40, Alg. MAXLOCATION] is unsuitable for the multilevel Monte Carlo estimation.

The main objective of the present chapter is to provide an operational framework for Lévy processes, which allows us to settle the issues raised in the previous paragraph, develop a general simulation algorithm for $(X_T, \overline{X}_T, \overline{\tau}_T(X))$ and analyse the computational complexity of its Monte Carlo (MC) and multilevel Monte Carlo (MLMC) estimators.

The main results of this chapter can be grouped up in two. **(I)** We establish bounds on the Wasserstein and Kolmogorov distances between the vector $\overline{\chi}_T$ and its Gaussian approximation $\overline{\chi}_T^{(\kappa)} = (X_T^{(\kappa)}, \overline{X}_T^{(\kappa)}, \overline{\tau}_T(X^{(\kappa)}))$, where $X^{(\kappa)}$ is a jump diffusion equal to the Lévy process X with all the jumps smaller than $\kappa \in (0,1]$ substituted by a Brownian motion (see definition (4.5) below), and $\overline{X}_T^{(\kappa)}$ (resp. $\overline{\tau}_T(X^{(\kappa)})$) is the supremum of $X^{(\kappa)}$ (resp. the time $X^{(\kappa)}$ attains the supremum) over the time interval $[0,T]$. **(II)** We introduce a simple and fast algorithm, SBG-Alg, which samples exactly the vector of interest for the Gaussian approximation of any Lévy process X, develop an MLMC estimator based on SBG-Alg (see [51] for an implementation in Julia) and analyse its complexity for discontinuous and locally Lipschitz payoffs arising in applications. We now briefly discuss each of the two groups of results.

(I) In Theorem 4.3 (see also Corollary 4.4) we bound the Wasserstein distance between $\overline{\chi}_T$ and $\overline{\chi}_T^{(\kappa)}$ (as κ tends to 0) under weak assumptions, typically satisfied by the models used in applications. The proof of Theorem 4.3 has two main ingredients. First, in §4.6.2 below, we construct a novel *SBG coupling* between $\overline{\chi}_T$ and $\overline{\chi}_T^{(\kappa)}$, based on the SB representation of $\overline{\chi}_T$ in (4.1) and the minimal transport coupling between the increments of X and its approximation $X^{(\kappa)}$. The second ingredient consists of new bounds on the Wasserstein and Kolmogorov distances, given in Theorems 4.1 and 4.2 respectively, between the laws of X_t and $X_t^{(\kappa)}$ for any $t > 0$.

Theorem 4.3 is our main tool for controlling the distance between $\overline{\chi}_T$ and $\overline{\chi}_T^{(\kappa)}$. The SBG coupling underlying it cannot be simulated, but it provides a bound on the bias of SBG-Alg. Dominating the bias of the time $\overline{\tau}_T(X)$, which is a non-Lipschitz functional of the path of X, requires (by SB representations (4.1)) the bound in Theorem 4.2 on the Kolmogorov distance between the marginals. Applications related to the duration of drawdown and the risk-management of barrier options require bounding the bias of certain discontinuous functions of $\overline{\chi}_T$. In §4.3.2 we develop such bounds. Their proofs are based on Theorem 4.3 and Lemma 4.18 of §4.6.3, which essentially converts Wasserstein distance into Kolmogorov distance for sufficiently regular distributions. We give explicit general sufficient conditions on the characteristic triplet of the Lévy process X (see Proposition 4.12 below), which guarantee the applicability of the results of §4.3.2 to models typically used in practice. Moreover, we obtain bounds on the Kolmogorov distance between the components of $(\overline{X}_T, \overline{\tau}_T(X))$ and $(\overline{X}_T^{(\kappa)}, \overline{\tau}_T(X^{(\kappa)}))$ (see Corollary 4.11 below), which we hope are of independent interest.

(II) Our main simulation algorithm for this chapter, SBG-Alg, samples jointly coupled Gaussian approximations of $\overline{\chi}_T$ at distinct approximation levels. The coupling in SBG-Alg exploits the following simple observations:

- Any Gaussian approximation $\overline{\chi}_T^{(\kappa)}$ has an SB representation in (4.2), where the law of Y in (4.2) must equal that of $X^{(\kappa)}$.
- For any two Gaussian approximations, the stick-breaking process in (4.2) can be shared.
- The increments in (4.2) over the shared sticks can be coupled using the definition of the Gaussian approximation $X^{(\kappa)}$ in (4.5).

We analyse the computational complexity of the MLMC estimator based on SBG-Alg for a variety of payoff functions arising in applications. Figure 4.1 shows the leading power of the resulting MC and MLMC complexities, summarised in Tables 4.2 and 4.3 below (see Theorem 4.29 for full details), for locally Lipschitz and discontinuous payoffs used in practice. To the best of our knowledge, neither locally Lipschitz nor discontinuous payoffs had been previously considered in the context of MLMC estimation under Gaussian approximation.

A key component of the analysis of the complexity of an MLMC estimator is the rate of decay of level variances (see §3.5.2 for details). In the case of SBG-Alg, the rate of decay is given in Theorem 4.22 below for locally Lipschitz and discontinuous payoffs of interest. Moreover, the proof of Theorem 4.22 shows that the decay of the level variances for Lipschitz payoffs under SBG-Alg is asymptotically equal to that of Algorithm 2, which samples jointly the increments at two distinct levels only. Furthermore, an improved coupling in Algorithm 2 for the increments of the Gaussian approximations (cf. the last bullet on the list above) would reduce the computational complexity the MLMC estimator for all payoffs considered in this chapter (including the discontinuous ones). To the best of our knowledge, SBG-Alg is the first exact simulation algorithm for coupled Gaussian approximations of $\overline{\chi}_T$ with vanishing level variances when X has a Gaussian component, see also §4.4.1.

In §4.5, using the code in repository [51], we test our theoretical findings against numerical results. We run SBG-Alg for models in the tempered stable and Watanabe classes. The former is a widely used class of processes whose increments cannot be sampled for all parameter values and the latter is a well-known class of processes with infinite activity but singular continuous increments. In both cases we find a reasonable agreement between the theoretical prediction and the estimated decays of the bias and level variance, see Figures 4.3 & 4.4 below.

In the context of MC estimation, a direct simulation algorithm based on [40, Alg. MAXLOCATION] (Algorithm 3 below) can be used instead of SBG-Alg. In §4.5.2 we compare numerically its cost with that of SBG-Alg. In the examples we considered, the speedup of SBG-Alg over Algorithm 3 is about 50, see Figure 4.5, remaining significant even for processes with small jump activity, see Figure 4.6.

Comparison with the literature

As we have explained before, approximations of the pair $(X_T, \overline{X}_T)$ abound. They include the random walk approximation, a Wiener-Hopf based approximation [45, 70], the jump-adapted Gaussian (JAG) approximation [38, 39] and, more recently, the SB approximation [56]. The SB approximation converges the fastest as its bias decays geometrically in its computational cost. However, the JAG approximation is the only method known to us that does not require the ability to simulate the increments of the Lévy process X. Indeed, the JAG approximation simulates all jumps above a cutoff level, together with their jump times, and then samples the transitions of the Brownian motion from the Gaussian approximation on a random grid containing all the jump times. In contrast, in the present chapter we approximate $\overline{\chi}_T = (X_T, \overline{X}_T, \overline{\tau}_T(X))$ with an *exact* sample from the law of the Gaussian approximation $\overline{\chi}_T^{(\kappa)} = (X_T^{(\kappa)}, \overline{X}_T^{(\kappa)}, \overline{\tau}_T(X^{(\kappa)}))$.

The JAG approximation has been analysed for Lipschitz payoffs of the pair

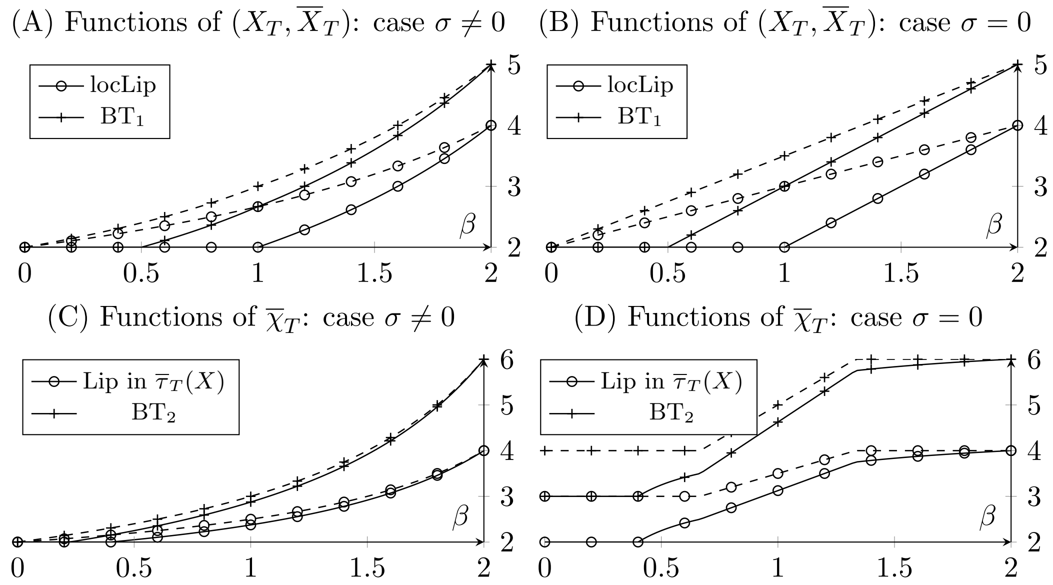

Figure 4.1: Dashed (resp. solid) line plots the power of ϵ^{-1} in the computational complexity of an MC (resp. MLMC) estimator, as a function of the BG index β defined in (4.6), for discontinuous functions in BT_1 (4.19) and BT_2 (4.21), locally Lipschitz payoffs as well as Lipschitz functions of $\overline{\tau}_T(X)$. The cases are split according to whether X is with ($\sigma \neq 0$) or without ($\sigma = 0$) a Gaussian component. The pictures are based on Tables 4.2 and 4.3 under assumptions typically satisfied in applications, see §4.4.2 below for details.

$(X_T, \overline{X}_T)$ in [38, 39]. The discontinuous and locally Lipschitz payoffs arising in applications, considered in this chapter (see Figure 4.1), have to the best of our knowledge not been analysed for the JAG approximation. Nor have the payoffs involving the time $\overline{\tau}_T(X)$ the supremum is attained. Within the class of Lipschitz payoffs of $(X_T, \overline{X}_T)$, the complexities of the MC and MLMC estimators based on SBG-Alg are asymptotically dominated by the complexities of the estimators based on the JAG approximation, see Figure 4.2. In fact, SBG-Alg with discontinuous payoffs outperforms the JAG approximation with Lipschitz payoffs by up to an order of magnitude in computational complexity, cf. Figure 4.1(A) & (B) and Figure 4.2.

In order to understand where the differences in Figure 4.2 come from, in Table 4.1 we summarise the bias and level variance for SBG-Alg and the JAG approximation as a function of the cutoff level κ in the Gaussian approximation (cf. (4.5) below).

Table 4.1 shows that both bias and level variance decay no slower (and typically faster) for SBG-Alg than for the JAG approximation. The large improvement in computational complexity of the MC estimator in Figure 4.2 is due to the faster decay of the bias under SBG-Alg. Put differently, the SBG coupling constructed

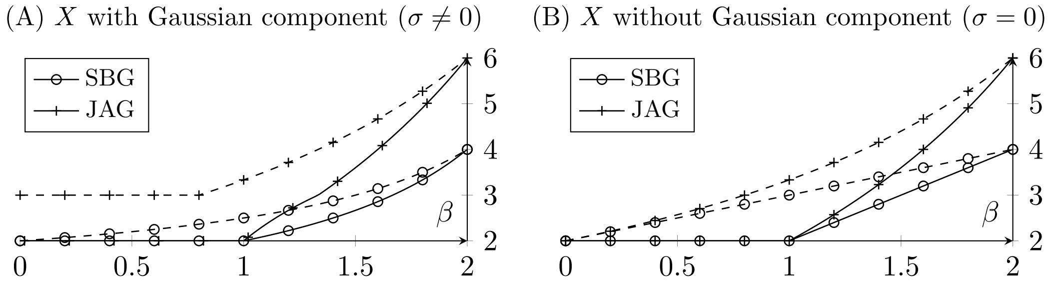

Figure 4.2: Dashed (resp. solid) lines represent the power of ϵ^{-1} in the computational complexity of the MC (resp. MLMC) estimator for the expectation of a Lipschitz functional $f(X_T, \overline{X}_T)$, plotted as a function of the BG index β defined in (4.6). The SBG plots are based on Tables 4.2 and 4.3 below. The JAG plots are based on [38, Cor. 3.2] for the MC cost, and [38, Cor. 1.2] if $\beta \geq 1$ (resp. [39, Cor. 1] if $\beta < 1$) for the MLMC cost.

Gaussian comp.	Approx.	Bias	Level variance
With ($\sigma \neq 0$)	JAG	$\max\{\kappa^{1-\beta/4}, \kappa^{\beta/2}\} \log^{1/2}(1/\kappa)$	$\max\{\kappa^{2-\beta}, \kappa^{\beta} \log(1/\kappa)\}$
	SBG	$\kappa^{3-\beta} \log(1/\kappa)$	$\kappa^{2-\beta}$
Without ($\sigma = 0$)	JAG	$\max\{\kappa^{1-\beta/4} \log^{1/2}(1/\kappa), \kappa^{\beta}\}$	$\max\{\kappa^{2-\beta}, \kappa^{2\beta}\}$
	SBG	$\kappa \log(1/\kappa)$	$\kappa^{2-\beta}$

Table 4.1: The rates (as $\kappa \to 0$) of decay of bias and level variance for Lipschitz payoffs of $(X_T, \overline{X}_T)$ under the JAG approximation are based on [38, Cor. 3.2] and [39, Thm 2], respectively. The rates on the bias and level variance for the SBG-Alg are given in Theorems 4.3 & 4.22 below.

in this chapter controls the Wasserstein distance much better than the KMT-based coupling in [38]. For the BG index $\beta > 1$, the improvement in the computational complexity of the MLMC estimator is mostly due to an faster bias decay. For $\beta < 1$, Figure 4.2(A) suggests that the computational complexity of the MLMC estimator under both algorithms is optimal. However, in this case, Table 4.1 and the equality in (3.46) imply that the MLMC estimator based on the JAG approximation has a computational complexity proportional to $\epsilon^{-2} \log^3(1/\epsilon)$ while that of SBG-Alg is proportional to ϵ^{-2}. This improvement is due solely to the faster decay of level variance under SBG-Alg. The numerical experiments in §4.5.1 suggest that our bounds for Lipschitz and locally Lipschitz functions are sharp, see graphs (A) & (C) in Figures 4.3 & 4.4.

To the best of our knowledge, in the literature there are no directly comparable results to either Theorem 4.3 or Proposition 4.9. Partial results in the direction of Theorem 4.3 are given in [24, 41, 76]. We will now briefly comment on these results. *Distance between the marginals* X_t *and* $X_t^{(\kappa)}$: Theorem 4.1 below, a key step in the proof of Theorem 4.3, improves the bounds in [76, Thm 9] on the Wasserstein distance. Theorem 4.2 below, a further key ingredient in the proof of Theorem 4.3, bounds the Kolmogorov distance with better rates than those of [41, Prop. 10 (part

1)] (as $\kappa \to 0$). Papers [24, 76] obtain bounds on the total variation distance between X_t and $X_t^{(\kappa)}$, dominating the Kolmogorov distance. However, Theorem 4.2 again yields faster decay. For more details about these comparisons see §4.3.1 below.
Distance between the suprema $\overline{X}_t$ *and* $\overline{X}_t^{(\kappa)}$: the rate of the bound in [41, Thm 2] on the Wasserstein distance is worse than that implied by the bound in Corollary 4.4 below on the Wasserstein distance between $(X_t, \overline{X}_t)$ and $(X_t^{(\kappa)}, \overline{X}_t^{(\kappa)})$. Proposition 4.6 below bounds the bias of locally Lipschitz functions, generalising [41, Prop. 9] and providing a faster decay rate. Proposition 4.7 and Corollary 4.11(a) below cover a class of discontinuous payoffs, including the up-and-in digital option considered in [41, Prop. 10 (part 3)], and provide a faster rate of decay as $\kappa \to 0$ if either X has a Gaussian component or the BG index $\beta > 2/3$.

The remainder of the chapter is organised as follows. In §4.2 we recall the SB representation (see (4.1)–(4.2)) for the infima of Lévy processes and the Gaussian approximation (see (4.5)) developed in [53] and [6], respectively. §4.3 presents bounds on Wasserstein and Kolmogorov distances between $\underline{\chi}_T$ and its Gaussian approximation $\underline{\chi}_T^{(\kappa)}$ and the biases of certain payoffs arising in applications. §4.3 also provides simple sufficient conditions, in terms of the Lévy triplet, under which these bounds hold. §4.4 constructs our main algorithm, SBG-Alg, and presents the computational complexity of the corresponding MC and MLMC estimators for all payoffs considered in this chapter. In §4.5 we illustrate numerically these results for a widely used class of Lévy models. The proofs and the technical results are found in §4.6. §3.5.1 gives a brief account of the complexity analysis of MC and MLMC (introduced in [48, 59]) estimators.

§4.2 The stick-breaking representation and the Gaussian approximation

We begin by restating (2.3) for the infimum of X on $[0, T]$, which is at the core of the bounds and algorithms developed in this chapter. Given a Lévy process X and a time horizon $t > 0$, there exists a coupling (X, Y), where $Y \overset{d}{=} X$, and a uniform stick-breaking process $\ell = (\ell_n)_{n\in\mathbb{N}}$ on $[0, t]$ (i.e. $L_0 = t$, $L_n = L_{n-1}U_n$, $\ell_n = L_n - L_{n-1}$ for $n \in \mathbb{N}$, where $(U_n)_{n\in\mathbb{N}}$ is an iid sequence following the uniform law $U_n \sim \mathrm{U}(0, 1)$), such that a.s.

$$\underline{\chi}_t = \sum_{k=1}^{\infty} \left(\xi_k, \min\{\xi_k, 0\}, \ell_k \cdot \mathbb{1}_{\{\xi_k \le 0\}}\right), \qquad \xi_k = Y_{L_{k-1}} - Y_{L_k}. \tag{4.1}$$

Since, given L_n, $(\ell_k)_{k>n}$ is a stick-breaking process on $[0, L_n]$, for any $n \in \mathbb{N}$, (4.1) implies

$$\underline{\chi}_t \overset{d}{=} (Y_{L_n}, \underline{Y}_{L_n}, \underline{\tau}_{L_n}(Y)) + \sum_{k=1}^{n} \big(\xi_k, \min\{\xi_k, 0\}, \ell_k \cdot \mathbb{1}_{\{\xi_k \le 0\}}\big). \tag{4.2}$$

We stress that (4.1) and (4.2) reduce the analysis of the path-functional $\underline{\chi}_t$ to that of the increments of X, since the "error term" $(Y_{L_n}, \underline{Y}_{L_n}, \underline{\tau}_{L_n}(Y))$ in (4.2) is typically exponentially small in n. More generally, for another Lévy process X', the vectors $\underline{\chi}_t$ and $(X'_t, \underline{X}'_t, \underline{\tau}_t(X'))$ will be close if the increments of Y and Y' over the intervals $[L_k, L_{k-1}]$ are close: apply (4.2) with a single stick-breaking process ℓ, independent of both Lévy processes $Y \overset{d}{=} X$ and $Y' \overset{d}{=} X'$, respectively. This observation constitutes a key step in the construction of the coupling used in the proof of Theorem 4.3 below, which in turn plays a crucial role in controlling the bias (see the subsequent results of §4.3) of our main simulation algorithm SBG-Alg described in §4.4 below. SBG-Alg is based on (4.2) with X' being the Gaussian approximation of a general Lévy process X introduced in [6] and recalled briefly next.

Recall the Lévy-Khintchine formula [91, Thm 8.1]: for $u \in \mathbb{R}$,

$$\frac{1}{t} \log \mathbb{E}\big[e^{iuX_t}\big] = iub - \frac{1}{2}u^2\sigma^2 + \int_{\mathbb{R}\setminus\{0\}} \big(e^{iux} - 1 - iux \cdot \mathbb{1}_{(-1,1)}(x)\big)\nu(dx), \tag{4.3}$$

where the Lévy measure ν satisfies $\int_{\mathbb{R}\setminus\{0\}} \min\{x^2, 1\}\nu(dx) < \infty$ and $\sigma \ge 0$ specifies the volatility of the Brownian component of X. Note that the 'drift' $b \in \mathbb{R}$ depends on the cutoff function $x \mapsto \mathbb{1}_{(-1,1)}(x)$. Thus the Lévy triplet (σ^2, ν, b), with respect to the cutoff function $x \mapsto \mathbb{1}_{(-1,1)}(x)$, determines the law of X. All the Lévy triplets in the present chapter use this cutoff function.

The *Lévy-Itô decomposition* at level $\kappa \in (0, 1]$ (see [91, Thms 19.2 & 19.3]) is given by

$$X_t = b_\kappa t + \sigma B_t + J_t^{1,\kappa} + J_t^{2,\kappa}, \qquad t \ge 0, \tag{4.4}$$

where $b_\kappa = b - \int_{(-1,1)\setminus(-\kappa,\kappa)} x\nu(dx)$, $B = (B_t)_{t\ge 0}$ is a standard Brownian motion and the processes $J^{1,\kappa} = (J_t^{1,\kappa})_{t\ge 0}$ and $J^{2,\kappa} = (J_t^{2,\kappa})_{t\ge 0}$ are Lévy with triplets $(0, \nu|_{(-\kappa,\kappa)}, 0)$ and $(0, \nu|_{\mathbb{R}\setminus(-\kappa,\kappa)}, b - b_\kappa)$, respectively. The processes B, $J^{1,\kappa}$, $J^{2,\kappa}$ in (4.4) are independent, $J^{1,\kappa}$ is an L^2-bounded martingale with jumps of magnitude less than κ and $J^{2,\kappa}$ is a driftless (i.e. piecewise constant) compound Poisson process with intensity $\overline{\nu}(\kappa) = \nu(\mathbb{R} \setminus (-\kappa, \kappa))$ and jump distribution $\nu|_{\mathbb{R}\setminus(-\kappa,\kappa)}/\overline{\nu}(\kappa)$.

In applications, the main problem lies in the user's inability to simulate the increments of $J^{1,\kappa}$ in (4.4), i.e. the small jumps of the Lévy process X. Instead of ignoring this component for a small value of κ, the Gaussian approximation [6]

$$X_t^{(\kappa)} = b_\kappa t + \sqrt{\overline{\sigma}_\kappa^2 + \sigma^2} W_t + J_t^{2,\kappa}, \quad \text{where } \overline{\sigma}_\kappa^2 = \int_{(-\kappa,\kappa)} x^2 \nu(\mathrm{d}x), \quad \kappa \in (0,1], \tag{4.5}$$

substitutes the martingale $\sigma B + J^{1,\kappa}$ in (4.4) with a Brownian motion with variance $\overline{\sigma}_\kappa^2 + \sigma^2$. In (4.5), the standard Brownian motion $W = (W_t)_{t\geq 0}$ is independent of $J^{2,\kappa}$. Let $\overline{\sigma}_\kappa$ denote the non-negative square root of $\overline{\sigma}_\kappa^2$. The *Gaussian approximation* of X at level κ, given by the Lévy process $X^{(\kappa)} = (X_t^{(\kappa)})_{t\geq 0}$, is natural in the following sense: the weak convergence $\overline{\sigma}_\kappa^{-1} J^{1,\kappa} \overset{d}{\to} W$ (in the Skorokhod space $\mathrm{d}[0,\infty)$) as $\kappa \to 0$ holds if and only if $\overline{\sigma}_{\min\{K\overline{\sigma}_\kappa,\kappa\}}/\overline{\sigma}_\kappa \to 1$ for every $K > 0$ (see [6]). This condition holds if $\overline{\sigma}_\kappa/\kappa \to \infty$ and the two conditions are equivalent if ν has no atoms in a neighbourhood of zero [6, Prop. 2.2].

Since $J^{2,\kappa}$ has an average of $\overline{\nu}(\kappa)t$ jumps on $[0,t]$, the expected complexity of simulating the increment $X_t^{(\kappa)}$ is a constant multiple of $1 + \overline{\nu}(\kappa)t$ (see Algorithm 2 below). Moreover, the user need only be able to sample from the normalised tails of ν, which can typically be achieved in multiple ways (see e.g. [90]). The behaviour of $\overline{\nu}(\kappa)$ and $\overline{\sigma}_\kappa$ as $\kappa \downarrow 0$, key in the analysis of the MC/MLMC complexity, can be described in terms of the *Blumenthal-Getoor* (BG) index [18] β, defined as

$$\beta = \inf\{p > 0 : I_0^p < \infty\}, \quad \text{where} \quad I_0^p = \int_{(-1,1)\setminus\{0\}} |x|^p \nu(\mathrm{d}x) \text{ for } p \geq 0. \tag{4.6}$$

Note that $\beta \in [0,2]$, since $I_0^2 < \infty$ by the definition of the Lévy measure ν. Furthermore, $I_0^1 < \infty$ if and only if the paths of $J^{1,\kappa}$ have finite variation. Moreover, $I_0^p < \infty$ for any $p > \beta$, but I_0^β can be either finite or infinite. If $q \in [0,2]$ satisfies $I_0^q < \infty$, the following inequalities hold for all $\kappa \in (0,1]$ (see Lemma 3.11):

$$\overline{\sigma}_\kappa^2 \leq I_0^q \kappa^{2-q} \qquad \text{and} \qquad \overline{\nu}(\kappa) \leq \overline{\nu}(1) + I_0^q \kappa^{-q}. \tag{4.7}$$

Finally we stress that the dependence between W in (4.5) and $\sigma B + J^{1,\kappa}$ in (4.4) has not been specified. This coupling will vary greatly, depending on the circumstance (e.g. the analysis of the Wasserstein distance between functionals of X and $X^{(\kappa)}$ in §4.3 or the minimisation of level variances in MLMC in §4.4). Thus, unless otherwise stated, no explicit dependence between $\sigma B + J^{1,\kappa}$ and W is assumed.

§4.3 Distance between the extrema and its Gaussian approximation

In this section we present bounds on the distance between the laws of the vectors $\underline{\chi}_t$, defined in §4.2 above, and its Gaussian approximation $\underline{\chi}_t^{(\kappa)} = (X_t^{(\kappa)}, \underline{X}_t^{(\kappa)}, \underline{\tau}_t(X^{(\kappa)}))$, based on the Lévy process $X^{(\kappa)}$ in (4.5). Our bounds on the Wasserstein distance (see Theorem 4.3 and Corollary 4.4 in §4.3.1) are based on a coupling constructed in §4.6.2 below, which in turn draws on the coupling in (4.1). Theorem 4.3 is then applied to control the bias of certain discontinuous and non-Lipschitz functions of $\underline{\chi}_t$ arising in applications (§4.3.2 below) as well as the Kolmogorov distances between the components of $(\underline{X}_t, \underline{\tau}_t(X))$ and $(\underline{X}_t^{(\kappa)}, \underline{\tau}_t(X^{(\kappa)}))$ (see §4.3.3 below).

§4.3.1 Bounds on the Wasserstein and Kolmogorov distances

In order to study the Wasserstein distance between $\underline{\chi}_t$ and $\underline{\chi}_t^{(\kappa)}$ via (4.1)–(4.2), we have to quantify the Wasserstein and Kolmogorov distances between the increments X_s and $X_s^{(\kappa)}$ for any time $s > 0$. With this in mind, we start with Theorems 4.1 and 4.2, which play a key role in the proofs of the main results of the subsection, Theorem 4.3 and Corollary 4.4 below, and are of independent interest.

Theorem 4.1. *There exist universal constants $K_1 = 1/2$ and $K_p > 0$, $p \in (1, 2]$, independent of (σ^2, ν, b), such that for any $t > 0$ and $\kappa \in (0, 1]$ there exists a coupling $(X_t, X_t^{(\kappa)})$ satisfying*

$$\mathbb{E}\big[\big|X_t - X_t^{(\kappa)}\big|^p\big]^{1/p} \le \min\big\{\sqrt{2t}\overline{\sigma}_\kappa, K_p\kappa\varphi_\kappa^{2/p}\big\}, \\ \text{where} \quad \varphi_\kappa = \overline{\sigma}_\kappa/\sqrt{\overline{\sigma}_\kappa^2 + \sigma^2}, \quad \text{for all} \quad p \in [1, 2]. \tag{4.8}$$

Theorem 4.1 bounds the L^p-Wasserstein distance (see (4.17) below for definition) between X_t and $X_t^{(\kappa)}$. The inequality in (4.8) sharpens the bound $\mathbb{E}[|X_t - X_t^{(\kappa)}|^p]^{1/p} \le \min\{\sqrt{2t}\overline{\sigma}_\kappa, K_p\kappa\}$ in [76, Thm 9]: the factor $\varphi_\kappa^{2/p} \in [0, 1]$ tends to zero (with $\kappa \to 0$) as a constant multiple of $\overline{\sigma}_\kappa^{2/p}$ if the Brownian component is present (i.e. $\sigma > 0$) and is equal to 1 when $\sigma = 0$. The bound in (4.8) cannot be improved in general in the sense that there exists a Lévy processes for which, up to constants, the reverse inequality holds (see [76, Rem. 3] and [47, Sec. 4]).

The proof of Theorem 4.1, given in §4.6.1 below, decomposes the increment $M_t^{(\kappa)}$ of the Lévy martingale $M^{(\kappa)} = \sigma B + J^{1,\kappa}$ into a sum of m iid copies of $M_{t/m}^{(\kappa)}$ and applies a Berry-Essen-type bound for the Wasserstein distance [89] in the context of a central limit theorem (CLT) as $m \to \infty$. The small-time moment asymptotics of $M_{t/m}^{(\kappa)}$ in [46] imply that $M_t^{(\kappa)}$ is much closer to the Gaussian limit in the CLT if the Brownian component is present than if $\sigma = 0$. This explains a vastly superior rate in (4.8) in the case $\sigma^2 > 0$.

Bounds on the Kolmogorov distance may require the following generalisation of Orey's condition, which makes the distribution of X_t sufficiently regular (see [91, Prop. 28.3]).

Assumption (O-δ)**.** *For some $\delta \in (0, 2]$ we have $\inf_{u\in(0,1]} u^{\delta-2}(\overline{\sigma}_u^2 + \sigma^2) > 0$.*

Theorem 4.2. (a) *There exists a constant $C_{\mathrm{BE}} \in (0, \frac{1}{2})$, such that for any $\kappa \in (0, 1]$, $t > 0$ we have:*

$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}\big(X_t \le x\big) - \mathbb{P}\big(X_t^{(\kappa)} \le x\big)\big| \le C_{\mathrm{BE}}(\kappa/\overline{\sigma}_\kappa)\varphi_\kappa^3/\sqrt{t}. \tag{4.9}$$

(b) *Let Assumption (O-δ) hold. Then for every $T > 0$ there exists a constant $C > 0$, depending only on (T, δ, σ, ν), such that for any $\kappa \in (0, 1]$ and $t \in (0, T]$ we have:*

$$\sup_{x\in\mathbb{R}} \big|\mathbb{P}\big(X_t \le x\big) - \mathbb{P}\big(X_t^{(\kappa)} \le x\big)\big| \le \big(Ct^{-1/\delta} \min\big\{\sqrt{t}\overline{\sigma}_\kappa, \kappa\varphi_\kappa\big\}\big)^{2/3}. \tag{4.10}$$

The proof of Theorem 4.2 is in §4.6.1 below. Part (a) follows the same strategy as the proof of Theorem 4.1, applying the Berry-Esseen theorem (instead of [89, Thm 4.1]) to bound the Kolmogorov distance. For the same reason as in (4.8), the rate in (4.9) is far better if $\sigma^2 > 0$. Proof of Theorem 4.2(b) bounds the density of X_t using results in [86] and applies (4.8).

Note that no assumption is made on the Lévy process X in Theorem 4.2(a). In particular, Assumption (O-δ) is not required in part (a); however, if (O-δ) is not satisfied, implying in particular that $\sigma = 0$, it is possible for the bound in (4.9) not to vanish as $\kappa \to 0$ even if the Lévy process has infinite activity, i.e. $\nu(\mathbb{R}\setminus\{0\}) = \infty$. In fact, if $\sigma = 0$, the bound in (4.9) vanishes (as $\kappa \to 0$) if and only if $\overline{\sigma}_\kappa/\kappa \to \infty$, which is also a necessary and sufficient condition for the weak limit $\overline{\sigma}_\kappa^{-1} J^{1,\kappa} \xrightarrow{d} W$ to hold whenever ν has no atoms in a neighbourhood of 0 (see [6, Prop. 2.2]).

If X has a Brownian component (i.e. $\sigma \neq 0$), the bound on the total variation distance between the laws of X_t and $X_t^{(\kappa)}$ established in [76, Prop. 8] implies the following upper bound on the Kolmogorov distance:

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(X_t \le x) - \mathbb{P}(X_t^{(\kappa)} \le x)| \le \min\{\sqrt{8t}\overline{\sigma}_\kappa, \kappa\}/\sqrt{2\pi\sigma^2 t}.$$

This inequality is both generalised and sharpened (as $\kappa \to 0$) by the bound in (4.9). Further improvements to the bound on the total variation were made in [24], but the implied rates for the Kolmogorov distance are worse than the ones in Theorem 4.2

and require model restrictions when $\sigma = 0$ (beyond those of Theorem 4.2(b)) that can be hard to verify (see [24, Subsec. 2.1.1]).

We stress that the dependence in t in the bounds of Theorem 4.2 is explicit. This is crucial in the proof of Theorem 4.3 as we need to apply (4.9)–(4.10) over intervals of small random lengths. A related result [41, Prop. 10] contains similar bounds, which are non-explicit in t and suboptimal in κ.

If Assumption (O-δ) is satisfied, the parameter δ in part (b) of Theorem 4.2 should be taken as large as possible to get the sharpest inequality in (4.10). If $\sigma \neq 0$ (equivalently $\delta = 2$), the bound in part (a) has a faster decay in κ than the bound in part (b). If $\sigma = 0$ (equivalently $0 < \delta < 2$), it is possible for the bound in part (a) to be sharper than the one in part (b) or vice versa. Indeed, it is easy to construct a Lévy measure ν such that $\delta \in (0, 2)$ in Theorem 4.2(b) satisfies $\lim_{u\downarrow 0} u^{\delta-2}\overline{\sigma}_u^2 = \inf_{u\in(0,1]} u^{\delta-2}\overline{\sigma}_u^2 = 1$. Then the bound in (4.9) is a multiple of $t^{-1/2}\kappa^{\delta/2}$ as $t, \kappa \to 0$, while the one in (4.10) behaves as $t^{-2/(3\delta)}\kappa^{2/3} \min\{1, t^{1/3}\kappa^{-\delta/3}\}$. Hence one bound may be sharper than the other depending on the value of δ, as t and/or κ tend to zero. In fact, we will use the bound in part (b) only when the maximal δ satisfying the assumption of Theorem 4.2(b) is smaller than $4/3$, bounding the activity of the Lévy measure around 0 away from maximal possible activity.

Denote $x^+ = \max\{x, 0\}$ for $x \in \mathbb{R}$. The next result quantifies the Wasserstein distance between the laws of the vectors $\underline{\chi}_t$ and $\underline{\chi}_t^{(\kappa)}$.

Theorem 4.3. *For any $\kappa \in (0, 1]$ and $t > 0$, there exists a coupling between X and $X^{(\kappa)}$ on the interval $[0, t]$ such that the following inequalities hold for $p \in \{1, 2\}$:*

$$\mathbb{E}\big[\max\big\{\big|X_t - X_t^{(\kappa)}\big|, \big|\underline{X}_t - \underline{X}_t^{(\kappa)}\big|\big\}^p\big]^{1/p} \le \mu_p(\kappa, t), \qquad \textit{where} \tag{4.11}$$

$$\mu_1(\kappa, t) := \min\big\{2\sqrt{2t}\overline{\sigma}_\kappa, \kappa\varphi_\kappa^2\big\}\big(1 + \log^+\big(2\sqrt{2t}(\overline{\sigma}_\kappa/\kappa)\varphi_\kappa^{-2}\big)\big),$$

$$\mu_2(\kappa, t) = \sqrt{2}\mu_1(\kappa, t) + \min\big\{\sqrt{2t}\overline{\sigma}_\kappa, K_2\kappa\varphi_\kappa\big\}\sqrt{1 + 2\log^+\big(K_2^{-1}\sqrt{2t}(\overline{\sigma}_\kappa/\kappa)\varphi_\kappa^{-1}\big)}, \tag{4.12}$$

with $\varphi_\kappa = \overline{\sigma}_\kappa/\sqrt{\overline{\sigma}_\kappa^2 + \sigma^2}$ and K_2 as in Theorem 4.1. Furthermore, we have

$$\mathbb{E}\big|\underline{\tau}_t(X) - \underline{\tau}_t(X^{(\kappa)})\big| \le \mu_0^\tau(\kappa, t) = \sqrt{t}(\kappa/\overline{\sigma}_\kappa)\varphi_\kappa^3. \tag{4.13}$$

Moreover, if Assumption (O-δ) holds, then for every $T > 0$ there exists a constant

$C > 0$, dependent only on (T, δ, σ, ν), such that for all $t \in [0, T]$ and $\kappa \in (0, 1]$,

$$\mathbb{E}|\underline{\tau}_t(X) - \underline{\tau}_t(X^{(\kappa)})| \le \mu_\delta^\tau(\kappa, t), \quad \textit{where} \quad \psi_\kappa = C\kappa\varphi_\kappa \quad \textit{and} \tag{4.14}$$

$$\mu_\delta^\tau(\kappa, t) = \begin{cases} t \wedge \psi_\kappa^\delta + t^{1-\frac{2}{3\delta}} \psi_\kappa^{\frac{2}{3}} \big(1 - \min\big\{1, t^{-\frac{1}{\delta}}\psi_\kappa\big\}^{\delta - \frac{2}{3}}\big), & \delta \in (0, 2] \setminus \{\frac{2}{3}\}, \\ \big(t \wedge \psi_\kappa^{\frac{2}{3}}\big)\big(1 + \log^+\big(t\psi_\kappa^{-\frac{2}{3}}\big)\big), & \delta = \frac{2}{3}. \end{cases} \tag{4.15}$$

The coupling in Theorem 4.3 satisfying the above inequalities will be hereafter referred to as the *SBG coupling* $(X, X^{(\kappa)})$. The SBG coupling is constructed in the proof of Theorem 4.3 (found in in §4.6.2 below) in terms of the distribution functions of the marginals X_s and $X_s^{(\kappa)}$ (for $s > 0$) and the coupling used in (4.1), see [56] for the latter. The key idea is to couple $\underline{\chi}_t$ and $\underline{\chi}_t^{(\kappa)}$ so that they share the stick-breaking process in their respective SB representations (4.1), while the increments of the associated Lévy processes over each interval $[L_n, L_{n-1}]$ are coupled so that they minimise appropriate Wasserstein distances. This coupling produces a bound on the distance between $\underline{\chi}_t$ and $\underline{\chi}_t^{(\kappa)}$ that depends only on the distances between the marginals of X_s and $X_s^{(\kappa)}$, $s > 0$, so that Theorems 4.1 and 4.2 above can be applied. We stress that the bound in (4.11) cannot be obtained from Doob's L^2-maximal inequality (see, e.g. [62, Prop. 7.16]) and Theorem 4.1: if the processes X and $X^{(\kappa)}$ are coupled in such a way that $X_t - X_t^{(\kappa)}$ satisfies the inequality in (4.8), the difference process $(X_s - X_s^{(\kappa)})_{s \in [0,t]}$ need not be a martingale.

Inequality (4.11) holds without assumptions on X and is at most a logarithmic factor worse than the marginal inequality (4.8) for $p \in \{1, 2\}$, with the upper bound satisfying $\mu_p(\kappa, t) \le 2\kappa \log(1/\kappa)$ for all sufficiently small κ. Moreover, by Jensen's inequality, for all $1 < p < 2$ the SBA coupling satisfies the following inequality: $\mathbb{E}[\max\{|X_t - X_t^{(\kappa)}|, |\underline{X}_t - \underline{X}_t^{(\kappa)}|\}^p]^{1/p} \le \mu_2(\kappa, t)$. In the absence of a Brownian component (i.e. $\sigma = 0$) we have $\varphi_\kappa = 1$, making the upper bound $\mu_2(\kappa, t)$ proportional to $\mu_1(\kappa, t)$ as $\kappa \to 0$. If $\sigma > 0$, then $\mu_1(\kappa, t) \le 2\kappa\overline{\sigma}_\kappa^2 \log(1/(\kappa\overline{\sigma}_\kappa))/\sigma^2$ for all small κ and, typically, $\mu_2(\kappa, t)$ is proportional to $\kappa\overline{\sigma}_\kappa\sqrt{\log(1/(\kappa\overline{\sigma}_\kappa))}$ as $\kappa \to 0$, which dominates $\mu_1(\kappa, t)$.

The bound in (4.13) holds without assumptions on X, while (4.14) requires Assumption (O-δ) and is sharper the larger the value of $\delta \in (0, 2]$, satisfying (O-δ), is. Note that, if $\sigma \neq 0$, (O-δ) holds with $\delta = 2$. If $\sigma = 0$ and δ satisfies (O-δ), we must have $\beta \ge \delta$, where β is the *Blumenthal–Getoor (BG) index* defined in (4.6) above. In fact, models typically used in applications either have $\sigma \neq 0$ or (O-δ) holds with $\delta = \beta$ (however, it is possible for (O-δ) to hold for some $\delta < \beta$ but not $\delta = \beta$, cf. [91, p. 362]).

If $\sigma > 0$, the inequality in (4.13) is sharper than (4.14), i.e. $\mu_0^\tau(t, \kappa) \le \mu_2^\tau(t, \kappa)$

for all small $\kappa > 0$. However, if $\sigma = 0$ and $\delta \in (0,2)$ satisfies (O-δ), then typically $\mu_0^\tau(\kappa,t)$ is proportional to $\kappa^{\delta/2}$, while $\mu_\delta^\tau(\kappa,t)$ is asymptotically proportional to $\kappa^{\min\{2/3,\delta\}}(1+\log(1/\kappa)\mathbb{1}_{\{2/3\}}(\delta))$ as $\kappa \to 0$, implying that (4.14) is sharper than (4.13) for $\delta < 4/3$. The following quantity is the smallest of the upper bounds in (4.13) and (4.14):

$$\mu_*^\tau(\kappa,t) = \min\left\{\mu_0^\tau(\kappa,t), \inf\left\{\mu_\delta^\tau(\kappa,t) : \delta \in (0,2] \text{ satisfies Assumption (O-}\delta)\right\}\right\}.$$

Under Assumption (O-δ), for some constant $c_t > 0$ and all $\kappa \in (0,1]$, we have

$$\mu_*^\tau(\kappa,t) \le c_t \kappa^{\max\{\delta/2,\min\{2/3,\delta\}\}}(1+\log(1/\kappa)\cdot \mathbb{1}_{\{2/3\}}(\delta)). \tag{4.16}$$

For any $a \in \mathbb{R}^d$, let $|a| = \sum_{i=1}^d |a_i|$ denote its ℓ^1-norm. Recall that for $p \ge 1$, the L^p-Wasserstein distance [99, Def. 6.1] between the laws of random vectors ξ and ζ in $\mathbb{R}^d$ can be defined as

$$\mathcal{W}_p(\xi,\zeta) = \inf\left\{\mathbb{E}\left[|\xi'-\zeta'|^p\right]^{1/p} : \xi' \overset{d}{=} \xi, \zeta' \overset{d}{=} \zeta\right\}. \tag{4.17}$$

Theorem 4.3 implies a bound on the L^p-Wasserstein distance between the vectors $\underline{\chi}_t$ and $\underline{\chi}_t^{(\kappa)}$, extending the bound on the distance between the laws of the marginals X_t and $X_t^{(\kappa)}$ in [76, Thm 9].

Corollary 4.4. *Fix $\kappa \in (0,1]$ and $t > 0$. Then we have*

$$\begin{aligned} \mathcal{W}_p((X_t,\underline{X}_t),(X_t^{(\kappa)},\underline{X}_t^{(\kappa)})) &\le 2(\mathbb{1}_{\{p=1\}}\mu_1(\kappa,t) + \mathbb{1}_{\{1<p\le 2\}}\mu_2(\kappa,t)), \qquad p \in [1,2], \\ \mathcal{W}_p(\underline{\tau}_t(X),\underline{\tau}_t(X^{(\kappa)})) &\le t^{1-1/p}\mu_*^\tau(\kappa,t)^{1/p}, \qquad p \ge 1. \end{aligned}$$

Moreover, for $p \in [1,2]$, we have

$$\mathcal{W}_p(\underline{\chi}_t,\underline{\chi}_t^{(\kappa)}) \le 2^{2-1/p}(\mathbb{1}_{\{p=1\}}\mu_1(\kappa,t) + \mathbb{1}_{\{1<p\le 2\}}\mu_2(\kappa,t)) + (2t)^{1-1/p}\mu_*^\tau(\kappa,t)^{1/p}.$$

Given the bounds in Corollary 4.4 and Theorem 4.2, it is natural to inquire about the convergence in Kolmogorov distance of the components of $(\underline{X}_t^{(\kappa)},\underline{\tau}_t(X^{(\kappa)}))$ to $(\underline{X}_t,\underline{\tau}_t(X))$ as $\kappa \to 0$. This question is addressed by Corollary 4.11 of §4.3.3.

The famous Kómlos-Major-Tusnády (KMT) coupling is used in [38, Thm 6.1] to bound the L^2-Wasserstein distance between the paths of X and $X^{(\kappa)}$ on $[0,t]$ in the supremum norm, implying a bound on $\mathcal{W}_2((X_t,\underline{X}_t),(X_t^{(\kappa)},\underline{X}_t^{(\kappa)}))$ proportional to $\kappa\log(1/\kappa)$ as $\kappa \to 0$, cf. [38, Cor. 6.2]. If $\sigma > 0$, $\mu_2(\kappa,t)$ in (4.11) is bounded by a multiple of $\kappa\overline{\sigma}_\kappa\log(1/(\kappa\overline{\sigma}_\kappa))$ for small κ and is thus smaller than a multiple of $\kappa^{2-q/2}$ for any $q \in (\beta, 2)$ (where β is the BG index defined in (4.6) above). As mentioned above, $\mu_2(\kappa, t)$ is bounded by a multiple of $\kappa \log(1/\kappa)$ for small κ in the case $\sigma = 0$. Unlike the SBG coupling, which underpins Theorem 4.3, the KMT coupling does not imply a bound on the distance between the times of the infima $\underline{\tau}_t(X)$ and $\underline{\tau}_t(X^{(\kappa)})$ as these are not Lipschitz functionals of the trajectories with respect to the supremum norm.

Remark 4.5. *The bounds on* $\mathbb{E}|\underline{\tau}_t(X)-\underline{\tau}_t(X^{(\kappa)})|$ *in Theorem 4.3 and Corollary 4.4, based on the SB representation in* (4.1)*, require the control on the expected difference between the signs of the components of* $(X_s, X_s^{(\kappa)})$ *as either* s *or* κ *tend to zero. This is achieved via the minimal transport coupling (see* (4.25) *and Lemma 4.15 below) and a general bound in Theorem 4.2 on the Kolmogorov distance. However, further improvements seem possible in the finite variation case if the* natural drift *(i.e. the drift of* X *when small jumps are not compensated) is nonzero. Intuitively, the sign of the natural drift determines the sign of both components of* $(X_s, X_s^{(\kappa)})$ *with overwhelming likelihood as* $s \to 0$*. This suggestion is left for future research.*

§4.3.2 Bounds on the bias

The main tool used to study and bound the bias of various Lipschitz, non-Lipschitz and discontinuous functions of $\underline{\chi}_t$ is the SBG coupling underpinning Theorem 4.3. The Lipschitz case is a direct consequence: for any $d \in \mathbb{N}$, let $\mathrm{Lip}_K(\mathbb{R}^d)$ denote the space of real-valued Lipschitz functions on $\mathbb{R}^d$ (under ℓ^1-norm given above display (4.17)) with Lipschitz constant $K \geq 0$ and note that the triangle inequality and Theorem 4.3 imply the following bounds on the bias

$$\begin{aligned} \left|\mathbb{E}f(X_T, \underline{X}_T) - \mathbb{E}f\big(X_T^{(\kappa)}, \underline{X}_T^{(\kappa)}\big)\right| &\leq 2K\mu_1(\kappa, T) \qquad \text{and} \\ \left|\mathbb{E}g(\underline{\tau}_T) - \mathbb{E}g\big(\underline{\tau}_T(X^{(\kappa)})\big)\right| &\leq K'\mu_*^{\tau}(\kappa, T) \end{aligned} \tag{4.18}$$

for any time horizon $T > 0$ and $f \in \mathrm{Lip}_K(\mathbb{R}^2)$, such that $\mathbb{E}|f(X_T, \underline{X}_T)| < \infty$, and $g \in \mathrm{Lip}_{K'}(\mathbb{R})$. Since in applications, the process X is often used to model log-returns of a risky asset $(S_0 e^{X_t})_{t\geq 0}$, it is natural to study the bias of a Monte Carlo estimator of a locally Lipschitz function $f \in \mathrm{locLip}_K(\mathbb{R}^2)$, satisfying $|f(x,y) - f(x',y')| \leq K\big(|e^x - e^{x'}| + |e^y - e^{y'}|\big)$ for any $x, x', y, y' \in \mathbb{R}$ (equivalently $(x,y) \mapsto f(\log x, \log y)$ is in $\mathrm{Lip}_K((0,\infty)^2)$). Such payoffs arise in risk management (e.g. absolute drawdown) and in the pricing of hindsight call, perpetual American call and lookback put options.

Proposition 4.6. *Let* $f \in \mathrm{locLip}_K(\mathbb{R}^2)$ *and assume* $\int_{[1,\infty)} e^{2x}\nu(dx) < \infty$*, where* ν *is the Lévy measure of* X*. For any* $T > 0$ *and* $\kappa \in (0,1]$ *and* $\mu_2(\kappa, T)$ *defined* *in* (4.12)*, the SBG coupling satisfies*

$$\mathbb{E}\big|f(X_T,\underline{X}_T)-f\big(X_T^{(\kappa)},\underline{X}_T^{(\kappa)}\big)\big|\le 4K\mathbb{E}[\mathrm{e}^{2X_T}]^{1/2}(1+\mathrm{e}^{\overline{\sigma}_\kappa^2 T})\mu_2(\kappa,T).$$

The assumption $\int_{[1,\infty)}\mathrm{e}^{2x}\nu(\mathrm{d}x)<\infty$ is equivalent to $\mathbb{E}[\mathrm{e}^{2X_T}]<\infty$ (see [91, Thm 25.3]), which is a natural requirement as the asset price model $(S_0\mathrm{e}^{X_t})_{t\ge0}$ ought to have finite variance. Moreover, via the Lévy-Khintchine formula, an explicit bound on the expectation $\mathbb{E}[\mathrm{e}^{2X_T}]$ (and hence the constant in the inequality of Proposition 4.6) in terms of the Lévy triplet of X can be obtained. If one instead considers $f(X_T,\overline{X}_T)$ (a function on the supremum $\overline{X}_T$), the proof of Proposition 4.6 in §4.6.3 below can be used to establish that $\mathbb{E}\big|f(X_T,\overline{X}_T)-f\big(X_T^{(\kappa)},\overline{X}_T^{(\kappa)}\big)\big|$ is bounded by $4K(\mathbb{E}[\mathrm{e}^{2\overline{X}_T}]+\mathbb{E}[\mathrm{e}^{2\overline{X}_T^{(\kappa)}}])^{1/2}\mu_2(\kappa,T)$, where both expectations $\mathbb{E}[\mathrm{e}^{2\overline{X}_T}]$ and $\mathbb{E}[\mathrm{e}^{2\overline{X}_T^{(\kappa)}}]$ are finite under our assumption $\int_{[1,\infty)}\mathrm{e}^{2x}\nu(\mathrm{d}x)<\infty$ and bounded explicitly in terms of the Lévy triplet of X, see the proof of Proposition 3.6. Thus, by Proposition 4.6, the bias for $f\in\mathrm{locLip}_K(\mathbb{R}^2)$ is at most a multiple of $\kappa\log(1/\kappa)$, as is the case for $f\in\mathrm{Lip}_K(\mathbb{R}^2)$ by (4.18), cf. discussion after Theorem 4.3.

In financial markets, the class of barrier-type functions arises naturally: for $K,M\ge0$, $y<0$ let

$$\mathrm{BT}_1(y,K,M)=\{f\,:\,f(x,z)=h(x)\mathbb{1}_{[y,\infty)}(z),\ h\in\mathrm{Lip}_K(\mathbb{R}),\ 0\le h\le M\}. \tag{4.19}$$

Note that the indicator function $\mathbb{1}_{[y,\infty)}$ lies in $\mathrm{BT}_1(y,0,1)$ and satisfies the identity $\mathbb{E}[\mathbb{1}_{[y,\infty)}(\underline{X}_T)]=\mathbb{P}(\underline{X}_T\ge y)$. Moreover, a down-and-out put option payoff $x\mapsto\max\{\mathrm{e}^k-\mathrm{e}^x,0\}\mathbb{1}_{[y,\infty)}(x)$, for constants $y<0<k$, is in $\mathrm{BT}_1(y,\mathrm{e}^k,\mathrm{e}^k-\mathrm{e}^y)$. Bounding the bias of the estimators for functions in $\mathrm{BT}_1(y,K,M)$ requires the following regularity of the distribution of $\underline{X}_T$ at y (analogous to Assumption (H) in Chapter 3).

Assumption (H)**.** *Given $C,\gamma>0$ and $y<0$, we have $\mathbb{P}(y<\underline{X}_T\le y+x)\le Cx^\gamma$ for all $x>0$.*

Proposition 4.7. *Let $f\in\mathrm{BT}_1(y,K,M)$ for some $K,M\ge0$ and $y<0$. If y and some $C,\gamma>0$ satisfy Assumption (H), then for any $T>0$ and $\kappa\in(0,1]$, the SBG coupling satisfies*

$$\begin{aligned}\mathbb{E}\big|f(X_T,\underline{X}_T)-f\big(X_T^{(\kappa)},\underline{X}_T^{(\kappa)}\big)\big|&\\ \le K\mu_1(\kappa,T)+M'\min\{\mu_1(\kappa,T)^{\gamma/(1+\gamma)},\mu_2(\kappa,T)^{2\gamma/(2+\gamma)}\},&\end{aligned} \tag{4.20}$$

where $M'=M\max\{(1+1/\gamma)(2C\gamma)^{1/(1+\gamma)},(1+2/\gamma)(C\gamma)^{2/(2+\gamma)}\}$.

Remark 4.8. *Since* $\mu_1(\kappa, T) \to 0$ *and* $\mu_2(\kappa, T) \to 0$ *as* $\kappa \to 0$ *and* $\gamma/(1+\gamma) < 2\gamma/(2+\gamma)$ *for all* $\gamma > 0$*, the bound in* (4.20) *is typically dominated by a multiple of* $\mu_1(\kappa, T)^{\gamma/(1+\gamma)}$*, if* $\sigma \neq 0$ *and* $\beta < 2 - \gamma$ *(recall the definition of the BG index* β *in* (4.6)*), or* $\mu_2(\kappa, T)^{2\gamma/(1+\gamma)}$*, otherwise. By Hölder's inequality,* f *in* (4.20) *need not be bounded if appropriate moments of* X *exist.*

The proof of Proposition 4.7 is in §4.6.3 below. Assumption (H) with $\gamma = 1$ requires the distribution function of $\underline{X}_T$ to be locally Lipschitz at y. By the Lebesgue differentiation theorem [33, Thm 6.3.3], any distribution function is differentiable Lebesgue-a.e., implying that Assumption (H) holds for $\gamma = 1$ and a.e. $y < 0$. However, there exist Lévy processes satisfying Assumption (H) for countably many levels y with $\gamma < 1$, but not with $\gamma = 1$, see [56, App. B]. Proposition 4.12 below provides simple sufficient conditions, in terms of the Lévy triplet of X, for (H) to hold with $\gamma = 1$ for all $y < 0$. In particular, this is the case if $\sigma \neq 0$.

The next class arises in the analysis of the duration of drawdown: for $K, M \geq 0$, $s \in (0, T)$ define the set:

$$\begin{aligned}&\mathrm{BT}_2(s, K, M)\\&\quad = \{f : f(x, z, t) = h(x, z)\mathbb{1}_{(s,T]}(t),\, h \in \mathrm{Lip}_K(\mathbb{R}^2),\, 0 \leq h \leq M\}.\end{aligned} \tag{4.21}$$

The biases of these functions include $|\mathbb{P}(\underline{\tau}_T(X) > s) - \mathbb{P}(\underline{\tau}_T(X^{(\kappa)}) > s)|$. Analogous to Proposition 4.7, we require the following regularity from the distribution function of $\underline{\tau}_T(X)$.

Assumption (Hτ). *Given* $C, \gamma > 0$ *and* $s \in (0, T)$*, the following inequality holds,*

$$|\mathbb{P}(\underline{\tau}_T(X) \leq s) - \mathbb{P}(\underline{\tau}_T(X) \leq s + t)| \leq C|t|^\gamma, \quad \textit{for all } t \in \mathbb{R}.$$

Proposition 4.9. *Let Assumption (Hτ) hold for some* $s \in (0, T)$ *and* $C, \gamma > 0$*. Let* $f \in \mathrm{BT}_2(s, K, M)$ *for some* $K, M \geq 0$*. Then for all* $\kappa \in (0, 1]$ *the SBG coupling satisfies*

$$\mathbb{E}\big|f(\underline{\chi}_T) - f\big(\underline{\chi}_T^{(\kappa)}\big)\big| \leq 2K\mu_1(\kappa, T) + M(2C\gamma)^{1/(1+\gamma)}(1 + 1/\gamma)\mu_*^\tau(\kappa, T)^{\gamma/(1+\gamma)}. \tag{4.22}$$

Remark 4.10. *As in Remark 4.8, the bound in* (4.22) *is asymptotically proportional to* $\mu_*^\tau(\kappa, T)^{\gamma/(1+\gamma)}$ *as* $\kappa \to 0$*. Inequality* (4.22) *can be generalised to unbounded function* f *if appropriate moments of* X *exist.*

If X is not a compound Poisson process, then Assumption $(H\tau)$ holds with $\gamma = 1$ for all $s \in (0, T)$, since, by Lemma 4.20 in §4.6.3 below, $\underline{\tau}_T(X)$ has a locally bounded density, making the distribution function of $\underline{\tau}_T(X)$ locally Lipschitz on $(0,T)$. Assumption $(H\tau)$ is satisfied if either $\nu(\mathbb{R}\setminus\{0\})=\infty$ or $\sigma\neq 0$. In particular, Assumption (O-δ) implies $(H\tau)$. The proof of Proposition 4.9 is in §4.6.3 below.

§4.3.3 Convergence in the Kolmogorov distance

As a consequence of Proposition 4.7 (resp. 4.9), if Assumption (H) (resp. $(H\tau)$) holds uniformly, then $\underline{X}_T^{(\kappa)}$ (resp. $\underline{\tau}_T(X^{(\kappa)})$) converges to $\underline{X}_T$ (resp. $\underline{\tau}_T(X)$) in Kolmogorov distance as $\kappa\to 0$.

Corollary 4.11. (a) *Suppose $C,\gamma>0$ satisfy (H) for all $y<0$. Then for any $\kappa\in(0,1]$ we have*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}(\underline{X}_T\le x)-\mathbb{P}\big(\underline{X}_T^{(\kappa)}\le x\big)\right|\le M'\min\{\mu_1(\kappa,T)^{\frac{\gamma}{1+\gamma}},\mu_2(\kappa,T)^{\frac{2\gamma}{2+\gamma}}\},\qquad(4.23)$$

where $M'=\max\{(1+1/\gamma)(2C\gamma)^{1/(1+\gamma)},(1+2/\gamma)(C\gamma)^{2/(2+\gamma)}\}$.
(b) *Suppose $C,\gamma>0$ satisfy $(H\tau)$ for all $s\in[0,T]$. Then for any $\kappa\in(0,1]$ we have*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}(\underline{\tau}_T(X)\le x)-\mathbb{P}\big(\underline{\tau}_T(X^{(\kappa)})\le x\big)\right|\le(2C\gamma)^{\frac{1}{1+\gamma}}(1+\tfrac{1}{\gamma})\mu_*^\tau(\kappa,T)^{\frac{\gamma}{1+\gamma}}.\qquad(4.24)$$

Proposition 4.12 gives sufficient conditions (in terms of the Lévy triplet (σ^2,ν,b) of X) for Assumptions (H) and $(H\tau)$ to hold for all $y<0$ and $s\in[0,T]$, respectively. Recall that a function $f(x)$ is said to be regularly varying with index r as $x\to 0$ if $\lim_{x\to 0}f(\lambda x)/f(x)=\lambda^r$ for every $\lambda>0$ (see [16, p. 18]).

Proposition 4.12. *Let $\overline{\nu}_+(x)=\nu([x,\infty))$ and $\overline{\nu}_-(x)=\nu((-\infty,-x])$ for $x>0$ and let β be the BG index of X defined in* (4.6) *above. Suppose that either (I) $\sigma>0$ or (II) the Lévy measure ν satisfies the following conditions: $\overline{\nu}_+(x)$ is regularly varying with index $-\beta$ as $x\to 0$ and*
- $\beta=2$ *and* $\liminf_{x\to 0}\overline{\nu}_+(x)/\overline{\nu}_-(x)>0$,
- $\beta\in(1,2)$ *and* $\lim_{x\to 0}\overline{\nu}_+(x)/\overline{\nu}_-(x)\in(0,\infty]$ *or*
- $\beta\in(0,1)$, $b=\int_{(-1,1)}x\nu(\mathrm{d}x)$ *and* $\lim_{x\to 0}\overline{\nu}_+(x)/\overline{\nu}_-(x)\in(0,\infty)$.

Then there exists constants $\gamma>0$ and C such that Assumption $(H\tau)$ holds with γ,C for all $s\in[0,T]$. Either (I) or (II) with $\beta>1$ imply that (H) holds with $\gamma=1$ and some constant C_I for all y in a compact $I\subset(-\infty,0)$.

Note that Proposition 4.12 holds if the roles of $\overline{\nu}_+$ and $\overline{\nu}_-$ are interchanged, i.e $\overline{\nu}_-(x)$ is regularly varying and the limit conditions are satisfied by the quotients $\overline{\nu}_-(x)/\overline{\nu}_+(x)$. The assumptions of Proposition 4.12 are satisfied by most models used in practice, including tempered stable and most subordinated Brownian motion processes. Excluded are Lévy processes without a Brownian component and with barely any jump activity (i.e. BG index $\beta = 0$, which includes compound Poisson and variance gamma processes), where the Gaussian approximation $X^{(\kappa)}$ is not useful.

Proposition 4.12 is a consequence of a more general result, Proposition 4.21 below, stating that Assumptions ($H\tau$) and (H) hold uniformly and locally uniformly, respectively, if over short time horizons, X is "attracted to" an α-stable process with non-monotone paths, see §4.6.3 below for details. In this case $\rho = \lim_{t\to 0}\mathbb{P}(X_t > 0)$ exists in $(0, 1)$ and γ in the conclusion of Proposition 4.12, satisfying ($H\tau$) on $[0, T]$, can be arbitrarily chosen in the interval $(0, \min\{\rho, 1-\rho\})$. In contrast to ($H\tau$), a simple sufficient condition for the uniform version of (H), required in Corollary 4.11(a), remains elusive beyond special cases such as stable or tempered stable processes with γ in the interval $(0, \alpha(1-\rho))$, where α is the stability parameter and ρ is as above.

§4.4 Simulation and the computational complexity of MC and MLMC

This section describes an MC and MLMC methods for the simulation of the vector $\underline{\chi}_T^{(\kappa)} = (X_T^{(\kappa)}, \underline{X}_T^{(\kappa)}, \underline{\tau}_T(X^{(\kappa)}))$ (SBG-Alg in §4.4.1) and analyses the computational complexities for various locally Lipschitz and discontinuous functions of $\underline{\chi}_T^{(\kappa)}$ (§4.4.2). The numerical performance of SBG-Alg, which is based on the SB representation in (4.1)-(4.2) of $\underline{\chi}_T^{(\kappa)}$, is far superior to that of the "obvious" algorithm for jump diffusions (see Algorithm 3 below), particularly when the jump intensity is large (cf. §4.4.1 and §4.4.1). Moreover, SBG-Alg is designed with MLMC in mind, which turns out not to be feasible in general for the "obvious" algorithm (see §4.4.1).

§4.4.1 Simulation of the extrema

The main aim of the subsection is to develop a simulation algorithm for the pair of vectors $(\underline{\chi}_T^{(\kappa)}, \underline{\chi}_T^{(\kappa')})$ at levels $\kappa, \kappa' \in (0, 1]$ over a time horizon $[0, T]$, such that the L^2-distance between $\underline{\chi}_T^{(\kappa)}$ and $\underline{\chi}_T^{(\kappa')}$ tends to zero as $\kappa, \kappa' \to 0$. SBG-Alg below, based on the SB representation in (4.2), achieves this aim: it applies Algorithm 2 for the increments over the stick-breaking lengths that arise in (4.2) and Algorithm 3 for the "error term" over the time horizon $[0, L_n]$. By Theorem 4.22 below the L^2-distance for the coupling given in SBG-Alg decays to zero, ensuring the feasibility of MLMC (see Theorem 4.29 for the computational complexity of MLMC).

Increments in the SB representation

A simulation algorithm for a coupling $\big(X_t^{(\kappa_1)}, X_t^{(\kappa_2)}\big)$ of Gaussian approximations (at levels $1 \geq \kappa_1 > \kappa_2 > 0$) of X_t at an arbitrary time $t > 0$ is based on the following observation: the compound Poisson processes J^{2,κ_1} and J^{2,κ_2} in the Lévy-Itô decomposition in (4.4) can be simulated jointly, as the jumps of J^{2,κ_1} are precisely those of J^{2,κ_2} with modulus of at least κ_1. By choosing the same Brownian motion W in representation (4.5) of $X_t^{(\kappa_1)}$ and $X_t^{(\kappa_2)}$, we obtain the coupling $\big(X_t^{(\kappa_1)}, X_t^{(\kappa_2)}\big)$ with law $\Pi_t^{\kappa_1,\kappa_2}$ given in Algorithm 2.

Algorithm 2. Simulation of the law $\Pi_t^{\kappa_1,\kappa_2}$

Require: Cutoff levels $1 \geq \kappa_1 > \kappa_2 > 0$ and time horizon $t > 0$.

1: Compute b_{κ_i} and $\overline{\sigma}^2_{\kappa_i}$ for $i \in \{1,2\}$ and $\overline{\nu}(\kappa_2)$
2: Sample $W_t \sim N(0,t)$, $N_t \sim \mathrm{Poi}(\overline{\nu}(\kappa_2)t)$ and $\lambda_k \sim \nu(\cdot \setminus (-\kappa_2,\kappa_2))/\overline{\nu}(\kappa_2)$ for $k \in \{1,\ldots,N_t\}$
3: Put $J_t^{2,\kappa_i} = \sum_{k=1}^{N_t} \lambda_k \cdot \mathbb{1}\{|\lambda_k| \geq \kappa_i\}$ for $i \in \{1,2\}$
4: **return** $\big(Z_t^{(\kappa_1)}, Z_t^{(\kappa_2)}\big)$, where $Z_t^{(\kappa_i)} = b_{\kappa_i}t + \sqrt{\sigma^2 + \overline{\sigma}^2_{\kappa_i}}W_t + J_t^{2,\kappa_i}$ for $i \in \{1,2\}$

Since $Z_t^{(\kappa_i)} \overset{d}{=} X_t^{(\kappa_i)}$, $i \in \{1,2\}$, Proposition 4.23(a) below implies that the coupling $\Pi_t^{\kappa_1,\kappa_2}$ provides a bound on the L^2-Wasserstein distance $\mathcal{W}_2\big(X_t^{(\kappa_1)}, X_t^{(\kappa_2)}\big) \leq (2t(\overline{\sigma}^2_{\kappa_1} - \overline{\sigma}^2_{\kappa_2}))^{1/2}$. This bound is suboptimal as the variables $J_t^{2,\kappa_2} - J_t^{2,\kappa_1}$ and $(\overline{\sigma}^2_{\kappa_2} - \overline{\sigma}^2_{\kappa_1})^{1/2}W_t$ in Algorithm 2 are independent. The minimal transport coupling, with L^2-distance equal to $\mathcal{W}_2\big(X_t^{(\kappa_1)}, X_t^{(\kappa_2)}\big)$, is not accessible from the perspective of simulation. Since the law $\mathrm{Poi}(\overline{\nu}(\kappa_2)t)$ of the variable N_t in line 2 of Algorithm 2 is Poisson with mean $\overline{\nu}(\kappa_2)t$, the expected number of steps of Algorithm 2 is bounded by a constant multiple of $1 + \overline{\nu}(\kappa_2)t$, which is in turn bounded by a negative power of κ_2 by (4.7). Since the computational complexity of sampling the law of $X_t^{(\kappa_2)}$ is of the same order as that of the law $\Pi_t^{\kappa_1,\kappa_2}$, in the complexity analysis of SBG-Alg below, we may apply Algorithm 2 with $\Pi_t^{1,\kappa}$ to sample $X_t^{(\kappa)}$ for any $\kappa \in (0,1]$.

"Error term" in the SB representation

Algorithm 3 samples from the law $\underline{\Pi}_t^{\kappa_1,\kappa_2}$ of a coupling $(\underline{\chi}_t^{(\kappa_1)}, \underline{\chi}_t^{(\kappa_2)})$ for levels $0 < \kappa_2 < \kappa_1 \leq 1$ and any (typically very small) $t > 0$. In particular, it requires the sampler [40, Alg. MAXLOCATION] for the law $\Phi_t(v,\mu)$ of $(B_t^*, \underline{B}_t^*, \tau_t(B^*))$ where $(B_s^*)_{s\geq 0} = (vB_s + \mu s)_{s \geq 0}$ is a Brownian motion with drift $\mu \in \mathbb{R}$ and volatility $v > 0$.

Algorithm 3 samples the jump times and sizes of the compound Poisson process J^{2,κ_2} on the interval $(0,t)$ and prunes the jumps to get J^{2,κ_1}. Then it samples

Algorithm 3. Simulation of the law $\underline{\Pi}_t^{\kappa_1,\kappa_2}$

Require: Cutoff levels $1 \ge \kappa_1 > \kappa_2 > 0$ and time horizon $t > 0$.

Compute b_{κ_i}, $\overline{\sigma}^2_{\kappa_i}$ and $\upsilon_{\kappa_i} = \sqrt{\sigma^2 + \overline{\sigma}^2_{\kappa_i}}$ for $i \in \{1,2\}$ and $\overline{\nu}(\kappa_2)$, see (4.4)–(4.5)
Sample $N_t \sim \mathrm{Poi}(\overline{\nu}(\kappa_2)t)$ and $U_k \sim \mathrm{U}(0,t)$ for $k \in \{1,\dots,N_t+1\}$
Set $s = \sum_{i=1}^{N_t+1} \log U_k$ and let $t_k = s^{-1}\sum_{i=1}^{k} \log U_i$ for $k \in \{0,\dots,N_t+1\}$
Set $(Z_0^{(\kappa_1)}, \underline{Z}_0^{(\kappa_1)}, \underline{\tau}_0^{(\kappa_1)}, Z_0^{(\kappa_2)}, \underline{Z}_0^{(\kappa_2)}, \underline{\tau}_0^{(\kappa_2)}) = (0,0,0,0,0,0)$
for $k \in \{1,\dots,N_t+1\}$ **do**
    Sample $\lambda_k \sim \nu(\cdot \setminus (-\kappa_2,\kappa_2))/\overline{\nu}(\kappa_2)$ if $k \le N_t$ and otherwise put $\lambda_k = 0$
    Let $\delta_k = t_k - t_{k-1}$ and sample $(\Delta^1_{k,i}, \Delta^2_{k,i}, \Delta^3_{k,i}) \sim \Phi_{\delta_k}(\upsilon_{\kappa_i}, b_{\kappa_i})$ independently for $i \in \{1,2\}$
    for $i \in \{1,2\}$ **do**
        if $\underline{Z}_{t_{k-1}}^{(\kappa_i)} > Z_{t_{k-1}}^{(\kappa_i)} + \Delta^2_{k,i}$ **then**
            Set $(Z_{t_k}^{(\kappa_i)}, \underline{Z}_{t_k}^{(\kappa_i)}, \underline{\tau}_{t_k}^{(\kappa_i)}) = (Z_{t_{k-1}}^{(\kappa_i)} + \Delta^1_{k,i} + \lambda_k \cdot \mathbb{1}_{\{|\lambda_k| \ge \kappa_i\}}, Z_{t_{k-1}}^{(\kappa_i)} + \Delta^2_{k,i}, t_{k-1} + \Delta^3_{k,i})$
        else
            Set $(Z_{t_k}^{(\kappa_i)}, \underline{Z}_{t_k}^{(\kappa_i)}, \underline{\tau}_{t_k}^{(\kappa_i)}) = (Z_{t_{k-1}}^{(\kappa_i)} + \Delta^1_{k,i} + \lambda_k \cdot \mathbb{1}_{\{|\lambda_k| \ge \kappa_i\}}, \underline{Z}_{t_{k-1}}^{(\kappa_i)}, \underline{\tau}_{t_{k-1}}^{(\kappa_i)})$
        end if
    end for
end for
return $(\underline{\zeta}^{(\kappa_1)}, \underline{\zeta}^{(\kappa_2)})$, where $\underline{\zeta}^{(\kappa_i)} = (Z_t^{(\kappa_i)}, \underline{Z}_t^{(\kappa_i)}, \underline{\tau}_t^{(\kappa_i)})$ for $i \in \{1,2\}$

the increment, infimum and the time the infimum is attained for the Brownian motion with drift on each interval between the jumps of J^{2,κ_2} and assembles the pair $(\underline{\zeta}^{(\kappa_1)}, \underline{\zeta}^{(\kappa_2)})$, clearly satisfying $\underline{\zeta}^{(\kappa_i)} \overset{d}{=} \underline{\chi}_t^{(\kappa_i)}$, $i \in \{1,2\}$. As [40, Alg. MAXLOCATION] samples the law $\Phi_t(v,\mu)$ with uniformly bounded expected runtime over the choice of parameters μ, v and t, the computational cost of sampling the pair of vectors $(\underline{\chi}_t^{(\kappa_1)}, \underline{\chi}_t^{(\kappa_2)})$ using Algorithm 3 is proportional to to the cost of sampling $X_t^{(\kappa)}$ via Algorithm 2.

In principle, Algorithm 3 is an exact algorithm for the simulation of a coupling $(\underline{\chi}_t^{(\kappa_1)}, \underline{\chi}_t^{(\kappa_2)})$. However, it cannot be applied within an MLMC simulation scheme for a function of $\underline{\chi}_T^{(\kappa)}$ at a fixed time horizon T (the next paragraph explains why). SBG-Alg below circumvents this issue via the SB representation in (4.2), which also makes SBG-Alg *paralellizable* and thus much faster in practice even in the context of MC simulation (see the discussion after Corollary 4.26 below).

To the best of our knowledge, there is no simulation algorithm for the increment, the infima and the times the infima are attained of a Brownian motion under different drifts, i.e. of the vector

$$\big(B_t, \underline{B}_t^{(c_1)}, \underline{\tau}_t(B^{(c_1)}), \underline{B}_t^{(c_2)}, \underline{\tau}_t(B^{(c_2)})\big), \text{ where } (B_s^{(c)})_{s\ge 0} = (B_s + cs)_{s\ge 0} \text{ and } c_1 \ne c_2.$$

Thus, in line 7 of Algorithm 3, we are forced to take independent samples from $\Phi_{\delta_k}(\upsilon_{\kappa_1}, b_{\kappa_1})$ and $\Phi_{\delta_k}(\upsilon_{\kappa_2}, b_{\kappa_2})$ at each step k. In particular, the coupling of the marginals $X_t^{(\kappa_1)}$ and $X_t^{(\kappa_2)}$ of $\underline{\Pi}_t^{\kappa_1,\kappa_2}$, given in line 16 of Algorithm 3, amounts to taking two independent Brownian motions in the respective representations in (4.5) of $X_t^{(\kappa_1)}$ and $X_t^{(\kappa_2)}$. Thus, unlike the coupling defined in Algorithm 2, here, by Proposition 4.23(b) below, the squared L^2-distance satisfies $\mathbb{E}[(X_t^{(\kappa_1)} - X_t^{(\kappa_2)})^2] \geq 2t\sigma^2$ for all levels $1 \geq \kappa_1 > \kappa_2 > 0$, where σ^2 is the Gaussian component of X. Hence, for a fixed time horizon, the coupling $\underline{\Pi}_t^{\kappa_1,\kappa_2}$ of $\underline{\chi}_t^{(\kappa_1)}$ and $\underline{\chi}_t^{(\kappa_2)}$ is not sufficiently strong for an MLMC scheme to be feasible if X has a Gaussian component, because the level variances do not decay to zero. However, by Proposition 4.23(b), the L^2-distance between $\underline{\zeta}^{(\kappa_1)}$ and $\underline{\zeta}^{(\kappa_2)}$ constructed in Algorithm 3 does tend to zero with $t \to 0$. Thus, SBG-Alg below, which applies Algorithm 3 over the time interval $[0, L_n]$ (recall $\mathbb{E}L_n = T/2^n$ from SB representation (4.2)), circumvents this issue.

The SBG sampler

For a time horizon T, we can now define the coupling $\underline{\Pi}_{n,T}^{\kappa_1,\kappa_2}$ of the vectors $\underline{\chi}_T^{(\kappa_1)}$ and $\underline{\chi}_T^{(\kappa_2)}$ via the following algorithm.

Algorithm 4. SBG-Alg

Require: Cutoffs $1 \geq \kappa_1 > \kappa_2 > 0$, number of sticks $n \geq 0$, time horizon $T > 0$.

1: Set $L_0 = T$, sample $U_k \sim \mathrm{U}(0,1)$, put $\ell_k = L_{k-1}U_k$ and $L_k = L_{k-1} - \ell_k$ for $k \in \{1, \ldots, n\}$

2: Sample $(\xi_{k,1}, \xi_{k,2}) \sim \Pi_{\ell_k}^{\kappa_1,\kappa_2}$ for $k \in \{1, \ldots, n\}$ and $(\underline{\xi}_1, \underline{\xi}_2) \sim \underline{\Pi}_{L_n}^{\kappa_1,\kappa_2}$ ▷ Algs 2&3

3: Put $\underline{\chi}_{n,T}^{(\kappa_i)} = \underline{\xi}_i + \sum_{k=1}^n \left(\xi_{k,i}, \min\{\xi_{k,i}, 0\}, \ell_k \cdot \mathbb{1}_{\{\xi_{k,i} \leq 0\}}\right)$ for $i \in \{1, 2\}$

4: **return** $\left(\underline{\chi}_{n,T}^{(\kappa_1)}, \underline{\chi}_{n,T}^{(\kappa_2)}\right)$

By SB representation (4.2), the law $\underline{\Pi}_{n,T}^{\kappa_1,\kappa_2}$ is indeed a coupling of the vectors $\underline{\chi}_T^{(\kappa_1)}$ and $\underline{\chi}_T^{(\kappa_2)}$ for any $n \in \mathbb{N} \cup \{0\}$. Note that if n equals zero, the set $\{1, \ldots, n\}$ in lines 1 and 2 of the algorithm is empty and the laws $\underline{\Pi}_{0,T}^{\kappa_1,\kappa_2}$ and $\underline{\Pi}_T^{\kappa_1,\kappa_2}$ coincide, implying that SBG-Alg may be viewed as a generalisation of Algorithm 3. The main advantage of SBG-Alg over Algorithm 3 is that it samples n increments of the Gaussian approximation over the interval $[L_n, T]$ using the fast Algorithm 2, with the "error term" contribution $\underline{\xi}_i$ being geometrically small.

The computational complexity of SBG-Alg and Algorithms 2 & 3 is simple to analyse. Assume throughout that all elementary mathematical operations (addition, multiplication, exponentiation, etc.), as well as the evaluation of $\overline{\nu}(\kappa)$ and $\overline{\sigma}_\kappa^2$ for all $\kappa \in (0, 1]$ have constant computational cost. Moreover, assume that the simulation of any of the following random variables has constant expected cost: standard normal $N(0,1)$, uniform $\mathrm{U}(0,1)$, Poisson random variable (independently of its mean) and any jump with distribution $\nu|_{\mathbb{R}\setminus(-\kappa,\kappa)}/\overline{\nu}(\kappa)$ (independently of the cutoff level $\kappa \in (0,1]$). Recall that [40, Alg. MAXLOCATION] samples the law $\Phi_t(v,\mu)$ with uniformly bounded expected cost for all values of the parameters $\mu \in \mathbb{R}$, $v > 0$ and $t > 0$. The next statement follows directly from the algorithms.

Corollary 4.13. *Under assumptions above, there exists a positive constant C_1 (resp. C_2; C_3), independent of $\kappa_1, \kappa_2 \in (0,1]$, $n \in \mathbb{N}$ and time horizon $t > 0$, such that the expected computational complexity of Algorithm 2 (resp. Algorithm 3; SBG-Alg) is bounded by $C_1(1+\overline{\nu}(\kappa_2)t)$ (resp. $C_2(1+\overline{\nu}(\kappa_2)t)$; $C_3(n+\overline{\nu}(\kappa_2)t)$).*

Up to a multiplicative constant, Algorithms 2 and 3 have the same expected computational cost. However, Algorithm 3 requires not only additional simulation of jump times of $X^{(\kappa_2)}$ and a sample from $\Phi_t(v,\mu)$ using [40, Alg. MAXLOCATION] between any two consecutive jumps, but also a sequential computation of the output (the "for-loop" in lines 5-15) due to the condition in line 9 of Algorithm 3. This makes it hard to parallelise Algorithm 3. SBG-Alg avoids this issue by using the fast Algorithm 2 over the stick lengths in SB representation (4.2) and calling Algorithm 3 only over the short time interval $[0, L_n]$, during which very few (if any) jumps of $X^{(\kappa_2)}$ occur. Moreover, SBG-Alg consists of several conditionally independent evaluations of Algorithm 2, which is paralellizable, leading to additional numerical benefits (see §4.5.2 below).

§4.4.2 Complexity of the MC/MLMC estimator based on SBG-Alg

This subsection gives an overview of the bounds on the computational complexity of the MC and MLMC estimators defined respectively in (4.52) and (4.53) of §4.6.5 below. Corollary 4.26 (for MC) and Theorem 4.29 (for MLMC) in §4.6.5 give the full analysis.

Assume (O-δ) holds with some $\delta \in (0,2]$ throughout the subsection. As discussed in §4.3.1 above, we take δ as large as possible. In particular, if $\sigma \neq 0$ then $\delta = 2$. Let $q \in (0,2]$ be as in (4.7) and thus $q \geq \delta$ if $\sigma = 0$. We take q as small as possible. For processes used in practice with $\sigma = 0$, we may typically take $\delta = q = \beta$, where β is the BG index defined in (4.6). Assumption $(H\tau)$, required for the analysis of the class BT_2 in (4.21) of discontinuous functions of $\underline{\tau}_T(X)$, holds with $\gamma = 1$ as (O-δ) is satisfied (see the discussion following Proposition 4.9 above). When analysing the class of discontinuous functions BT_1 in (4.19), assume (H) holds throughout with some $\gamma > 0$.

Monte Carlo.

An MC estimator is L^2-accurate at level $\epsilon > 0$, if its bias is smaller than $\epsilon/\sqrt{2}$ and the number N of independent samples is proportional to ϵ^{-2}, see §3.5.1. The following table contains a summary of the values κ, as a function of ϵ, such that the bias of the estimator in (4.52) is at most $\epsilon/\sqrt{2}$, and the associated Monte Carlo cost $\mathcal{C}_{\text{MC}}(\epsilon)$ (up to a constant) for various classes of functions of $\underline{\chi}_T$ analysed in §4.3.2 (see also Corollary 4.26 below for full details).

Family of functions f	Case	κ	$\epsilon^2 \cdot \mathcal{C}_{\text{MC}}(\epsilon)$
Lip in $(X_T, \underline{X}_T)$	$\sigma \neq 0$	$-1/(3-q)$	$q/(3-q)$
locLip in $(X_T, \underline{X}_T)$	$\sigma \neq 0$	$-2/(4-q)$	$2q/(4-q)$
Lip $\cup$ locLip in $(X_T, \underline{X}_T)$	$\sigma = 0$	ϵ	q
BT_1 defined in (4.19)	$\sigma \neq 0$	$-\min\{\frac{3}{4-q}, \frac{2}{3-q}\}$	$\min\{\frac{3q}{4-q}, \frac{2q}{3-q}\}$
	$\sigma = 0$	$-\frac{1}{2} - \frac{1}{\gamma}$	$q(\frac{1}{2} + \frac{1}{\gamma})$
Lip in $\underline{\tau}_T(X)$	$\sigma \neq 0$	$-1/(3-q)$	$q/(3-q)$
	$\delta \in (0,2) \setminus \{\frac{2}{3}\}$	$-\min\{\frac{2}{\delta}, \max\{\frac{3}{2}, \frac{1}{\delta}\}\}$	$q\min\{\frac{2}{\delta}, \max\{\frac{3}{2}, \frac{1}{\delta}\}\}$
	$\delta = \frac{2}{3}$	$-3/2$	$3q/2$
BT_2 defined in (4.21)	$\sigma \neq 0$	$-2/(3-q)$	$2q/(3-q)$
	$\delta \in (0,2) \setminus \{\frac{2}{3}\}$	$-\min\{\frac{4}{\delta}, \max\{3, \frac{2}{\delta}\}\}$	$q\min\{\frac{4}{\delta}, \max\{3, \frac{2}{\delta}\}\}$
	$\delta = \frac{2}{3}$	-3	$3q$

Table 4.2: The table presents the power of ϵ^{-1} in the asymptotic behaviour of the level κ and the complexity $\mathcal{C}_{\text{MC}}(\epsilon)$ as $\epsilon \to 0$ for the MC estimator in (4.52).

The number of sticks $n \in \mathbb{N} \cup \{0\}$ in SBG-Alg does not affect the law of $\underline{\chi}_T^{(\kappa)}$. It only impacts the MC estimator in (4.52) through numerical stability and the reduction of simulation cost. It is hard to determine the optimal choice for n. Clearly, the choice $n = 0$ (i.e. Algorithm 3) is not a good one as discussed in §4.4.1 above. A balance needs to be struck between (i) having a vanishingly small number of jumps in the time interval $[0, L_n]$, so that Algorithm 3 behaves in a numerically stable way, and (ii) not having too many sticks so that line 2 of SBG-Alg does not execute redundant computation of many geometrically small increments of $X^{(\kappa)}$, which are not detected in the final output. A good rule of thumb is $n = n_0 + \lceil \log^2(1 + \overline{\nu}(\kappa)T) \rceil$, where $\lceil x \rceil = \inf\{j \in \mathbb{Z} : j \geq x\}$, $x \in \mathbb{R}$, and the initial value n_0 is chosen so that some sticks are present if for large κ the total expected number of jumps $\overline{\nu}(\kappa)T$ is small (e.g. $n_0 = 5$ works well in §4.5.2 for jump diffusions with low activity, see Figures 4.6 and 4.5), ensuring that the expected number of jumps in $[0, L_n]$ vanishes as ϵ (and hence κ) tends to zero.

Multilevel Monte Carlo.

The MLMC estimator in (4.53) is based on the coupling in SBG-Alg for consecutive levels of a geometrically decaying sequence $(\kappa_j)_{j\in\mathbb{N}}$ and an increasing sequence of the numbers of sticks $(n_j)_{j\in\mathbb{N}}$. Table 4.3 summarises the resulting MLMC complexity up to logarithmic factors, with full results available in Theorem 4.29 below.

There are two key ingredients in the proof of Theorem 4.29: (I) the bounds in Theorem 4.22 on the L^2-distance (i.e. the level variance, see §3.5.2) between the functions of the marginals of the coupling $\underline{\Pi}_{n_j,T}^{\kappa_j,\kappa_{j+1}}$ constructed by SBG-Alg; (II) the bounds on the bias of various functions in §4.3 above. The number of levels m in the MLMC estimator in (4.53) is chosen to ensure that its bias, equal to the bias of $\underline{\chi}_T^{(\kappa_m)}$ at the top cutoff level κ_m, is bounded by $\epsilon/\sqrt{2}$. Thus, the value of m can be expressed in terms of ϵ using Table 4.2 and the explicit formula for the cutoff κ_j, given in the caption of Table 4.3. The formula for κ_j at level j in the MLMC estimator in (4.53) is established in the proof of Theorem 4.29 by minimising the multiplicative constant in the computational complexity $\mathcal{C}_{\mathrm{ML}}(\epsilon)$ over all possible rates of the geometric decay of the sequence $(\kappa_j)_{j\in\mathbb{N}}$.

We stress that the analysis of the level variances for the various payoff functions of the coupling $\underline{\Pi}_{n_j,T}^{\kappa_j,\kappa_{j+1}}$ in Theorem 4.22 is carried out directly for locally Lipschitz payoffs, see Propositions 4.23. However, in the case of the discontinuous payoffs in BT_1 (see (4.19)) and BT_2 (see (4.21)), the analysis requires a certain regularity (uniformly in the cutoff levels) of the coupling $(\underline{\chi}_T^{(\kappa_j)}, \underline{\chi}_T^{(\kappa_{j+1})})$. This leads to a construction of a further coupling $(\underline{\chi}_T^{(\kappa_j)}, \underline{\chi}_T^{(\kappa_{j+1})}, \underline{\chi}_T)$ where the components of the pair $(\underline{\chi}_T^{(\kappa_j)}, \underline{\chi}_T^{(\kappa_{j+1})})$ can be compared to the limiting object $\underline{\chi}_T$, which can be shown to possess the necessary regularity (see Proposition 4.25 below for details).

§4.5 Numerical examples

In this section we study numerically the performance of SBG-Alg. All the results are based on the code available in repository [51]. In §4.5.1 we apply SBG-Alg to two families of Lévy models (tempered stable and Watanabe processes) and verify numerically the decay of the bias (established in §4.3.2 above) and level variance (see Theorem 4.22 below) of the Gaussian approximations. In §4.5.2 we study numerically the cost reduction of SBG-Alg, when compared to Algorithm 3, for the simulation of the vector $\underline{\chi}_T^{(\kappa)}$.

Family of functions f	Case	a	The power of ϵ^{-1} in $\epsilon^2 \cdot \mathcal{C}_{\mathrm{ML}}(\epsilon)$
Lip in $(X_T, \underline{X}_T)$	$\sigma \neq 0$	$2(q-1)$	$2(q-1)^+/(3-q)$
locLip in $(X_T, \underline{X}_T)$	$\sigma \neq 0$	$2(q-1)$	$4(q-1)^+/(4-q)$
Lip $\cup$ locLip in $(X_T, \underline{X}_T)$	$\sigma = 0$	$2(q-1)$	$2(q-1)^+$
BT_1 defined in (4.19)	$\sigma \neq 0$	$2(2q-1)/3$	$(2q-1)^+ \min\{\frac{2}{4-q}, \frac{4}{9-3q}\}$
	$\sigma = 0$	$2(q(1+\gamma)-\gamma)/(2+\gamma)$	$(q(1+1/\gamma)-1)^+$
Lip in $\underline{\tau}_T(X)$	$\sigma \neq 0$	$\frac{5}{4}q - \frac{1}{2}$	$(\frac{5}{4}q - \frac{1}{2})^+$
	$\sigma = 0$	$q - (1-\frac{q}{2})\min\{\frac{1}{2}, \frac{2\delta}{2-\delta}\}$	$\frac{(2q-(2-q)\min\{1, 4\delta/(2-\delta)\})^+}{\max\{\delta, \min\{4/3, 2\delta\}\}}$
BT_2 defined in (4.21)	$\sigma \neq 0$	$\frac{9}{8}q - \frac{1}{4}$	$(\frac{9}{4}q - \frac{1}{2})^+$
	$\sigma = 0$	$q - (1-\frac{q}{2})\min\{\frac{1}{4}, \frac{\delta}{2-\delta}\}$	$\frac{(2q-(2-q)\min\{\frac{1}{2}, 2\delta/(2-\delta)\})^+}{\max\{\delta/2, \min\{2/3, \delta\}\}}$

Table 4.3: The table presents the power of ϵ^{-1} in $\epsilon^2 \cdot \mathcal{C}_{\mathrm{ML}}(\epsilon)$ as $\epsilon \to 0$, neglecting only the logarithmic factors (see Theorem 4.29 below for the complete result). Parameter a in the table determines the decreasing sequence of cutoff levels $(\kappa_j)_{j\in\mathbb{N}}$ as follows: $\kappa_j = (1+|a|/q)^{-2(j-1)/|a|}$ if $a \neq 0$ and $\kappa_j = \exp(-(2/q)(j-1))$ otherwise. The corresponding increasing number of sticks n_j in the definition of the law $\underline{\Pi}_{n_j,T}^{\kappa_j,\kappa_{j+1}}$ can be taken to grow asymptotically as $\log^2(1+\overline{\nu}(\kappa_j)T)$ for large j, see Theorem 4.29.

§4.5.1 Numerical performance of SBG-Alg for tempered stable and Watanabe processes

To illustrate numerically our results, we consider two classes of exponential Lévy models $S = S_0 \mathrm{e}^X$. The first is the tempered stable class, containing the CGMY (or KoBoL) model, a widely used process for modeling risky assets in financial mathematics (see e.g. [36] and the references therein), which satisfies the regularity assumptions from §4.3.2 above. The second is the Watanabe class, which has diffuse but singular transition laws [91, Thm 27.19], making it a good candidate to stress test our results.

We numerically study the decay of the bias and level variance of the MLMC estimator in (4.53) for the prices of a lookback put $\mathbb{E}[\overline{S}_T - S_T]$ and up-and-out call $\mathbb{E}[(S_T - K)^+ \mathbb{1}_{\{\overline{S}_T \le M\}}]$ as well as the values of the ulcer index (UI), given by $100\mathbb{E}[(S_T/\overline{S}_T - 1)^2]^{1/2}$ [44, 77], and a modified ulcer index (MUI), given by $100\mathbb{E}[(S_T/\overline{S}_T - 1)^2 \mathbb{1}_{\{\overline{\tau}_T(S) < T/2\}}]^{1/2}$. The first three quantities are commonplace in applications, see [36, 44]. The MUI refines the UI by incorporating the information on the drawdown duration, weights trends more heavily than short-time fluctuations.

In §4.5.1 and §4.5.1 we use $N = 10^5$ independent samples to estimate the means and variances of the variables D_j^1 in (4.53) (with $\underline{\chi}_T^{(\kappa_j)}$ substituted by $\overline{\chi}_T^{(\kappa_j)}$), where $\kappa_j = \mathrm{e}^{-r(j-1)}$ and $n_j = \lceil \max\{j, \log^2(1+\overline{\nu}(\kappa_{j+1}))\} \rceil$, $j \in \mathbb{N}$, discussed in §4.6.5.

Tempered stable.

The characteristic triplet (σ^2, ν, b) of the tempered stable Lévy process X is given by $\sigma = 0$, drift $b \in \mathbb{R}$ and Lévy measure $\nu(\mathrm{d}x) = |x|^{-1-\alpha_{\mathrm{sgn}(x)}} c_{\mathrm{sgn}(x)} \mathrm{e}^{-\lambda_{\mathrm{sgn}(x)}|x|} \mathrm{d}x$, where $\alpha_\pm \in [0,2)$, $c_\pm \geq 0$ and $\lambda_\pm > 0$, cf. (4.3). Exact simulation of increments is currently out of reach if either $\alpha_+ > 1$ or $\alpha_- > 1$ (see e.g. [58]) and requires the Gaussian approximation.

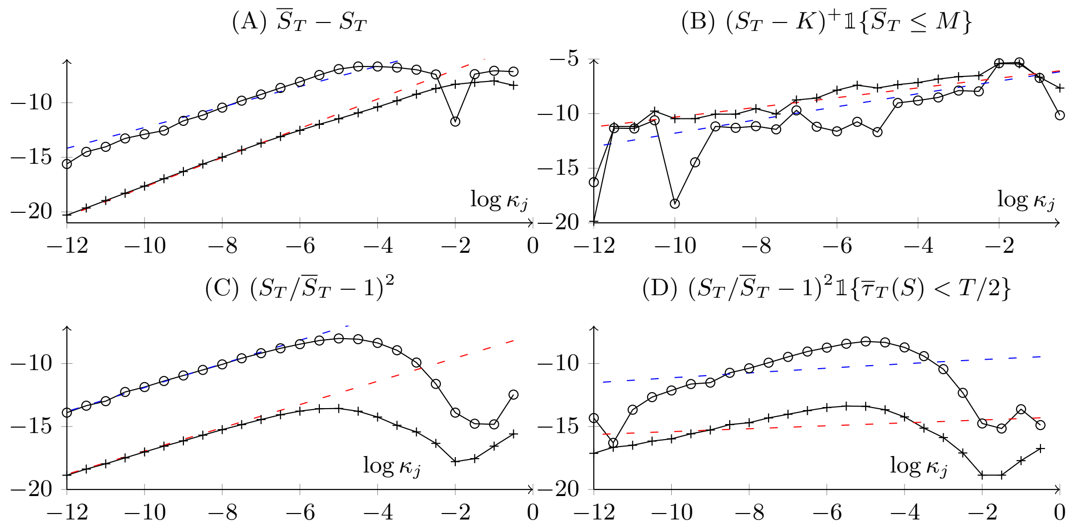

Figure 4.3: Gaussian approximation of a tempered stable process: log-log plot of the bias and level variance for various payoffs as a function of $\log \kappa_j$. Circle (∘) and plus (+) correspond to $\log |\mathbb{E}[D_j^1]|$ and $\log \mathbb{V}[D_j^1]$, respectively, where D_j^1 is given in (4.53) with $\kappa_j = \exp(-r(j-1))$ for $r = 1/2$. The dashed lines in all the graphs plot the rates of the theoretical bounds in §4.3.2 (blue for the bias) and Theorem 4.22 (red for level variances). In plots (A)–(D) the initial value of the risky asset is normalised to $S_0 = 1$ and the time horizon is set to $T = 1/6$. In plot (B) we set $K = 1$ and $M = 1.2$. The model parameters are given in Table 4.4 below.

Parameter set	b	α_+	α_-	c_+	c_-	λ_+	λ_-	Graphs
1	0	.66	.66	.1305	.0615	6.5022	3.0888	(A) and (B)
2	.1274	1.0781	1.0781	.41077	.41077	49.663	59.078	(C) and (D)

Table 4.4: The parameters used for Figure 4.3. The first set of parameters corresponds to the risk-neutral calibration to vanilla options on the USD/JPY exchange rate, see [3, Table 3]. The second set is the maximum likelihood estimate based on the real-world S&P stock prices, see [63, Table 1].

In Figure 4.3, our bounds are close to the exhibited numerical behaviour for continuous payoff functions. However, in the discontinuous case, $\overline{\chi}_T^{(\kappa_j)}$ appears to be much closer to $\overline{\chi}_T$ (resp. $\overline{\chi}_T^{(\kappa_{j+1})}$), than predicted by Propositions 4.7 & 4.9 (resp. Theorem 4.22(b)&(d)).

Watanabe model.

The characteristic triplet (σ, ν, b) of the Watanabe process is given by $\sigma = 0$, the Lévy measure ν equals $\sum_{n\in\mathbb{N}} c_+\delta_{a^{-n}} + c_-\delta_{-a^{-n}}$, where $a \in \mathbb{N}\setminus\{1\}$ and δ_x is the Dirac measure at x, and the drift $b \in \mathbb{R}$ is arbitrary. The increments of the Watanabe process are diffuse but have no density (see [91, Thm 27.19]). Since the process has very little jump activity, the bound in Proposition 4.9 (see also (4.13)) is non-vanishing and the bounds in Theorem 4.22(c) & (d) are not applicable, meaning that we have no theoretical control on the approximation of $\overline{\tau}_T(S)$. This is not surprising as such acute lack of jump activity makes the Gaussian approximation unsuitable (cf. [6, Prop. 2.2]).

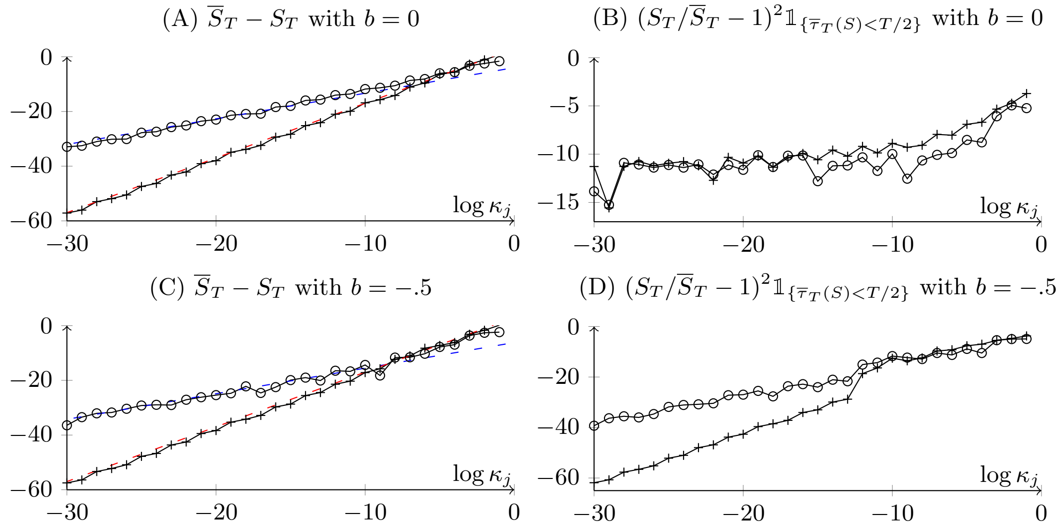

Figure 4.4: Gaussian approximation of a Watanabe process: log-log plot of the bias and level variance for various payoffs as a function of $\log\kappa_j$. Circle ($\circ$) and plus ($+$) correspond to $\log|\mathbb{E}[D_j^1]|$ and $\log\mathbb{V}[D_j^1]$, respectively, where D_j^1 is given in (4.53) with $\kappa_j = \exp(-r(j-1))$ for $r = 1$. The dashed lines in graphs (A) & (C) plot the rates of the theoretical bounds in §4.3.2 (blue for the bias) and Theorem 4.22 (red for level variances). In plots (A)–(D) the initial value of the risky asset is normalised to $S_0 = 1$ and the time horizon is set to $T = 1$. The model parameters are given by $a = 2$, $c_+ = c_- = 1$.

The pictures in Figure 4.4 (A) & (C) suggest that our bounds on the bias and level variance in §4.3.2 and Theorem 4.22 are robust for continuous payoff functions even if the underlying Lévy process has no transition densities. There are no dashed lines in Figure 4.4 (B) & (D) as there are no results for discontinuous functions of $\overline{\tau}_T(S)$ in this case. In fact, Figure 4.4(B) suggests that the decay rate of the bias and level variance for functions of $\overline{\tau}_T(S)$ can be arbitrarily slow if the process does not have sufficient activity. Figure 4.4(D), however, suggests that this decay is still fast if the underlying finite variation process X has a nonzero natural drift (see also Remark 4.5).

§4.5.2 The cost reduction of SBG-Alg over Algorithm 3

Recall that Algorithm 3 and SBG-Alg both draw exact samples of a Gaussian approximation $\underline{\chi}_T^{(\kappa)}$. However, in practice, SBG-Alg may be many times faster than Algorithm 3: Figure 4.5 plots the speedup factor in the case of a tempered stable process, defined in §4.5.1 above, as a function of κ. In conclusion, one should use SBG-Alg instead of Algorithm 3 for the MC estimator in (4.52) (recall that Algorithm 3 is not suitable for the MLMC estimator, as discussed in §4.4.1).

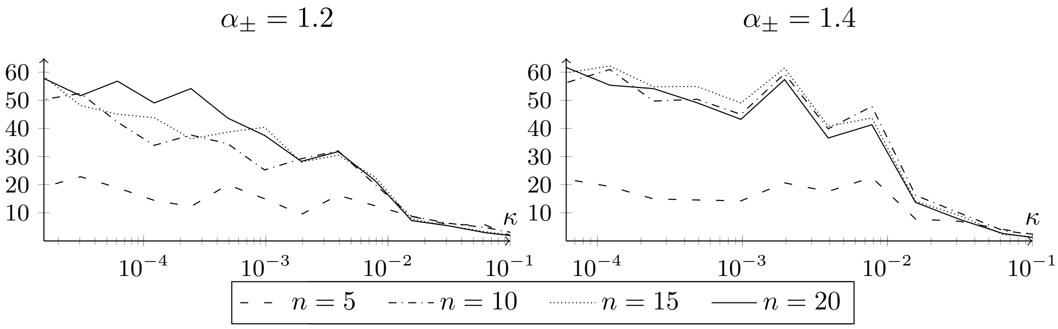

Figure 4.5: The pictures show the ratio of the cost of Algorithm 3 over the cost of SBG-Alg (both in seconds) for the Gaussian approximations of a tempered stable process as a function of the cutoff level κ. The parameters used are $\lambda_\pm = 5$, $c_\pm = 2$. The number of sticks n in SBG-Alg varies between 5 and 20. The ratio for $n = 20$ is 57.8 (resp. 61.7) in the case $\alpha_\pm = 1.2$ (resp. $\alpha_\pm = 1.4$) for $\kappa = 2^{-16}$ (resp. $\kappa = 2^{-14}$).

If the Lévy process X is a jump diffusion, i.e. $\nu(\mathbb{R} \setminus \{0\}) < \infty$, we may apply Algorithms 2 & 3 and SBG-Alg with $\kappa_1 = \kappa_2 = 0$. In that case SBG-Alg still outperforms Algorithm 3 by a constant factor, with computational benefits being more pronounced when the total expected number of jumps $\lambda = \nu(\mathbb{R} \setminus \{0\})T$ is large. The cost reduction is most drastic when λ is large, but the improvement is already significant for $\lambda = 2$.

§4.5.3 Estimating the Greeks: Delta and Gamma for barrier options via Monte Carlo

A fundamental problem in mathematical finance is to compute the sensitivity of the price of a derivative security to the various underlying parameters in order to construct appropriate hedging strategies. These sensitivities are known as the *Greeks* and are in practice given by the partial derivatives of the option price $e^{-rT}\mathbb{E}[P]$ (where r is the discount rate over the time horizon T and P is a random payoff).

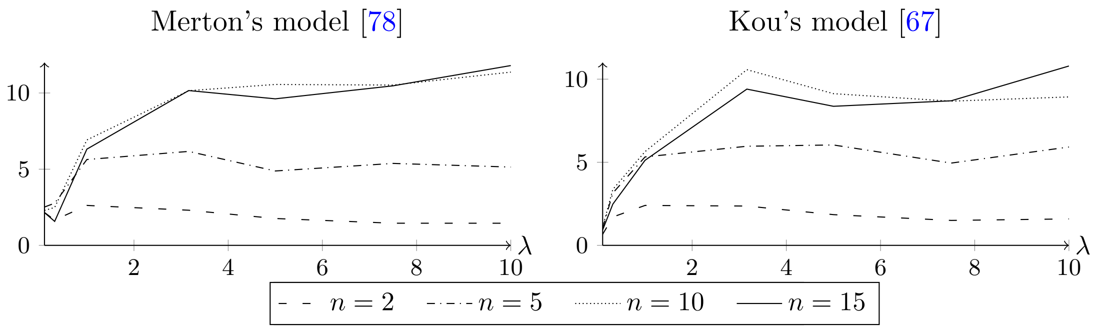

Figure 4.6: The pictures show, for multiple number of sticks n, the ratio of the cost of Algorithm 3 over the cost of SBG-Alg (both in seconds) for jump diffusions as a function of the mean number of jumps $\lambda = \nu(\mathbb{R} \setminus \{0\})T$. The ratio for $n = 15$ is 11.8 (resp. 10.8) in Merton's (resp. Kou's) model when $\lambda = 10$.

The most common of the Greeks are Delta and Gamma, given by the first and second derivatives of the price $\mathrm{e}^{-rT}\mathbb{E}[P]$ with respect to the spot S_0.

If the risk-neutral dynamics of the risky asset is described by an exponential Lévy model $S = S_0\mathrm{e}^X$, SBG-Alg provides a simple procedure for the Monte Carlo estimation of Delta and Gamma for any payoff $P = g(\overline{\chi}_T)$ (recall that $f'(x)$ and $f''(x)$ of a function $f(x)$ are approximated by $(f(x+h)-f(x-h))/(2h)$ and $(f(x+h)-2f(x)+f(x-h))/h^2$, respectively, for a small $h > 0$). This approach, widely used in practice, requires the evaluation of $\mathrm{e}^{-rT}\mathbb{E}[g(\overline{\chi}_T)]$ by SBG-Alg on a grid of current spot prices S_0, where the simulated stick-breaking sequence and the corresponding increments of $X^{(\kappa)}$ can be reused over the grid points of S_0.

In order to test numerically the performance of SBG-Alg in this context, we compute Delta and Gamma of an up-and-out digital option with barrier M and payoff $g(\overline{\chi}_T) = \mathbb{1}\{\overline{S}_T \leq M\} = \mathbb{1}\{\overline{X}_T \leq \log(M/S_0)\}$ under an exponential Lévy model with numerically accassible Delta and Gamma. Let X be an α-stable process of infinite variation without positive jumps. Then $\overline{X}_T$ has the same law as X_T conditioned to be positive and Delta and Gamma are thus equal to the first and second derivatives of $s \mapsto \mathrm{e}^{-rT}\mathbb{P}(X_T \leq \log(M/s))/\mathbb{P}(X_T > 0)$, which can be numerically evaluated via definite integrals and power series [97, Ch. 4].

The parameters were chosen as follows: X has unit scale (in Zolotarev's (C) parametrisation) and $\alpha = 1.5$, while the market data is $r = 0.05$, $T = 1$, $M = 1$ and $S_0 \in [1/2, 1)$. The cutoff is set at $\kappa = 0.1$ and the grid spacing at $h = 0.01$. We used $n = 20$ sticks and $N = 10^7$ samples. This resulted in a total simulation time [1] of 1 minute. The estimation of Delta and Gamma is accurate and numerically stable

[1]We used an HP Pavilion laptop 15-cw0xxx containing an AMD Ryzen 5 2500U with Radeon Vega Mobile and 12 GB of RAM

(see Figure 4.7). Surprisingly the error in Delta remains bounded all the way to the barrier M.

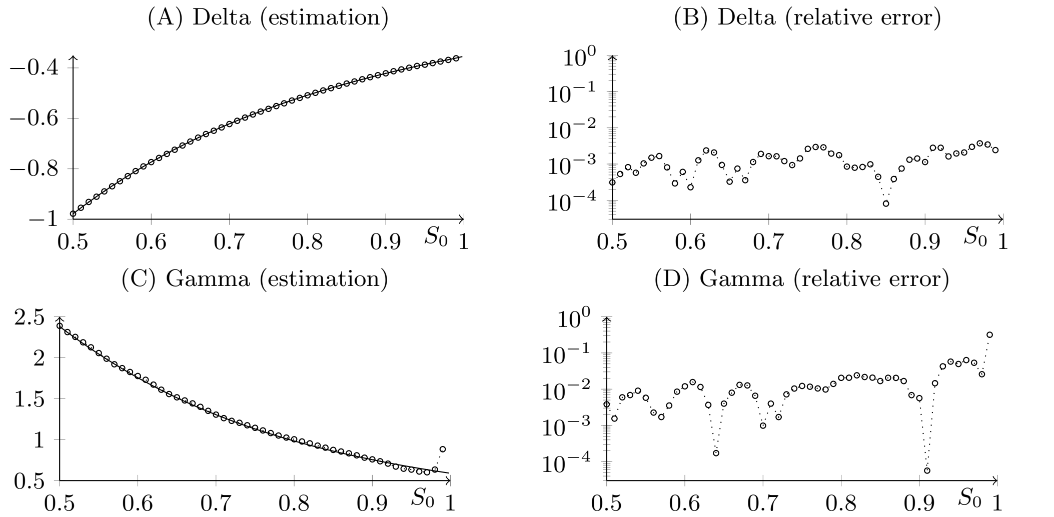

Figure 4.7: Monte Carlo estimation of Delta and Gamma for the up-and-out digital option with payoff $\mathbb{1}\{\overline{S}_T \leq M\}$. Dotted lines with ($\circ$) correspond to the output based on SBG-Alg. Subfigures (A) and (C) contain a solid line corresponding to the true values of Delta and Gamma.

§4.6 Proofs and technical results of Chapter 4

In the remainder of the chapter we use the notation $\underline{\tau}_t = \underline{\tau}_t(X)$, $\underline{\tau}_t^{(\kappa)} = \underline{\tau}_t(X^{(\kappa)})$ for all $t > 0$.

§4.6.1 Proof of Theorems 4.1 and 4.2

In this subsection we establish bounds on the Wasserstein and Kolmogorov distances between the increment X_t and its Gaussian approximation $X_t^{(\kappa)}$ in (4.5).

Proof of Theorem 4.1. Recall the Lévy-Itô decomposition of X at level κ in (4.4) and the martingale $M^{(\kappa)} = \sigma B + J^{1,\kappa}$. Set $Z = X - M^{(\kappa)}$ and note $X^{(\kappa)} = Z + \sqrt{\overline{\sigma}_\kappa^2 + \sigma^2} W$, where W is a standard Brownian motion in (4.5), independent of Z. Hence any coupling $(W_t, M_t^{(\kappa)})$ yields a coupling of $(X_t, X_t^{(\kappa)})$ satisfying $\mathbb{E}[|X_t - X_t^{(\kappa)}|^p] = \mathbb{E}[|M_t^{(\kappa)} - \sqrt{\overline{\sigma}_\kappa^2 + \sigma^2} W_t|^p]$. Setting $W = B$, which amounts to the independence coupling $(W, J^{1,\kappa})$, and applying Jensen's inequality for $p \in [1, 2]$

yields

$$\mathbb{E}[|X_t - X_t^{(\kappa)}|^p]^{2/p} \le \mathbb{E}\big[\big|J_t^{1,\kappa} - \big(\sqrt{\overline{\sigma}_\kappa^2+\sigma^2}-\sigma\big)W_t\big|^2\big]$$
$$= \mathbb{E}[||J_t^{1,\kappa}|^2] + \big(\sqrt{\overline{\sigma}_\kappa^2+\sigma^2}-\sigma\big)^2 t \le 2t\overline{\sigma}_\kappa^2.$$

For $m\in\mathbb{N}$ we have $M_t^{(\kappa)} \overset{d}{=} \sum_{i=1}^m \xi_i$, where $\xi_1,\dots,\xi_m$ are iid with $\xi_1 \overset{d}{=} M_{t/m}^{(\kappa)}$. Hence [84, Thm 16] and [89, Thm 4.1] imply the existence of universal constants K_p, $p\in[1,2]$, with $K_1 = 1/2$, satisfying

$$\mathcal{W}_p^p\big(M_t^{(\kappa)}, \sqrt{\overline{\sigma}_\kappa^2+\sigma^2}W_t\big) \le K_p^p \tfrac{[t(\overline{\sigma}_\kappa^2+\sigma^2)]^{p/2}\mathbb{E}[|\xi_1|^{p+2}]}{m^{p/2}\mathbb{E}[\xi_1^2]^{(p+2)/2}} = K_p^p \tfrac{(m/t)\mathbb{E}[|M_{t/m}^{(\kappa)}|^{p+2}]}{\overline{\sigma}_\kappa^2+\sigma^2}, \quad m\in\mathbb{N}.$$

According to [46, Thm 1.1], the limit as $m\to\infty$ of the right-hand side of the display above equals $K_p^p\int_{(-\kappa,\kappa)}|x|^{p+2}\nu(\mathrm{d}x)/(\overline{\sigma}_\kappa^2+\sigma^2) \le K_p^p\kappa^p\varphi_\kappa^2$, implying the claim in the theorem. □

Proof of Theorem 4.2. (a) Define $d_\kappa = \sup_{x\in\mathbb{R}}|\mathbb{P}(M_t^{(\kappa)}\le x) - \mathbb{P}(\sqrt{\overline{\sigma}_\kappa^2+\sigma^2}W_t \le x)|$ and note that

$$|\mathbb{P}(X_t\le x) - \mathbb{P}(X_t^{(\kappa)}\le x)| = \big|\mathbb{E}\big[\mathbb{P}(M_t^{(\kappa)}\le x - Z_t|Z_t) - \mathbb{P}\big(\sqrt{\overline{\sigma}_\kappa^2+\sigma^2}W_t\le x-Z_t\big|Z_t\big)\big]\big|$$

is bounded by d_κ, where the processes Z and $M^{(\kappa)}$ are as in the proof of Theorem 4.1. Since $M^{(\kappa)}$ is a Lévy process, for any $m\in\mathbb{N}$ we have $M_t^{(\kappa)} \overset{d}{=} \sum_{i=1}^m\xi_i$, where $\xi_1,\dots,\xi_m$ are iid with $\xi_1 \overset{d}{=} M_{t/m}^{(\kappa)}$. By the Berry-Esseen inequality [66, Thm 1], there exists a constant $C_{\mathrm{BE}}\in(0,\frac{1}{2})$ such that

$$d_\kappa \le \frac{C_{\mathrm{BE}}\mathbb{E}[|\xi_1|^3]}{\sqrt{m}\mathbb{E}[\xi_1^2]^{3/2}} = \frac{C_{\mathrm{BE}}t}{m\sqrt{m}}\cdot\frac{(m/t)\mathbb{E}[|M_{t/m}^{(\kappa)}|^3]}{(t/m)^{3/2}(\overline{\sigma}_\kappa^2+\sigma^2)^{3/2}} = C_{\mathrm{BE}}\frac{(m/t)\mathbb{E}[|M_{t/m}^{(\kappa)}|^3]}{\sqrt{t}(\overline{\sigma}_\kappa^2+\sigma^2)^{3/2}}, \quad m\in\mathbb{N}.$$

According to [46, Thm 1.1], the limit as $m\to\infty$ of the right-hand side of the display above equals $C_{\mathrm{BE}}\int_{(-\kappa,\kappa)}|x|^3\nu(\mathrm{d}x)/(\sqrt{t}(\overline{\sigma}_\kappa^2+\sigma^2)^{3/2}) \le C_{\mathrm{BE}}(\kappa/\overline{\sigma}_\kappa)\varphi_\kappa^3/\sqrt{t}$, implying (a).

(b) By [86, Thm 3.1(a)], X_t has a smooth density f_t and, given $T>0$, the constant $C' = \sup_{(t,x)\in(0,T]\times\mathbb{R}} t^{1/\delta}f_t(x)$ is finite. Applying (4.8) and (4.33) in Lemma 4.18 with $p=2$ gives (4.10). □

§4.6.2 Proof of Theorem 4.3

We recall an elementary result for stick-breaking processes.

Lemma 4.14. *Let* $(\varpi_n)_{n\in\mathbb{N}}$ *be a stick-breaking process on* $[0,1]$ *based on the law* $\mathrm{U}(0,1)$*. For any measurable function* $\phi \geq 0$*, we have*

$$\sum_{n\in\mathbb{N}} \mathbb{E}[\phi(\varpi_n)] = \int_0^1 \frac{\phi(x)}{x}\mathrm{d}x.$$

In particular, for any $a_1, a_2 > 0$ *and* $b_1 < b_2$ *with* $b_2 > 0$*, we have*

$$\sum_{n\in\mathbb{N}} \mathbb{E}[\min\{a_1\varpi_n^{b_1}, a_2\varpi_n^{b_2}\}]$$
$$= \begin{cases} \frac{a_2}{b_2}\min\left\{1, \frac{a_1}{a_2}\right\}^{b_2/(b_2-b_1)} + \frac{a_1}{b_1}\left(1 - \min\left\{1, \frac{a_1}{a_2}\right\}^{b_1/(b_2-b_1)}\right), & b_1 \neq 0, \\ b_2^{-1}\min\{a_2, a_1\}\left(1 + \log^+\left(\frac{a_2}{a_1}\right)\right), & b_1 = 0. \end{cases}$$

Proof. The law of $-\log \varpi_n$ is gamma with shape n and scale 1. Applying Fubini's theorem, implies

$$\sum_{n\in\mathbb{N}} \mathbb{E}[\phi(\varpi_n)] = \sum_{n\in\mathbb{N}} \int_0^\infty \frac{x^{n-1}}{(n-1)!}\mathrm{e}^{-x}\phi(\mathrm{e}^{-x})\mathrm{d}x = \int_0^\infty \phi(\mathrm{e}^{-x})\mathrm{d}x = \int_0^1 \frac{\phi(x)}{x}\mathrm{d}x.$$

The formula for $\phi(x) = \min\{a_1 x^{b_1}, a_2 x^{b_2}\}$ follows by a direct calculation. □

The L^p-Wasserstein distance, defined in above (4.17), satisfies $\mathcal{W}_p^p(\xi, \xi_*) = \int_0^1 |F^{-1}(u) - F_*^{-1}(u)|^p du$, where F^{-1} (resp. F_*^{-1}) is the right inverse of the distribution function F (resp. F_*) of the real-valued random variable ξ (resp. ξ_*) (see [19, Thm 2.10]). Thus the *comonotonic* (or *minimal transport*) coupling, defined by

$$(\xi, \xi_*) = (F^{-1}(U), F_*^{-1}(U)) \qquad \text{for some } U \sim \mathrm{U}(0,1), \tag{4.25}$$

attains the infimum in definition (4.17).

Lemma 4.15. *If the random variables* ξ *and* ξ_* *are comonotonically coupled, then*

$$\mathbb{E}[|\mathbb{1}\{\xi \leq x\} - \mathbb{1}\{\xi_* \leq x\}|] = |\mathbb{E}[\mathbb{1}\{\xi \leq x\} - \mathbb{1}\{\xi_* \leq x\}]| \qquad \text{for any } x \in \mathbb{R}.$$

Proof. Suppose $(\xi, \xi_*) = (F^{-1}(U), F_*^{-1}(U))$ for some $U \sim \mathrm{U}(0,1)$, where F and F_* are the distribution functions of ξ and ξ_*. Suppose $y = F(x) \leq F_*(x) =: y_*$. Since F^{-1} and F_*^{-1} are monotonic functions, it follows that $\mathbb{1}\{\xi \leq x\} - \mathbb{1}\{\xi_* \leq x\} \leq 0$ a.s. since this difference equals -1 or 0 according to $U \in (y, y_*]$ or $U \in (0,1) \setminus (y, y_*]$, respectively. If $y \geq y_*$, we have $\mathbb{1}\{\xi \leq x\} - \mathbb{1}\{\xi_* \leq x\} \geq 0$ a.s. In either case, the result follows. □

For any $t > 0$, let G_t^κ denote the joint law of the comonotonic coupling of X_t and $X_t^{(\kappa)}$ defined in (4.25). Note that a coupling $(X_t, X_t^{(\kappa)})$ with law G_t^κ satisfies the inequality in Theorem 4.1. The following lemma is crucial in the proof of Theorem 4.3.

Lemma 4.16. *Let $\ell = (\ell_n)_{n\in\mathbb{N}}$ be a stick-breaking process on $[0,t]$ and $(\xi_n, \xi_n^{(\kappa)})$, $n \in \mathbb{N}$, a sequence of random vectors that, conditional on ℓ, are independent and satisfy $(\xi_n, \xi_n^{(\kappa)}) \sim G_{\ell_n}^\kappa$ for all $n \in \mathbb{N}$. Then for any $p \in [1,2]$ and $x \in \mathbb{R}$ we have*

$$\begin{aligned} \mathbb{E}\bigg[\bigg(\sum_{n=1}^\infty |\xi_n - \xi_n^{(\kappa)}|\bigg)^p\bigg]^{1/p} &\le \mu_p(\kappa, t) \quad \text{and} \\ \mathbb{E}\bigg[\sum_{n=1}^\infty \ell_n \big|\mathbb{1}\{\xi_n \le x\} - \mathbb{1}\{\xi_n^{(\kappa)} \le x\}\big|\bigg] &\le \mu_0^\tau(\kappa, t), \end{aligned} \tag{4.26}$$

where μ_p and μ_0^τ are as in (4.12) and (4.13), respectively. Moreover, if (O-δ) holds, then for every $T > 0$ there exists a constant $C > 0$, dependent only on (T, δ, σ, ν), such that for all $t \in [0,T]$, $\kappa \in (0,1]$ and $x \in \mathbb{R}$ we have

$$\mathbb{E}\bigg[\sum_{n=1}^\infty \ell_n \big|\mathbb{1}\{\xi_n \le x\} - \mathbb{1}\{\xi_n^{(\kappa)} \le x\}\big|\bigg] \le \mu_\delta^\tau(\kappa, t), \tag{4.27}$$

where μ_δ^τ is defined in (4.15).

Proof. Note that $\mu_p(\kappa,t) = \mu_2(\kappa,t)$ for all $p \in (1,2]$. Hence, by Jensen's inequality, in (4.26) we need only consider $p \in \{1,2\}$. Pick $n \in \mathbb{N}$ and set $\kappa_p = K_p^p \kappa^p \varphi_\kappa^2$, $p \in \{1,2\}$, where K_p and φ_κ are as in the statement of Theorem 4.1. Condition on ℓ_n and apply the bound in (4.8) to obtain

$$\mathbb{E}[|\xi_n^{(\kappa)} - \xi_n|^p | \ell_n] \le \min\big\{2^{p/2}\overline{\sigma}_\kappa^p \ell_n^{p/2}, \kappa_p\big\}, \qquad p \in \{1,2\}. \tag{4.28}$$

An application of (4.28) and Lemma 4.14 yield the first inequality in (4.26) for $p = 1$:

$$\begin{aligned} \sum_{n=1}^\infty \mathbb{E}\big[|\xi_n - \xi_n^{(\kappa)}|\big] &\le \sum_{n=1}^\infty \mathbb{E}[\min\big\{\sqrt{2\ell_n}\overline{\sigma}_\kappa, \kappa_1\big\}] \\ &= 2\min\big\{\sqrt{2t}\overline{\sigma}_\kappa, \kappa_1\big\}\big(1 + \log^+\big(\sqrt{2t}\overline{\sigma}_\kappa/\kappa_1\big)\big). \end{aligned}$$

Consider the case $p = 2$. A simple expansion yields

$$\mathbb{E}\bigg[\bigg(\sum_{n=1}^\infty |\xi_n - \xi_n^{(\kappa)}|\bigg)^2\bigg] = \sum_{n=1}^\infty \mathbb{E}\big[(\xi_n - \xi_n^{(\kappa)})^2\big] + 2\sum_{n=1}^\infty \sum_{m=n+1}^\infty \mathbb{E}\big[|\xi_n - \xi_n^{(\kappa)}||\xi_m - \xi_m^{(\kappa)}|\big].$$

We proceed to bound the two sums. The inequality in (4.28) for $p = 2$ and Lemma 4.14 imply

$$\sum_{n=1}^{\infty} \mathbb{E}\big[\big(\xi_n - \xi_n^{(\kappa)}\big)^2\big] \le \sum_{n=1}^{\infty} \mathbb{E}\big[\min\big\{2\overline{\sigma}_\kappa^2 \ell_n, \kappa_2\big\}\big]$$
$$= \min\big\{2t\overline{\sigma}_\kappa^2, \kappa_2\big\}\big(1 + 2\log^+\big(\sqrt{2t}\overline{\sigma}_\kappa/\sqrt{\kappa_2}\big)\big).$$

Define the σ-algebra $\mathcal{F}_n = \sigma(\ell_1, \dots, \ell_n)$ and use the conditional independence to obtain

$$\mathbb{E}\big[\big|\xi_n - \xi_n^{(\kappa)}\big|\big|\xi_m - \xi_m^{(\kappa)}\big|\big|\mathcal{F}_m\big] \le \min\{\sqrt{2\ell_n}\overline{\sigma}_\kappa, \kappa_1\} \min\{\sqrt{2\ell_m}\overline{\sigma}_\kappa, \kappa_1\}, \qquad n < m.$$

Note that $(\ell_m/L_n)_{m=n+1}^{\infty}$ is a stick-breaking process on $[0,1]$ independent of $\mathcal{F}_n$. Use the tower property and apply (4.8) and Lemma 4.14 to get

$$\sum_{m=n+1}^{\infty} \mathbb{E}\big[\big|\xi_n - \xi_n^{(\kappa)}\big|\big|\xi_m - \xi_m^{(\kappa)}\big|\big|\mathcal{F}_n\big]$$
$$\le \min\{\sqrt{2\ell_n}\overline{\sigma}_\kappa, \kappa_1\} \sum_{m=n+1}^{\infty} \mathbb{E}\big[\min\big\{\sqrt{2\ell_m}\overline{\sigma}_\kappa, \kappa_1\big\}\big|\mathcal{F}_n\big]$$
$$= 2\min\big\{\sqrt{2\ell_n}\overline{\sigma}_\kappa, \kappa_1\big\}\min\big\{\sqrt{2L_n}\overline{\sigma}_\kappa, \kappa_1\big\}\Big(1 + \log^+\Big(\tfrac{\sqrt{2L_n}\overline{\sigma}_\kappa}{\kappa_1}\Big)\Big)$$
$$\le 2\min\big\{2L_{n-1}\overline{\sigma}_\kappa^2, \kappa_1^2\big\}\big(1 + \log^+\big(\sqrt{2t}\overline{\sigma}_\kappa/\kappa_1\big)\big),$$

where $\max\{L_n, \ell_n\} \le L_{n-1} \le t$ is used in the last step. Since $\ell_n \overset{d}{=} L_n$, $n \in \mathbb{N}$, Lemma 4.14 yields

$$2\sum_{n=1}^{\infty}\sum_{m=n+1}^{\infty} \mathbb{E}\big[\big|\xi_n - \xi_n^{(\kappa)}\big|\big|\xi_m - \xi_m^{(\kappa)}\big|\big]$$
$$\le 4\sum_{n=1}^{\infty} \mathbb{E}\left[\min\big\{2L_{n-1}\overline{\sigma}_\kappa^2, \kappa_1^2\big\}\right]\big(1 + \log^+\big(\sqrt{2t}\overline{\sigma}_\kappa/\kappa_1\big)\big)$$
$$= 2\mu_1(\kappa, t)^2.$$

Putting everything together yields the first inequality in (4.26) for $p = 2$.

Next we prove the second inequality in (4.26). By Lemma 4.15, we have

$$\mathbb{E}[|\mathbb{1}\{\xi_n \le x\} - \mathbb{1}\{\xi_n^{(\kappa)} \le x\}||\ell_n] = \big|\mathbb{P}(X_{\ell_n} \le x|\ell_n) - \mathbb{P}\big(X_{\ell_n}^{(\kappa)} \le x\big|\ell_n\big)\big|. \qquad (4.29)$$

Applying (4.9) in Theorem 4.2(a) implies $\ell_n\big|\mathbb{P}(X_{\ell_n} \le x|\ell_n) - \mathbb{P}\big(X_{\ell_n}^{(\kappa)} \le x\big|\ell_n\big)\big| \le \frac{1}{2}(\kappa/\overline{\sigma}_\kappa)\varphi_\kappa^3\ell_n^{1/2}$. By Fubini's theorem, conditioning each summand on ℓ_n, applying equality (4.29) and Lemma 4.14, we have

$$\mathbb{E}\bigg[\sum_{n\in\mathbb{N}}\ell_n\big|\mathbb{1}\{\xi_n\le x\}-\mathbb{1}\{\xi_n^{(\kappa)}\le x\}\big|\bigg]\le\frac{1}{2}\sqrt{t}(\kappa/\overline{\sigma}_\kappa)\varphi_\kappa^3\sum_{n\in\mathbb{N}}\mathbb{E}\big[(\ell_n/t)^{1/2}\big]=\mu_0^\tau(\kappa,t).$$

Let $\delta\in(0,2]$ satisfy $\inf_{u\in(0,1]}u^{\delta-2}(\overline{\sigma}_u^2+\sigma^2)>0$. By (4.10) in Theorem 4.2(b), we see that $\ell_n\big|\mathbb{P}(X_{\ell_n}\le x|\ell_n)-\mathbb{P}\big(X_{\ell_n}^{(\kappa)}\le x\big|\ell_n\big)\big|\le\psi_\kappa^{2/3}\ell_n^{1-2/(3\delta)}$, where $\psi_\kappa=C\kappa\varphi_\kappa$ as defined in (4.14). Moreover, we have $\ell_n\big|\mathbb{P}(X_{\ell_n}\le x|\ell_n)-\mathbb{P}\big(X_{\ell_n}^{(\kappa)}\le x\big|\ell_n\big)\big|\le\ell_n$. Hence by (4.29) and Lemma 4.14, we obtain

$$\begin{aligned}\sum_{n=1}^\infty\mathbb{E}\big[\ell_n\big|\mathbb{1}\{\xi_n\le x\}-\mathbb{1}\{\xi_n^{(\kappa)}\le x\}\big|\big]&\le\sum_{n=1}^\infty\mathbb{E}\big[\min\big\{\ell_n,\psi_\kappa^{2/3}\ell_n^{1-2/(3\delta)}\big\}\big]\\&=\begin{cases}t\wedge\psi_\kappa^\delta+\frac{3\delta}{3\delta-2}\psi_\kappa^{2/3}t^{1-\frac{2}{3\delta}}\big(1-\min\big\{1,t^{-1/\delta}\psi_\kappa\big\}^{\delta-2/3}\big), & \delta\in(0,2]\setminus\{\frac{2}{3}\},\\(t\wedge\psi_\kappa^{2/3})(1+\log^+(t\psi_\kappa^{-2/3})), & \delta=\frac{2}{3}.\end{cases}\end{aligned}\qquad\square$$

Proof of Theorem 4.3. Let $\ell=(\ell_n)_{n\in\mathbb{N}}$ and $(\xi_n,\xi_n^{(\kappa)})$, $n\in\mathbb{N}$, be as in Lemma 4.16. Define the vector

$$\big(\zeta_1,\zeta_2,\zeta_3,\zeta_1^{(\kappa)},\zeta_2^{(\kappa)},\zeta_3^{(\kappa)}\big)=\sum_{n=1}^\infty\big(\xi_n,\xi_n\wedge 0,\ell_n\mathbb{1}_{\{\xi_n\le 0\}},\xi_n^{(\kappa)},\xi_n^{(\kappa)}\wedge 0,\ell_n\mathbb{1}_{\{\xi_n^{(\kappa)}\le 0\}}\big).$$

By (4.1) and (4.25), we have $(\zeta_1,\zeta_2,\zeta_3)\overset{d}{=}\underline{\chi}_t$ and $(\zeta_1^{(\kappa)},\zeta_2^{(\kappa)},\zeta_3^{(\kappa)})\overset{d}{=}\underline{\chi}_t^{(\kappa)}$. Hence, it suffices to show that these vectors satisfy (4.11), (4.13) and (4.14). Since $x\mapsto\min\{x,0\}$ is in $\mathrm{Lip}_1(\mathbb{R})$, the inequalities

$$\max\big\{\big|\zeta_1-\zeta_1^{(\kappa)}\big|,\big|\zeta_2-\zeta_2^{(\kappa)}\big|\big\}\le\sum_{n=1}^\infty\big|\xi_n-\xi_n^{(\kappa)}\big|\qquad\text{and}$$

$$\big|\zeta_3-\zeta_3^{(\kappa)}\big|\le\sum_{n=1}^\infty\ell_n\big|\mathbb{1}\{\xi_n\le 0\}-\mathbb{1}\{\xi_n^{(\kappa)}\le 0\}\big|$$

follow from the triangle inequality. The theorem follows from Lemma 4.16. □

Remark 4.17. *Let C_t and $C_t^{(\kappa)}$ denote the convex minorants of X and $X^{(\kappa)}$ on $[0,t]$, respectively. Couple X and $X^{(\kappa)}$ in such a way that the stick-breaking processes describing the lengths of the faces of their convex minorants (see [87, Thm 1] and [56, Sec. 4.1]) coincide. (The Skorokhod space $\mathcal{D}[0,t]$ and the space of sequences on $\mathbb{R}$ are both Borel spaces by [62, Thms A1.1, A1.2 & A2.2], so the existence of such a coupling is guaranteed by [62, Thm 6.10].) Geometric arguments (see [53, §3]), show that the sequences of heights of the faces of the convex minorants, denoted by*

$(\xi_n)_{n\in\mathbb{N}}$ and $(\xi_n^{(\kappa)})_{n\in\mathbb{N}}$, satisfy

$$\sup_{s\in[0,t]}\big|C_t(s)-C_t^{(\kappa)}(s)\big|\le\sum_{n=1}^{\infty}\big|\xi_n-\xi_n^{(\kappa)}\big| \qquad \textit{and}$$
$$\big|\underline{\tau}_t-\underline{\tau}_t^{(\kappa)}\big|\le\sum_{n=1}^{\infty}\ell_n\big|\mathbb{1}\{\xi_n\le 0\}-\mathbb{1}\{\xi_n^{(\kappa)}\le 0\}\big|.$$

Hence, if $(\xi_n,\xi_n^{(\kappa)})$, $n\in\mathbb{N}$, are coupled as in Lemma 4.16, the inequalities in (4.26) *and* (4.27) *yield the same bounds as in Theorem 4.3 but in a stronger metric (namely, the distance between the convex minorants in the supremum norm), while retaining the control on the time of the infimum.*

§4.6.3 The proofs of Propositions 4.6, 4.7, 4.9 and 4.12

The Lévy-Khintchine formula for X_t in (4.3), the definition of $X_t^{(\kappa)}$ in (4.5) and the inequality $\mathrm{e}^z\ge 1+z$ (for all $z\in\mathbb{R}$) imply

$$\begin{aligned} t^{-1}\log\mathbb{E}\big[\mathrm{e}^{uX_t^{(\kappa)}}\big] &\\ &= bu+(\sigma^2+\overline{\sigma}_\kappa^2)\frac{u^2}{2}+\int_{\mathbb{R}\setminus(-\kappa,\kappa)}(\mathrm{e}^{ux}-1-ux\mathbb{1}_{(-1,1)}(x))\nu(\mathrm{d}x) \qquad (4.30)\\ &\le \overline{\sigma}_\kappa^2u^2/2+t^{-1}\log\mathbb{E}\big[\mathrm{e}^{uX_t}\big] \quad \text{for any } u\in\mathbb{R},\ t>0 \text{ and } \kappa\in(0,1]. \end{aligned}$$

Thus $\mathbb{E}[\exp(uX_t^{(\kappa)})]\le\mathbb{E}[\exp(uX_t)]\exp(\overline{\sigma}_\kappa^2u^2t/2)$ and, in particular, the Gaussian approximation $X^{(\kappa)}$ has as many exponential moments as the Lévy process X.

Proof of Proposition 4.6. By [99, Thm 6.16], there exists a coupling between $(\xi,\zeta)\overset{d}{=}(X_T,\underline{X}_T)$ and $(\xi',\zeta')\overset{d}{=}(X_T^{(\kappa)},\underline{X}_T^{(\kappa)})$, such that

$$\mathbb{E}[(|\xi-\xi'|+|\zeta-\zeta'|)^2]^{1/2}=\mathcal{W}_2((X_T,\underline{X}_T),(X_T^{(\kappa)},\underline{X}_T^{(\kappa)})).$$

The identity $\mathrm{e}^b-\mathrm{e}^a=\int_a^b\mathrm{e}^z dz$ implies that, for $x\ge y$ and $x'\ge y'$, we have

$$|f(x,y)-f(x',y')|\le K(|\mathrm{e}^x-\mathrm{e}^{x'}|+|\mathrm{e}^y-\mathrm{e}^{y'}|)\le K(|x-x'|+|y-y'|)\mathrm{e}^{\max\{x,x'\}}. \quad (4.31)$$

Apply this inequality, the Cauchy-Schwartz inequality, the elementary inequalities, which hold for all $a,b\ge 0$, $(a+b)^2\le 2(a^2+b^2)$ and $(a+b)^{1/2}\le a^{1/2}+b^{1/2}$ and the bound in (4.30) to obtain

$$\begin{aligned}\mathbb{E}|f(\xi,\zeta)-f(\xi',\zeta')| &\le K\mathbb{E}[(|\xi-\xi'|+|\zeta-\zeta'|)^2]^{1/2}\mathbb{E}[(\mathrm{e}^{\xi}+\mathrm{e}^{\xi'})^2]^{1/2}\\ &\le 2^{1/2}K\mathcal{W}_2((X_T,\underline{X}_T),(X_T^{(\kappa)},\underline{X}_T^{(\kappa)}))\mathbb{E}[\mathrm{e}^{2\xi}+\mathrm{e}^{2\xi'}]^{1/2}\\ &\le 2K\mathcal{W}_2((X_T,\underline{X}_T),(X_T^{(\kappa)},\underline{X}_T^{(\kappa)}))\mathbb{E}[\mathrm{e}^{2X_T}]^{1/2}(1+\mathrm{e}^{\overline{\sigma}_\kappa^2 T}).\end{aligned}$$

Applying Corollary 4.4 gives the desired inequality, concluding the proof of the proposition. □

We now introduce a tool that uses the L^p-distance $\mathbb{E}[|\zeta-\zeta'|^p]^{1/p}$ between random variables ζ and ζ' to bound the L^1-distance $\mathbb{E}|\mathbb{1}_{[y,\infty)}(\zeta)-\mathbb{1}_{[y,\infty)}(\zeta')|$ between the indicators.

Lemma 4.18. *Let (ξ,ζ) and (ξ',ζ') be random vectors in $\mathbb{R}^n\times\mathbb{R}$. Fix $y\in\mathbb{R}$ and let $h\in\mathrm{Lip}_K(\mathbb{R}^n)$ satisfy $0\le h\le M$ for some constants $K,M\ge 0$. Then for any $p,r>0$, the difference $H=h(\xi)\mathbb{1}_{[y,\infty)}(\zeta)-h(\xi')\mathbb{1}_{[y,\infty)}(\zeta')$ satisfies*

$$\mathbb{E}|H|\le K\mathbb{E}\|\xi-\xi'\|+M\mathbb{P}(|\zeta-y|\le r)+Mr^{-p}\mathbb{E}[|\zeta-\zeta'|^p]. \tag{4.32}$$

In particular, if $|\mathbb{P}(\zeta\le y)-\mathbb{P}(\zeta\le y+r)|\le C|r|^\gamma$ for some $C,\gamma>0$ and all $r\in\mathbb{R}$,

$$\mathbb{E}|H|\le K\mathbb{E}\|\xi-\xi'\|+M(2C\gamma/p)^{\frac{p}{p+\gamma}}(1+p/\gamma)\mathbb{E}[|\zeta-\zeta'|^p]^{\frac{\gamma}{p+\gamma}}. \tag{4.33}$$

Remark 4.19. *An analogous bound to the one in* (4.32) *holds for the indicator $\mathbb{1}_{(-\infty,y]}$. Moreover, it follows from the proof below that the boundedness of the function h assumed in Lemma 4.18 may be replaced with a moment assumption $\xi,\xi'\in L^q$ for some $q>1$. In such a case, Hölder's inequality could be invoked to obtain an analogue to* (4.34) *below. Similar arguments may be used to simultaneously handle multiple indicators.*

Proof of Lemma 4.18. Applying the assumed local γ-Hölder continuous property of the distribution function of ζ to (4.32) and optimising over $r>0$ yields (4.33). Thus, it remains to establish (4.32).

Elementary set manipulation yields

$$\begin{aligned}&|\mathbb{1}_{\{y\le\zeta\}}-\mathbb{1}_{\{y\le\zeta'\}}|=|\mathbb{1}_{\{\zeta'<y\le\zeta\}}-\mathbb{1}_{\{\zeta<y\le\zeta'\}}|\\ &\quad\le\mathbb{1}_{\{|\zeta-\zeta'|>r,\zeta'<y\le\zeta\}}+\mathbb{1}_{\{|\zeta-\zeta'|\le r,\zeta'<y\le\zeta\}}+\mathbb{1}_{\{|\zeta-\zeta'|>r,\zeta<y\le\zeta'\}}+\mathbb{1}_{\{|\zeta-\zeta'|\le r,\zeta<y\le\zeta'\}}\\ &\quad\le\mathbb{1}_{\{|\zeta-\zeta'|>r\}}+\mathbb{1}_{\{|\zeta-y|\le r\}}.\end{aligned}$$

Hence, the triangle inequality and the Lipschitz property gives

$$\begin{aligned}
&|h(\xi)\mathbb{1}_{[y,\infty)}(\zeta) - h(\xi')\mathbb{1}_{[y,\infty)}(\zeta')| \\
&\qquad \le |h(\xi)||\mathbb{1}_{[y,\infty)}(\zeta) - \mathbb{1}_{[y,\infty)}(\zeta')| + |h(\xi) - h(\xi')|\mathbb{1}_{[y,\infty)}(\zeta') \\
&\qquad \le M(\mathbb{1}_{\{|\zeta-y|\le r\}} + \mathbb{1}_{\{|\zeta-\zeta'|>r\}}) + K\|\xi - \xi'\|.
\end{aligned} \tag{4.34}$$

Taking expectations and using Markov's inequality $\mathbb{P}(|\zeta - \zeta'| > r) \le r^{-p}\mathbb{E}[|\zeta - \zeta'|^p]$ yields (4.32). □

Proof of Proposition 4.7. Theorem 4.3 and (4.33) in Lemma 4.18 (with C and γ given in Assumption (H) and $p = 2$) applied to $(X_T, \underline{X}_T)$ and $(X_T^{(\kappa)}, \underline{X}_T^{(\kappa)})$ under the SBG coupling give the claim. □

Proof of Proposition 4.9. Analogous to the proof of Proposition 4.7, Theorem 4.3 and (4.33) in Lemma 4.18 (with C and γ given in Assumption ($H\tau$) and $p = 1$), give the result. □

Lemma 4.20. *Suppose X is not a compound Poisson process. Then the law of $\underline{\tau}_T$ is absolutely continuous on $(0,T)$ and its density is locally bounded on $(0,T)$.*

Proof. If X or $-X$ is a subordinator then $\underline{\tau}_T$ is a.s. 0 or T, respectively. In either case, the result follows immediately. Suppose now that neither X nor $-X$ is a subordinator. Denote by $\overline{n}(\zeta > \cdot)$ (resp. $\underline{n}(\zeta > \cdot)$) the intensity measures of the lengths ζ of the excursions away from 0 of the Markov process $\overline{X} - X$ (resp. $X - \underline{X}$). Then, by Theorem 5 in [28] with $F \equiv K \equiv 1$, the law of $\underline{\tau}_T$ can only have atoms at 0 or T, is absolutely continuous on $(0,T)$ and its density is given by $s \mapsto \underline{n}(\zeta > s)\overline{n}(\zeta > T - s)$, $s \in (0,T)$. The maps $s \mapsto \underline{n}(\zeta > s)$ and $s \mapsto \overline{n}(\zeta > s)$ are non-increasing, so the density is bounded on any compact subset of $(0,T)$, completing the proof. □

In preparation for the next result, we introduce the following assumption.

Assumption (S-α)**.** *There exists some function $a : (0,\infty) \to (0,\infty)$ such that $X_t/a(t)$ converges in distribution to an α-stable law as $t \to 0$.*

Proposition 4.21. *Let Assumption (S-α) hold for some $\alpha \in (0,2]$.*
(a) *If $\alpha > 1$, then Assumption (H) holds uniformly on compact subsets of $(-\infty,0)$ with $\gamma = 1$.*
(b) *Suppose $\rho = \lim_{t\to0}\mathbb{P}(X_t > 0) \in (0,1)$. Then for any $\gamma \in (0,\min\{\rho, 1-\rho\})$, there exists some constant $C > 0$ such that Assumption ($H\tau$) holds for all $s \in [0,T]$.*

Note that ρ is well defined under Assumption (S-α) and that $X_t/a(t)$ can only have a nonzero weak limit as $t \to 0$ if the limit is α-stable. Moreover, in that case, a is necessarily regularly varying at 0 with index $1/\alpha$ and α is given in terms of the Lévy triplet (σ^2, ν, b) of X:

$$\alpha = \begin{cases} 2, & \sigma \neq 0, \\ 1, & \beta \in (0,1) \text{ and } b \neq \int_{(-1,1)} x\nu(\mathrm{d}x), \\ \beta, & \text{otherwise,} \end{cases}$$

where β is the BG index in (4.6). In fact, the assumptions of Proposition 4.12 imply Assumption (S-α) by [17, Prop. 2.3], so Proposition 4.21 generalises Proposition 4.12. We refer the reader to [60, Sec. 3 & 4] for conditions that are equivalent to (S-α).

Assumption (S-α) allows for the cases $\rho = 0$ or $\rho = 1$ when $\alpha \leq 1$, correspond to the stable limit being a.s. negative or a.s. positive, respectively. In these cases, the distribution of $\overline{\tau}_T(X)$ may have an atom at 0 or T, while the law of $\overline{\tau}_T(X^{(\kappa)})$ is absolutely continuous, making the convergence in Kolmogorov distance impossible. This is the reason for excluding $\rho \in \{0, 1\}$ in Proposition 4.21.

Proof of Proposition 4.21. By [17, Lem. 5.7], under the assumptions in part (a) of the proposition, $\underline{X}_T$ has a continuous density on $(-\infty, 0)$, implying the conclusion in (a).

Since $\rho = \lim_{t\to 0} \mathbb{P}(X_t > 0) \in (0,1)$, 0 is regular for both half-lines by Rogozin's criterion [91, Thm 47.2]. [28, Thm 6] then asserts that the law of $\underline{\tau}_T$ is absolutely continuous with density given by $s \mapsto \underline{n}(\zeta > s)\overline{n}(\zeta > T - s)$, $s \in (0,T)$. The maps $s \mapsto \underline{n}(\zeta > s)$ and $s \mapsto \overline{n}(\zeta > s)$ are non-increasing and, by [17, Prop. 3.5], regularly varying with indices $\rho - 1$ and $-\rho$, respectively. Thus for any $\gamma \in (0, \min\{\rho, 1-\rho\})$ there exists some $C > 0$ such that $\underline{n}(\zeta > s)\overline{n}(\zeta > T - s) \leq Cs^{\gamma-1}(T-s)^{\gamma-1}$ for all $s \in (0,T)$. Thus, for any $s, t \in [0, T/2]$ with $t \geq s$, we have

$$\begin{aligned} \mathbb{P}(\underline{\tau}_T \leq t) - \mathbb{P}(\underline{\tau}_T \leq s) &\leq \int_s^t Cu^{\gamma-1}(T-u)^{\gamma-1}du \leq C\int_s^t u^{\gamma-1}(T/2)^{\gamma-1}du \\ &\leq C\gamma^{-1}(T/2)^{\gamma-1}(t^\gamma - s^\gamma) \leq C\gamma^{-1}(T/2)^{\gamma-1}(t-s)^\gamma. \end{aligned}$$

since the map $x \mapsto x^\gamma$ is concave. A similar bound holds for $s, t \in [T/2, T]$.

Moreover, when $s \in [0,T/2]$ and $t \in [T/2,T]$ we have

$$\begin{aligned}\mathbb{P}(\underline{\tau}_T \le t) - \mathbb{P}(\underline{\tau}_T \le s) &\le \mathbb{P}(\underline{\tau}_T \le t) - \mathbb{P}(\underline{\tau}_T \le T/2) + \mathbb{P}(\underline{\tau}_T \le T/2) - \mathbb{P}(\underline{\tau}_T \le s)\\ &\le C\gamma^{-1}(T/2)^{\gamma-1}[(T/2-s)^\gamma + (t-T/2)^\gamma]\\ &\le C\gamma^{-1}(T/2)^{2\gamma-2}(t-s)^\gamma.\end{aligned}$$

This gives part (b) of the proposition. □

§4.6.4 Level variances under SBG-Alg

In the present subsection we establish bounds on the level variances under the coupling $\underline{\Pi}_{n,T}^{\kappa_1,\kappa_2}$ (constructed in SBG-Alg) for Lipschitz, locally Lipschitz and even discontinuous payoff functions (see BT_1 in (4.19) and BT_2 in (4.21)) of $\underline{\chi}_T$.

Theorem 4.22. *Fix $T > 0$, $n \in \mathbb{N}$ and $1 \ge \kappa_1 > \kappa_2 > 0$. Denote the vector $(Z_{n,T}^{(\kappa_i)}, \underline{Z}_{n,T}^{(\kappa_i)}, \underline{\tau}_{n,T}^{(\kappa_i)}) = \underline{\chi}_{n,T}^{(\kappa_i)}$, $i \in \{1,2\}$, where the vector $\big(\underline{\chi}_{n,T}^{(\kappa_1)}, \underline{\chi}_{n,T}^{(\kappa_2)}\big)$, constructed in SBG-Alg, follows the law $\underline{\Pi}_{n,T}^{\kappa_1,\kappa_2}$.*

(a) *For any Lipschitz function $f \in \mathrm{Lip}_K(\mathbb{R}^2)$, $K > 0$, we have*

$$\mathbb{E}\big[\big(f\big(Z_{n,T}^{(\kappa_2)}, \underline{Z}_{n,T}^{(\kappa_2)}\big) - f\big(Z_{n,T}^{(\kappa_1)}, \underline{Z}_{n,T}^{(\kappa_1)}\big)\big)^2\big] \le K^2T\big(27\sigma^2 2^{-n} + 40\overline{\sigma}_{\kappa_1}^2\big). \tag{4.35}$$

For $f \in \mathrm{locLip}_K(\mathbb{R}^2)$, defined in §4.3.2 above, if $\int_{[1,\infty)} \mathrm{e}^{4x}\nu(\mathrm{d}x) < \infty$ then there exists a constant $C > 0$ independent of (n,κ_1,κ_2) such that

$$\mathbb{E}\big[\big(f\big(Z_{n,T}^{(\kappa_2)}, \underline{Z}_{n,T}^{(\kappa_2)}\big) - f\big(Z_{n,T}^{(\kappa_1)}, \underline{Z}_{n,T}^{(\kappa_1)}\big)\big)^2\big] \le C\big(\big(\tfrac{2}{3}\big)^{n/2}\mathbb{1}_{\{\sigma\ne 0\}} + \overline{\sigma}_{\kappa_1}^2 + \overline{\sigma}_{\kappa_1}\kappa_1\big). \tag{4.36}$$

(b) *Suppose Assumption (H) is satisfied by some $y < 0$ and $C,\gamma > 0$. Then for any $f \in \mathrm{BT}_1(y,K,M)$, $K,M \ge 0$, there exists some $K' > 0$ independent of (n,κ_1,κ_2) such that*

$$\mathbb{E}\big[\big(f\big(Z_{n,T}^{(\kappa_2)}, \underline{Z}_{n,T}^{(\kappa_2)}\big) - f\big(Z_{n,T}^{(\kappa_1)}, \underline{Z}_{n,T}^{(\kappa_1)}\big)\big)^2\big] \le K'\big(\sigma^2 2^{-n} + \overline{\sigma}_{\kappa_1}^2\big)^{\frac{\gamma}{2+\gamma}}. \tag{4.37}$$

(c) *If $\delta \in (0,2]$ satisfies Assumption (O-δ), then there exists some $C > 0$ such that for any $K > 0$, $f \in \mathrm{Lip}_K(\mathbb{R})$, $n \in \mathbb{N}$, $\kappa_1 > \kappa_2$ and $p \in \{1,2\}$, we have*

$$\frac{\mathbb{E}\big[\big|f\big(\underline{\tau}_{n,T}^{(\kappa_1)}\big) - f\big(\underline{\tau}_{n,T}^{(\kappa_2)}\big)\big|^p\big]}{2K^pT^p} \le 2^{-n} + C\overline{\sigma}_{\kappa_1}^{\min\{\frac{2\delta}{2-\delta},\frac{1}{2}\}}\big(1 + |\log\kappa_1|\mathbb{1}_{\{\delta=\frac{2}{5}\}}\big). \tag{4.38}$$

(d) *Fix $s \in (0,T)$ and let Assumption (O-δ) hold for some $\delta \in (0,2]$, then for any $f \in \mathrm{BT}_2(s,K,M)$, $K,M \ge 0$, there exists a constant $C > 0$ such that for any $n \in \mathbb{N}$,*

$p \in \{1,2\}$ *and* $\kappa_1 > \kappa_2$, *we have*

$$\mathbb{E}\big[\big|f\big(\underline{\chi}_{n,T}^{(\kappa_1)}\big) - f\big(\underline{\chi}_{n,T}^{(\kappa_2)}\big)\big|^p\big] \le C\big[2^{-\frac{n}{2}} + \overline{\sigma}_{\kappa_1}^{\min\{\frac{\delta}{2-\delta},\frac{1}{4}\}}\big(1 + \sqrt{|\log \kappa_1|}\mathbb{1}_{\{\delta=\frac{2}{5}\}}\big)\big]. \quad (4.39)$$

The synchronous coupling of the large jumps of the Gaussian approximations, implicit in SBG-Alg, ensures that no moment assumption on the large jumps of X is necessary for (4.35) to hold. For locally Lipschitz payoffs, however, the function may magnify the distance when a large jump occurs. This leads to the moment assumption $\int_{[1,\infty)} \mathrm{e}^{4x}\nu(\mathrm{d}x) < \infty$ for $f \in \mathrm{locLip}_K(\mathbb{R}^2)$.

The proof of Theorem 4.22 requires bounds on certain moments of the differences of the components of the output of Algorithms 2 & 3 and SBG-Alg, given in Proposition 4.23.

Proposition 4.23. *For any* $1 \ge \kappa_1 > \kappa_2 > 0$, $t > 0$ *and* $n \in \mathbb{N}$, *the following statements hold.*
(a) *The pair* $\big(Z_t^{(\kappa_1)}, Z_t^{(\kappa_2)}\big) \sim \Pi_t^{\kappa_1,\kappa_2}$, *constructed in Algorithm 2, satisfies*

$$\mathbb{E}\big[\big(Z_t^{(\kappa_1)} - Z_t^{(\kappa_2)}\big)^2\big] \le 2(\overline{\sigma}_{\kappa_1}^2 - \overline{\sigma}_{\kappa_2}^2)t,$$
$$\mathbb{E}\big[\big(Z_t^{(\kappa_1)} - Z_t^{(\kappa_2)}\big)^4\big] \le 12(\overline{\sigma}_{\kappa_1}^2 - \overline{\sigma}_{\kappa_2}^2)^2t^2 + (\overline{\sigma}_{\kappa_1}^2 - \overline{\sigma}_{\kappa_2}^2)\kappa_1^2 t.$$

Moreover, we have $\mathbb{E}[(\underline{Z}_t^{(\kappa_1)} - \underline{Z}_t^{(\kappa_2)})^{2p}] \le 4\mathbb{E}[(Z_t^{(\kappa_1)} - Z_t^{(\kappa_2)})^{2p}]$, *for any* $p \in \{1,2\}$.
(b) *The vector* $\big(Z_t^{(\kappa_1)}, \underline{Z}_t^{(\kappa_1)}, \underline{\tau}_t^{(\kappa_1)}, Z_t^{(\kappa_2)}, \underline{Z}_t^{(\kappa_2)}, \underline{\tau}_t^{(\kappa_2)}\big) \sim \underline{\Pi}_t^{\kappa_1,\kappa_2}$ *in Algorithm 3 satisfies the following inequalities*

$$\mathbb{E}\big[\big(Z_t^{(\kappa_1)} - Z_t^{(\kappa_2)}\big)^2\big] = 2(\sigma^2 + \overline{\sigma}_{\kappa_1}^2)t,$$
$$\mathbb{E}\big[\big(Z_t^{(\kappa_1)} - Z_t^{(\kappa_2)}\big)^4\big] \le 12(\sigma^2 + \overline{\sigma}_{\kappa_1}^2)^2t^2 + (\overline{\sigma}_{\kappa_1}^2 - \overline{\sigma}_{\kappa_2}^2)\kappa_1^2 t.$$

Moreover, we have $\mathbb{E}[(\underline{Z}_t^{(\kappa_1)} - \underline{Z}_t^{(\kappa_2)})^{2p}] \le 4\mathbb{E}[(Z_t^{(\kappa_1)} - Z_t^{(\kappa_2)})^{2p}]$, *for any* $p \in \{1,2\}$.
(c) *The coupling* $\big(\underline{\chi}_{n,t}^{(\kappa_1)}, \underline{\chi}_{n,t}^{(\kappa_2)}\big) \sim \underline{\Pi}_{n,t}^{\kappa_1,\kappa_2}$, *constructed in SBG-Alg, with components* $\underline{\chi}_{n,t}^{(\kappa_i)} = (Z_{n,t}^{(\kappa_i)}, \underline{Z}_{n,t}^{(\kappa_i)}, \underline{\tau}_{n,t}^{(\kappa_i)})$, $i \in \{1,2\}$, *satisfies the following inequalities:*

$$\mathbb{E}\big[\big(Z_{n,t}^{(\kappa_1)} - Z_{n,t}^{(\kappa_2)}\big)^2\big] \le 2(\sigma^2 2^{-n} + \overline{\sigma}_{\kappa_1}^2)t, \quad (4.40)$$
$$\mathbb{E}\big[\big(Z_{n,t}^{(\kappa_1)} - Z_{n,t}^{(\kappa_2)}\big)^4\big] \le (25\overline{\sigma}_{\kappa_1}^4 + 24\sigma^4 3^{-n})t^2 + \overline{\sigma}_{\kappa_1}^2\kappa_1^2 t, \quad (4.41)$$
$$\mathbb{E}\big[\big(\underline{Z}_{n,t}^{(\kappa_1)} - \underline{Z}_{n,t}^{(\kappa_2)}\big)^2\big] \le (2+3\pi)(\sigma^2 + \overline{\sigma}_{\kappa_1}^2)2^{-n}t + (2+5\pi)\overline{\sigma}_{\kappa_1}^2 t, \quad (4.42)$$
$$\mathbb{E}\big[\big(\underline{Z}_{n,t}^{(\kappa_1)} - \underline{Z}_{n,t}^{(\kappa_2)}\big)^4\big] \le 2\cdot 10^3\big[(\sigma^2 + \overline{\sigma}_{\kappa_1})^2 3^{-n} + \overline{\sigma}_{\kappa_1}^4\big]t^2 + 2\pi\overline{\sigma}_{\kappa_1}^{\frac{5}{2}}\kappa_1^{\frac{3}{2}}t^{\frac{5}{4}} + 4\overline{\sigma}_{\kappa_1}^2\kappa_1^2 t. \quad (4.43)$$

Remark 4.24. (i) By Proposition 4.23, the L^2-norms of the differences $Z_{n,t}^{(\kappa_1)} - Z_{n,t}^{(\kappa_2)}$ and $\underline{Z}_{n,t}^{(\kappa_1)} - \underline{Z}_{n,t}^{(\kappa_2)}$ of the components of $(\underline{\chi}_{n,t}^{(\kappa_1)}, \underline{\chi}_{n,t}^{(\kappa_2)})$, constructed in SBG-Alg, decay at the same rate as the L^2-norm of $Z_t^{(\kappa_1)} - Z_t^{(\kappa_2)}$, constructed in Algorithm 2. Indeed, assume that $\kappa_1 = c\kappa_2$ for some $c > 1$, $\kappa_2 \to 0$ and, for some $c', r > 0$ and all $x > 0$, we have $\overline{\nu}(x) = \nu(\mathbb{R} \setminus (-x, x)) \geq c' x^{-r}$. Then, for $n = \lceil \log^2(1 + \overline{\nu}(\kappa_2)) \rceil$ we have $2^{-n} \leq \overline{\sigma}_{\kappa_1}^2$ for all sufficiently small κ_1, implying the claim by Proposition 4.23(a) & (c). Moreover, by Corollary 4.13, the corresponding expected computational complexities of Algorithm 2 and SBG-Alg are proportional as $\kappa_2 \to 0$. Furthermore, since the decay of the bias of SBG-Alg is, by Theorem 4.3, at most a logarithmic factor away from that of Algorithm 2, the MLMC estimator based on Algorithm 2 for $\mathbb{E}f(X_t)$ has the same computational complexity (up to logarithmic factors) as the MLMC estimator for $\mathbb{E}f(X_t, \underline{X}_t)$ based on SBG-Alg (see Table 4.3 above for the complexity of the latter).
(ii) The proof of Proposition 4.23 implies that an improvement in Algorithm 2 (i.e. a simulation procedure for a coupling with a smaller L^2-norm of $Z_t^{(\kappa_1)} - Z_t^{(\kappa_2)}$) would result in an improvement in SBG-Alg for the simulation of a coupling $(\underline{\chi}_t^{(\kappa_1)}, \underline{\chi}_t^{(\kappa_2)})$. Interestingly, this holds in spite of the fact that SBG-Alg calls Algorithm 3 whose coupling $\underline{\Pi}_t^{\kappa_1,\kappa_2}$ is inefficient in terms of the L^2-distance but is applied over the short interval $[0, L_n]$.
(iii) A nontrivial bound on the moments of the difference $\underline{\tau}_t^{(\kappa_1)} - \underline{\tau}_t^{(\kappa_2)}$ under the coupling of Algorithm 3, which would complete the statement in Proposition 4.23(b), appears to be out of reach. By the SB representation in (4.2), such a bound is not necessary for our purposes. The corresponding bound on the moments of the difference $\underline{\tau}_{n,t}^{(\kappa_1)} - \underline{\tau}_{n,t}^{(\kappa_2)}$, constructed in SBG-Alg, follows from Proposition 4.25 below, see (4.49).
(iv) The bounds on the fourth moments in (4.41) and (4.43) are required to control the level variances of the MLMC estimator in the case of locally Lipschitz payoff functions and are applied in the proof of Theorem 4.22(a).

Proof of Proposition 4.23. (a) The difference $Z_t^{(\kappa_1)} - Z_t^{(\kappa_2)}$ equals, by (4.5), a sum of two independent martingales: $((\overline{\sigma}_{\kappa_1}^2 + \sigma^2)^{1/2} - (\overline{\sigma}_{\kappa_2}^2 + \sigma^2)^{1/2}) W_t$ and $J_t^{2,\kappa_1} - J_t^{2,\kappa_2} + (b_{\kappa_1} - b_{\kappa_2})t$. Thus, we obtain the identity

$$\mathbb{E}\big[\big(Z_t^{(\kappa_1)} - Z_t^{(\kappa_2)}\big)^2\big] = \Big[\Big(\sqrt{\sigma^2 + \overline{\sigma}_{\kappa_1}^2} - \sqrt{\sigma^2 + \overline{\sigma}_{\kappa_2}^2}\Big)^2 + \overline{\sigma}_{\kappa_1}^2 - \overline{\sigma}_{\kappa_2}^2\Big] t.$$

The first inequality follows since $0 < (\sigma^2 + \overline{\sigma}_{\kappa_1}^2)^{1/2} - (\sigma^2 + \overline{\sigma}_{\kappa_2}^2)^{1/2} \leq (\overline{\sigma}_{\kappa_1}^2 - \overline{\sigma}_{\kappa_2}^2)^{1/2}$. Since $Z_t^{(\kappa_1)} - Z_t^{(\kappa_2)}$ is a Lévy process, differentiating its Lévy-Khintchine formula in (4.3) yields the identity

$$\mathbb{E}\big[\big(Z_t^{(\kappa_1)} - Z_t^{(\kappa_2)}\big)^4\big] = 3\Big[\Big(\sqrt{\sigma^2 + \overline{\sigma}_{\kappa_1}^2} - \sqrt{\sigma^2 + \overline{\sigma}_{\kappa_2}^2}\Big)^2 + \overline{\sigma}_{\kappa_1}^2 - \overline{\sigma}_{\kappa_2}^2\Big]^2 t^2 + t\int_{(-\kappa_1,\kappa_1)\setminus(-\kappa_2,\kappa_2)} x^4\nu(\mathrm{d}x),$$

which implies the second inequality. Since $|\underline{Z}_t^{(\kappa_1)} - \underline{Z}_t^{(\kappa_2)}| \le \sup_{s\in[0,t]} |Z_s^{(\kappa_1)} - Z_s^{(\kappa_2)}|$, Doob's maximal martingale inequality [62, Prop. 7.16] applied to the martingale $(Z_s^{(\kappa_1)} - Z_s^{(\kappa_2)})_{s\in[0,t]}$ yields

$$\mathbb{E}\big[|\underline{Z}_t^{(\kappa_1)} - \underline{Z}_t^{(\kappa_2)}|^p\big] \le \big(1-1/p\big)^{-p}\mathbb{E}\big[|Z_t^{(\kappa_1)} - Z_t^{(\kappa_2)}|^p\big], \qquad p>1.$$

The corresponding inequalities follow because $(p/(p-1))^p \le 4$ for $p\in\{2,4\}$.

(b) Analogous to part (a), the difference $Z_t^{(\kappa_1)} - Z_t^{(\kappa_2)}$ constructed in Algorithm 3 is a sum of two independent martingales: $(\overline{\sigma}_{\kappa_1}^2 + \sigma^2)^{1/2}B_t - (\overline{\sigma}_{\kappa_2}^2 + \sigma^2)^{1/2}W_t$ and $J_t^{2,\kappa_1} - J_t^{2,\kappa_2} + (b_{\kappa_1} - b_{\kappa_2})t$, where B and W are independent standard Brownian motions. Thus the statements follow as in part (a).

(c) Let $(\xi_{1,k},\xi_{2,k}) \sim \Pi_{\ell_k}^{\kappa_1,\kappa_2}$, $k\in\{1,\dots,n\}$, and $(\underline{\zeta}_1,\underline{\zeta}_2) \sim \underline{\Pi}_{L_n}^{\kappa_1,\kappa_2}$ be independent draws as in line 2 of SBG-Alg above. Denote by $(\xi_{i,n+1},\underline{\xi}_{i,n+1})$ the first two coordinates of $\underline{\zeta}_i$, $i\in\{1,2\}$. Since the variables $\{\xi_{1,k}-\xi_{2,k}\}_{k=1}^{n+1}$ have zero mean and are uncorrelated, by conditioning on $\{\ell_k\}_{k=1}^n$ and L_n and applying parts (a) and (b) we obtain

$$\begin{aligned}\mathbb{E}\big[\big(Z_{n,t}^{(\kappa_1)} - Z_{n,t}^{(\kappa_2)}\big)^2\big] = \mathbb{V}\big[Z_{n,t}^{(\kappa_1)} - Z_{n,t}^{(\kappa_2)}\big] &= \mathbb{V}\big[\xi_{1,n+1} - \xi_{2,n+1}\big] + \sum_{k=1}^n \mathbb{V}\big[\xi_{1,k} - \xi_{2,k}\big]\\ &\le 2(\sigma^2 + \overline{\sigma}_{\kappa_1}^2)\mathbb{E}[L_n] + 2\overline{\sigma}_{\kappa_1}^2\sum_{k=1}^n \mathbb{E}[\ell_k]\\ &= 2(\sigma^2 + \overline{\sigma}_{\kappa_1}^2)2^{-n}t + 2\overline{\sigma}_{\kappa_1}^2(1-2^{-n})t.\end{aligned}$$

implying (4.40). Similarly, by conditioning on $\{\ell_k\}_{k=1}^n$ and L_n, we deduce that the expectations of

$$(\xi_{1,k_1} - \xi_{2,k_1})^3(\xi_{1,k_2} - \xi_{2,k_2}), \quad (\xi_{1,k_1} - \xi_{2,k_1})^2\prod_{i=2}^3(\xi_{1,k_i} - \xi_{2,k_i}), \quad \text{and} \quad \prod_{i=1}^4(\xi_{1,k_i} - \xi_{2,k_i}),$$

vanish for any distinct $k_1, k_2, k_3, k_4 \in \{1, \ldots, n+1\}$. Thus, by expanding, we obtain

$$\mathbb{E}\big[\big(Z_{n,t}^{(\kappa_1)} - Z_{n,t}^{(\kappa_2)}\big)^4\big] = \sum_{k=1}^{n+1} \mathbb{E}\big[\big(\xi_{1,k} - \xi_{2,k}\big)^4\big] + 6 \sum_{m=1}^{n} \sum_{k=m+1}^{n+1} \mathbb{E}\big[\big(\xi_{1,m} - \xi_{2,m}\big)^2 \big(\xi_{1,k} - \xi_{2,k}\big)^2\big].$$

The summands in the first sum are easily bounded by parts (a) and (b). To bound the summands of the second sum, condition on $\{\ell_k\}_{k=1}^n$ and L_n and apply parts (a) and (b):

$$\mathbb{E}\big[\big(\xi_{1,k} - \xi_{2,k}\big)^2 \big(\xi_{1,m} - \xi_{2,m}\big)^2\big] \le \begin{cases} 4\overline{\sigma}_{\kappa_1}^4 \mathbb{E}[\ell_m \ell_k], & m < k \le n, \\ 4(\sigma^2 + \overline{\sigma}_{\kappa_1}^2)\overline{\sigma}_{\kappa_1}^2 \mathbb{E}[\ell_m L_n], & m < k = n+1. \end{cases}$$

Inequality (4.41) follows since $\mathbb{E}[\ell_m \ell_k] = 3^{-m} 2^{m-k-1} t^2$, $\mathbb{E}[\ell_k L_n] = 3^{-k} 2^{k-n-1} t^2$ for $m < k \le n$ and $\sigma^2 2^{-n} \overline{\sigma}_\kappa^2 \le \sigma^2 3^{-n/2} \overline{\sigma}_\kappa^2 \le (\sigma^4 3^{-n} + \overline{\sigma}_\kappa^4)/2$.

The representation in line 3 of SBG-Alg and the elementary inequality: for all $a, b \in \mathbb{R}$, $|\min\{a, 0\} - \min\{b, 0\}| \le |a - b|$, imply

$$\begin{aligned} \mathbb{E}\big[\big(\underline{Z}_{n,t}^{(\kappa_1)} - \underline{Z}_{n,t}^{(\kappa_2)}\big)^2\big] \le\ & \mathbb{E}\Big[\big(\underline{\xi}_{1,n+1} - \underline{\xi}_{2,n+1}\big)^2 + \sum_{k=1}^{n} \big(\xi_{1,k} - \xi_{2,k}\big)^2\Big] \\ & + 2\mathbb{E} \sum_{k=1}^{n} \big|\underline{\xi}_{1,n+1} - \underline{\xi}_{2,n+1}\big| \big|\xi_{1,k} - \xi_{2,k}\big| \\ & + \mathbb{E} \sum_{m=1}^{n-1} \sum_{k=m+1}^{n} \big|\xi_{1,m} - \xi_{2,m}\big| \big|\xi_{1,k} - \xi_{2,k}\big|. \end{aligned} \tag{4.44}$$

The first term on the right-hand side of this inequality is easily bounded via the inequalities in parts (a) and (b). To bound the second term, condition on $\{\ell_k\}_{k=1}^n$ and L_n, apply the Cauchy-Schwarz inequality, denote $\upsilon = \sqrt{\sigma^2 + \overline{\sigma}_{\kappa_1}^2}$ and observe that for $m < k \le n$ we get

$$\begin{aligned} \mathbb{E}\big[\big|\underline{\xi}_{1,n+1} - \underline{\xi}_{2,n+1}\big| \big|\xi_{1,k} - \xi_{2,k}\big|\big] &\le \mathbb{E}\Big[\sqrt{16(\sigma^2 + \overline{\sigma}_{\kappa_1}^2)\overline{\sigma}_{\kappa_1}^2 \ell_k L_n}\Big] = \pi \upsilon \overline{\sigma}_{\kappa_1} \big(\tfrac{2}{3}\big)^n \big(\tfrac{3}{4}\big)^k t, \\ \mathbb{E}\big[\big|\xi_{1,m} - \xi_{2,m}\big| \big|\xi_{1,k} - \xi_{2,k}\big|\big] &\le \mathbb{E}\Big[\sqrt{4\overline{\sigma}_{\kappa_1}^4 \ell_m \ell_k}\Big] = \pi \overline{\sigma}_{\kappa_1}^2 \big(\tfrac{1}{2}\big)^{m+1} \big(\tfrac{2}{3}\big)^{k-m} t, \end{aligned}$$

where the equalities follow from the definition of the stick-breaking process. By (4.44) we have

$$\mathbb{E}\big[\big(\underline{Z}_{n,t}^{(\kappa_1)} - \underline{Z}_{n,t}^{(\kappa_2)}\big)^2\big] \leq \upsilon^2 2^{1-n} t + 2\overline{\sigma}_{\kappa_1}^2 t \sum_{k=1}^{\infty} 2^{-k} + 2\pi\upsilon\overline{\sigma}_{\kappa_1}\big(\tfrac{2}{3}\big)^n t \sum_{k=1}^{\infty}\big(\tfrac{3}{4}\big)^k$$
$$+ \pi\overline{\sigma}_{\kappa_1}^2 t \sum_{m=1}^{\infty}\sum_{k=1}^{\infty} 2^{-m}\big(\tfrac{2}{3}\big)^k,$$

so (4.42) follows from the inequalities $\upsilon(2/3)^n\overline{\sigma}_\kappa \leq \upsilon 2^{-n/2}\overline{\sigma}_\kappa \leq (\upsilon^2 2^{-n} + \overline{\sigma}_\kappa^2)/2$.

As before, $|\min\{a,0\} - \min\{b,0\}| \leq |a-b|$ for $a,b \in \mathbb{R}$, yields the inequality

$$\mathbb{E}\big[\big(\underline{Z}_{n,t}^{(\kappa_1)} - \underline{Z}_{n,t}^{(\kappa_2)}\big)^4\big] \leq \mathbb{E}\Bigg[\bigg(\big|\underline{\xi}_{1,n+1} - \underline{\xi}_{2,n+1}\big| + \sum_{k=1}^{n}\big|\xi_{1,k} - \xi_{2,k}\big|\bigg)^4\Bigg]. \tag{4.45}$$

By Jensen's inequality, $\mathbb{E}[|\vartheta|^3] \leq \mathbb{E}[\vartheta^4]^{3/4}$ and $\mathbb{E}[\vartheta] \leq \sqrt{\mathbb{E}[\vartheta^2]}$ for any random variable ϑ. Hence, we may bound the first and third conditional moments of $|\xi_{1,k} - \xi_{2,k}|$ and $|\underline{\xi}_{1,n+1} - \underline{\xi}_{2,n+1}|$ given $\{\ell_k\}_{k=1}^n$ and L_n. Thus, by expanding (4.45), conditioning on $\{\ell_k\}_{k=1}^n$ and L_n, and using elementary estimates as in all the previously developed bounds, we obtain (4.43). □

In order to control the level variances of the MLMC estimator in (4.53) for discontinuous payoffs of $\underline{\chi}_t$ and functions of $\underline{\tau}_t$, we would need to apply Lemma 4.18 to the components of $\big(\underline{\chi}_{n,t}^{(\kappa_1)}, \underline{\chi}_{n,t}^{(\kappa_2)}\big)$ constructed in SBG-Alg. In particular, the assumption in Lemma 4.18 requires a control on the constants in the locally Lipschitz property of the distribution functions of the various components of $\big(\underline{\chi}_{n,t}^{(\kappa_1)}, \underline{\chi}_{n,t}^{(\kappa_2)}\big)$ in terms of the cutoff levels κ_1 and κ_2. As such a uniform bound in the cutoff level appears to be out of reach, we establish Proposition 4.25, which allows us to compare the sampled quantities $\underline{\chi}_{n,t}^{(\kappa_1)}$ and $\underline{\chi}_{n,t}^{(\kappa_2)}$ with their limit $\underline{\chi}_t$ (as $\kappa_1, \kappa_2 \to 0$). Since, under mild assumptions, the distribution functions of the components of the limit $\underline{\chi}_t$ possess the necessary regularity and do not depend on the cutoff level, the application of Lemma 4.18 in the proof of Theorem 4.22 becomes feasible using Proposition 4.25.

Proposition 4.25. *There is a coupling between* $\underline{\chi}_t = (X_t, \underline{X}_t, \underline{\tau}_t)$ *and* $\big(\underline{\chi}_{n,t}^{(\kappa_1)}, \underline{\chi}_{n,t}^{(\kappa_2)}\big) \sim \underline{\Pi}_{n,t}^{\kappa_1,\kappa_2}$ *such that for any* $i \in \{1,2\}$ *and* $p \geq 1$*, the vector* $(Z_{n,t}^{(\kappa_i)}, \underline{Z}_{n,t}^{(\kappa_i)}, \underline{\tau}_{n,t}^{(\kappa_i)}) = \underline{\chi}_{n,t}^{(\kappa_i)}$ *satisfies*

$$\mathbb{E}\big[\big(X_t - Z_{n,t}^{(\kappa_i)}\big)^2\big] \leq (4\sigma^2 2^{-n} \cdot \mathbb{1}_{\{1\}}(i) + 2\overline{\sigma}_{\kappa_i}^2)t, \tag{4.46}$$
$$\mathbb{E}\big[\big(\underline{X}_t - \underline{Z}_{n,t}^{(\kappa_i)}\big)^2\big] \leq (48\sigma^2 2^{-n} \cdot \mathbb{1}_{\{1\}}(i) + 42\overline{\sigma}_{\kappa_i}^2)t. \tag{4.47}$$

Moreover, if $\delta \in (0,2]$ satisfies Assumption (O-δ), we have

$$\mathbb{E}\big[\big|\underline{\tau}_t - \underline{\tau}_{n,t}^{(\kappa_i)}\big|^p\big] \le 2^{-n}t^p + t^{p-1}\theta(t,\kappa_i), \tag{4.48}$$

where, given $T \ge t$, there exists a constant $C > 0$ dependent only on (T, σ^2, ν, b) such that for all $\kappa \in (0,1]$, the function $\theta(t,\kappa)$ is defined as

$$\theta(t,\kappa) = \begin{cases} (1 \wedge \sqrt{C\overline{\sigma}_\kappa})t, & \delta = 2, \\ t \wedge (C\overline{\sigma}_\kappa)^{\frac{2\delta}{2-\delta}} + \frac{4\delta}{5\delta-2}\sqrt{C\overline{\sigma}_\kappa}\big(t^{\frac{5\delta-2}{4\delta}} - t^{\frac{5\delta-2}{4\delta}} \wedge (C\overline{\sigma}_\kappa)^{\frac{5\delta-2}{4-2\delta}}\big), & \delta \in (0,2) \setminus \{\frac{2}{5}\}, \\ t \wedge \sqrt{C\overline{\sigma}_\kappa} + \sqrt{C\overline{\sigma}_\kappa}\log^+\big(t/\sqrt{C\overline{\sigma}_\kappa}\big), & \delta = \frac{2}{5}. \end{cases}$$

As a simple consequence of (4.48) (with $p = 1$) in Proposition 4.25 and the inequality $|\underline{\tau}_{n,t}^{(\kappa_1)} - \underline{\tau}_{n,t}^{(\kappa_2)}| \le t$, we deduce that the coupling in SBG-Alg satisfies

$$\mathbb{E}\big[\big|\underline{\tau}_{n,t}^{(\kappa_1)} - \underline{\tau}_{n,t}^{(\kappa_2)}\big|^p\big] \le 2^{1-n}t^p + 2t^{p-1}\theta(t,\kappa_1), \quad \text{for any } p \ge 1. \tag{4.49}$$

The bounds in (4.46) and (4.47) of Proposition 4.25 imply the inequalities in (4.40) and (4.42) of Proposition 4.23(c) with slightly worse constants.

Proof. The proof and construction of the random variables is analogous to that of Proposition 4.23(c), where, for $i \in \{1,2\}$, we compare the increment $Z_s^{(\kappa_i)}$ defined in Algorithm 2 with the Lévy-Itô decomposition $X_s = bs + \sigma W_s + J_s^{1,\kappa_i} + J_s^{2,\kappa_i}$ (W is as in Algorithm 2, independent of J^{1,κ_i} and J^{2,κ_i}) over the time horizons $s \in \{\ell_1, \ldots, \ell_{n-1}\}$. Similarly, we compare the pair of vectors $(\underline{\chi}_s^{(\kappa_1)}, \underline{\chi}_s^{(\kappa_2)})$ output by Algorithm 3 with $\underline{\chi}_s$ for $s = L_n$, where we assume that the (standardised) Brownian component of X equals that of $\underline{\chi}_s^{(\kappa_2)}$ (and is thus independent of the one in $\underline{\chi}_s^{(\kappa_1)}$) and all jumps in J^{2,κ_2} are synchronously coupled.

Denote the first and fourth components of the vector $(\underline{\chi}_s^{(\kappa_1)}, \underline{\chi}_s^{(\kappa_2)})$ by $Z_s^{(\kappa_1)}$ and $Z_s^{(\kappa_2)}$, respectively. Hence, it is enough to obtain the analogous bounds and identities to those presented in parts (a) and (b) for the expectations $\mathbb{E}[(X_t - Z_t^{(\kappa_i)})^2]$, $i \in \{1,2\}$ under both couplings: $\Pi_t^{\kappa_1,\kappa_2}$ and $\underline{\Pi}_t^{\kappa_1,\kappa_2}$. Such bounds may be obtained using the proofs of parts (a) and (b), resulting in the following: for $i \in \{1,2\}$, we have

$$\mathbb{E}\big[\big(X_t - Z_t^{(\kappa_i)}\big)^2\big] = \Big[\Big(\sqrt{\sigma^2 + \overline{\sigma}_{\kappa_i}^2} - \sigma\Big)^2 + \overline{\sigma}_{\kappa_i}^2\Big]t \le 2\overline{\sigma}_{\kappa_i}^2 t, \quad \text{under } \Pi_t^{\kappa_1,\kappa_2}, \tag{4.50}$$
$$\mathbb{E}\big[\big(X_t - Z_t^{(\kappa_i)}\big)^2\big] = 2(\sigma^2 \cdot \mathbb{1}_{\{1\}}(i) + \overline{\sigma}_{\kappa_1}^2)t, \quad \text{under } \underline{\Pi}_t^{\kappa_1,\kappa_2}.$$

Thus Doob's martingale inequality and elementary inequalities give (4.46) and (4.47).

By the construction of the law $\underline{\Pi}_{n,t}^{\kappa_1,\kappa_2}$ in SBG-Alg, there exist random variables

$(\xi_k')_{k=1}^n$ such that for $k \in \{1, \dots, n\}$, conditional on $\ell_k = s$ and independently of $\{\ell_j\}_{j \neq k}$, the distributional equality $(\xi_k', \xi_{1,k}, \xi_{2,k}) \overset{d}{=} (X_s, Z_s^{(\kappa_1)}, Z_s^{(\kappa_2)})$ holds, where $(Z_t^{(\kappa_1)}, Z_t^{(\kappa_2)}) \sim \Pi_t^{\kappa_1,\kappa_2}$ and W in Algorithm 2 equals the Brownian component of X in (4.4). Note that by (4.2) we have

$$\big|\underline{\tau}_t - \underline{\tau}_{n,t}^{(\kappa_i)}\big| \le L_n + \sum_{k=1}^n \ell_k \big|\mathbb{1}\{\xi_k' < 0\} - \mathbb{1}\{\xi_{i,k} < 0\}\big|, \quad \text{for } i \in \{1,2\}. \tag{4.51}$$

Let $\delta \in (0,2]$ be as in the statement of the proposition. By [86, Thm 3.1(a)], as in the proof of Theorem 4.3, we know that the density f_t of X_t exists, is smooth and, given $T > 0$, the constant $C' = 2^{3/2} \sup_{(s,x)\in(0,T]\times\mathbb{R}} s^{1/\delta} f_s(x)$ is finite. Thus, (4.33) in Lemma 4.18 (with constants $\gamma = 1$ & $C = 2^{-3/2}\ell_k^{-1/\delta} C'$ and $M = 1$, $K = 0$ & $p = 1$) gives

$$\begin{aligned}\mathbb{E}\big[\big|\mathbb{1}\{\xi_k' < 0\} - \mathbb{1}\{\xi_{i,k} < 0\}\big|\big|\ell_k\big] &\le \min\big\{1, 2^{-1/4}\sqrt{C'}\ell_k^{-\frac{1}{2\delta}}\mathbb{E}\big[|\xi_k' - \xi_{i,k}|\big|\ell_k\big]^{1/2}\big\} \\ &\le \min\big\{1, 2^{-1/4}\sqrt{C'}\ell_k^{-\frac{1}{2\delta}}(2\overline{\sigma}_{\kappa_i}^2\ell_k)^{1/4}\big\},\end{aligned}$$

for any $i \in \{1,2\}$ and $k \in \{1,\dots,n\}$, where the second inequality follows from Jensen's inequality and (4.50). Hence, elementary inequalities, together with (4.51) and Lemma 4.14 imply the following: for $i \in \{1,2\}$,

$$\begin{aligned}\mathbb{E}|\underline{\tau}_t - \underline{\tau}_{n,t}^{(\kappa_i)}| &\le \mathbb{E}L_n + \sum_{k=1}^n \mathbb{E}\big[\ell_k\big|\mathbb{1}\{\xi_k' < 0\} - \mathbb{1}\{\xi_{i,k} < 0\}\big|\big] \\ &\le 2^{-n}t + \sum_{k=1}^\infty \mathbb{E}\big[\min\big\{\sqrt{C'\overline{\sigma}_{\kappa_i}}\ell_k^{\frac{5}{4}-\frac{1}{2\delta}}, \ell_k\big\}\big] \le 2^{-n}t + \theta(t,\kappa_i).\end{aligned}$$

For $p > 1$, the result follows from the case $p = 1$ and the inequality $|\underline{\tau}_t - \underline{\tau}_{n,t}^{(\kappa_i)}|^p \le t^{p-1}|\underline{\tau}_t - \underline{\tau}_{n,t}^{(\kappa_i)}|$. □

Proof of Theorem 4.22. (a) Proposition 4.23(c) and elementary inequalities yield the bound in (4.35), so it remains to consider the case $f \in \mathrm{locLip}_K(\mathbb{R}^2)$. As in the proof of Proposition 4.6, by the inequality in (4.31) and the Cauchy-Schwarz inequality, we have

$$\mathbb{E}\big[\big(f(Z_{n,T}^{(\kappa_1)}, \underline{Z}_{n,T}^{(\kappa_1)}) - f(Z_{n,T}^{(\kappa_2)}, \underline{Z}_{n,T}^{(\kappa_2)})\big)^2\big]^2 \le K'\mathbb{E}\big[(|Z_{n,T}^{(\kappa_1)} - Z_{n,T}^{(\kappa_2)}| + |\underline{Z}_{n,T}^{(\kappa_1)} - \underline{Z}_{n,T}^{(\kappa_2)}|)^4\big],$$

where $K' = K^4\mathbb{E}[(\exp(Z_{n,T}^{(\kappa_1)}) + \exp(Z_{n,T}^{(\kappa_2)}))^4] \le 8\mathbb{E}[\exp(4X_T^{(\kappa_1)}) + \exp(4X_T^{(\kappa_2)})]$. Applying (4.30), we get $\mathbb{E}[\exp(4X_T^{(\kappa_i)})] \le \mathbb{E}[\exp(4X_T)]\exp(4T\overline{\sigma}_{\kappa_i}^2)$ and $\overline{\sigma}_{\kappa_i}^2 \le \overline{\sigma}_1^2$,

$i \in \{1,2\}$, where $\mathbb{E}[\exp(4X_T)]$ is finite since $\int_{[1,\infty)} \mathrm{e}^{4x}\nu(\mathrm{d}x) < \infty$. The concavity of $x \mapsto \sqrt{x}$ and Inequalities (4.41) & (4.43) Proposition 4.23(c) imply the existence of a constant $C > 0$ satisfying

$$\sqrt{\mathbb{E}\big[(|Z_{n,T}^{(\kappa_1)} - Z_{n,T}^{(\kappa_2)}| + |\underline{Z}_{n,T}^{(\kappa_1)} - \underline{Z}_{n,T}^{(\kappa_2)}|)^4\big]} \\ \le C(2/3)^{n/2} + 11T\overline{\sigma}_{\kappa_1}^2 + \sqrt{2\pi}T^{5/8}\overline{\sigma}_{\kappa_1}^{5/4}\kappa_1^{3/4} + \sqrt{5T}\overline{\sigma}_{\kappa_1}\kappa_1.$$

Inequality (4.36) then follows from the fact that $\overline{\sigma}_{\kappa_1}^{1/4}\kappa_1^{3/4} \le \max\{\overline{\sigma}_{\kappa_1}, \kappa_1\} \le \overline{\sigma}_{\kappa_1} + \kappa_1$.

(b) Let $(\underline{\chi}_T, \underline{\chi}_{n,T}^{(\kappa_1)}, \underline{\chi}_{n,T}^{(\kappa_2)})$ be coupled as in Proposition 4.25 with the notation $\underline{\chi}_T = (X_T, \underline{X}_T, \underline{\tau}_T)$ and $\underline{\chi}_{n,T}^{(\kappa_i)} = (Z_{n,T}^{(\kappa_i)}, \underline{Z}_{n,T}^{(\kappa_i)}, \underline{\tau}_{n,T}^{(\kappa_i)})$, $i \in \{1,2\}$. The triangle inequality and the inequalities $0 \le f \le M$ give

$$\mathbb{E}\big[\big(f\big(Z_{n,T}^{(\kappa_1)}, \underline{Z}_{n,T}^{(\kappa_1)}\big) - f\big(Z_{n,T}^{(\kappa_2)}, \underline{Z}_{n,T}^{(\kappa_2)}\big)\big)^2\big] \le M\mathbb{E}\big|f\big(Z_{n,T}^{(\kappa_1)}, \underline{Z}_{n,T}^{(\kappa_1)}\big) - f\big(Z_{n,T}^{(\kappa_2)}, \underline{Z}_{n,T}^{(\kappa_2)}\big)\big| \\ \le M\sum_{i=1}^{2}\mathbb{E}\big|f\big(Z_{n,T}^{(\kappa_i)}, \underline{Z}_{n,T}^{(\kappa_i)}\big) - f\big(X_T, \underline{X}_T\big)\big|.$$

Apply (4.33) in Lemma 4.18 with C and γ from Assumption (H) to $(X_T, \underline{X}_T)$ and $\big(Z_{n,T}^{(\kappa_i)}, \underline{Z}_{n,T}^{(\kappa_i)}\big)$ to get

$$\mathbb{E}\big|f(X_T, \underline{X}_T) - f\big(Z_{n,T}^{(\kappa_i)}, \underline{Z}_{n,T}^{(\kappa_i)}\big)\big| \\ \le K\mathbb{E}\big[\big|Z_{n,T}^{(\kappa_i)} - X_T\big|\big] + M(1+2/\gamma)(C^2\gamma^2\mathbb{E}\big[\big|\underline{Z}_{n,T}^{(\kappa_1)} - \underline{X}_T\big|^2\big]^\gamma)^{\frac{1}{2+\gamma}} \\ \le K\sqrt{T(4\sigma^2 2^{-n}\cdot \mathbb{1}_{\{1\}}(i) + 2\overline{\sigma}_{\kappa_i}^2)} + K''(\sigma^2 2^{-n}\cdot\mathbb{1}_{\{1\}}(i) + \overline{\sigma}_{\kappa_i}^2)^{\frac{\gamma}{2+\gamma}},$$

for any $i \in \{1,2\}$, where $K'' = M(1+2/\gamma)(48C^2\gamma^2T^\gamma)^{1/(2+\gamma)}$. In the second inequality we used the bounds (4.46) & (4.47) in Proposition 4.25. Since $\overline{\sigma}_{\kappa_1} \ge \overline{\sigma}_{\kappa_2}$, the result follows.

(c) Recall that the inequality in (4.49) follows from (4.48) of Proposition 4.25. The inequality in (4.38) in the proposition is a direct consequence of the Lipschitz property and (4.49).

(d) The proof follows along the same lines as in part (b): we apply (4.33) in Lemma 4.18 with C and γ from Assumption ($H\tau$) and bounds (4.46)–(4.48) in Proposition 4.25. □

§4.6.5 MC and MLMC estimators

In the present subsection we address the application of our previous results to estimate the expectation $\mathbb{E}[f(\underline{\chi}_T)]$ for various real-valued functions f satisfying

$\mathbb{E}[f(\underline{\chi}_T)^2] < \infty$. By definition, an estimator Υ of $\mathbb{E}[f(\underline{\chi}_T)]$ has L^2-accuracy of level $\epsilon > 0$ if $\mathbb{E}[(\Upsilon - \mathbb{E}f(\underline{\chi}_T))^2] < \epsilon^2$. We assume in this subsection that X has jumps of infinite activity, i.e. $\nu(\mathbb{R}\setminus\{0\}) = \infty$. If the jumps of X are finite activity, both Algorithm 3 and SBG-Alg are exact with the latter outperforming the former in practice by a constant factor, which is a function of the total number of jumps $T\nu(\mathbb{R}\setminus\{0\}) < \infty$, see §4.5.2 for a numerical example.

MC estimator

Pick $\kappa \in (0,1]$ and let the sequence $\underline{\chi}_T^{\kappa,i}$, $i \in \mathbb{N}$, be iid with the same distribution as $\underline{\chi}_T^{(\kappa)}$ simulated by SBG-Alg with $n \in \mathbb{N}\cup\{0\}$ sticks. Note that the choice of n does not affect the asymptotic behaviour as $\epsilon \searrow 0$ of the computational complexity $\mathcal{C}_{\mathrm{MC}}(\epsilon)$. The MC estimator based on $N \in \mathbb{N}$ independent samples is given by

$$\Upsilon_{\mathrm{MC}} = \frac{1}{N}\sum_{i=1}^{N} f\big(\underline{\chi}_T^{\kappa,i}\big). \tag{4.52}$$

The requirements on the bias and variance of the estimator Υ_{MC} (see §3.5.1), together with Theorem 4.3 and the bounds in (4.18) as well as Propositions 4.6, 4.7 & 4.9, imply Corollary 4.26. By expressing κ in terms of ϵ via Corollary 4.26 and (4.12), (4.15)–(4.16), the formulae for the expected computational complexity $\mathcal{C}_{\mathrm{MC}}(\epsilon)$ in Table 4.2 (of §4.4.2 above) follow.

Corollary 4.26. *For any $\epsilon \in (0,1)$, define κ as in (a)–(d) below and set $N = \big\lceil 2\epsilon^{-2}\mathbb{V}\big[f\big(\underline{\chi}_T^{(\kappa)}\big)\big]\big\rceil$ as in §3.5.1. Then the MC estimator Υ_{MC} of $\mathbb{E}[f(\underline{\chi}_T)]$ has L^2-accuracy of level ϵ and expected computational cost $\mathcal{C}_{\mathrm{MC}}(\epsilon)$ bounded by a constant multiple of $(1+\overline{\nu}(\kappa)T)N$.*

(a) *For any $K > 0$, $g \in \mathrm{Lip}_K(\mathbb{R}^2)$ (resp. $g \in \mathrm{locLip}_K(\mathbb{R}^2)$) and $f : (x,z,t) \mapsto g(x,z)$, set*

$$\kappa = \sup\{\kappa' \in (0,1] : 2\mu_1(\kappa',T) < \epsilon/\sqrt{2}\}$$
$$(\textit{resp. } \kappa = \sup\{\kappa' \in (0,1] : 8K^2\mu_2(\kappa',T)(1+\exp(2T\overline{\sigma}^2_{\kappa'}))\mathbb{E}[\exp(2X_T)] < \epsilon^2/2\}).$$

(b) *Pick $y < 0$ and let Assumption (H) hold for some $C, \gamma > 0$. Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is given by $f(x,z,t) = h(x)\mathbb{1}_{[y,\infty)}(z)$ where $h \in \mathrm{Lip}_K(\mathbb{R})$ and $0 \le h \le M$ for some $K, M > 0$. Then*

$$\kappa = \sup\{\kappa' \in (0,1] : M(C\gamma)^{2/(2+\gamma)}(1+2/\gamma)\mu_2(\kappa',T)^{2\gamma/(2+\gamma)} + K\mu_1(\kappa',T) < \epsilon/\sqrt{2}\}.$$

(c) *Let $\delta \in (0,2]$ satisfy Assumption (O-δ). Let $f : (x,z,t) \mapsto g(t)$, where $g \in$*

$\mathrm{Lip}_K(\mathbb{R})$, $K > 0$, *then*

$$\kappa = \sup\{\kappa' \in (0,1] : K\mu_*^\tau(\kappa', T) < \epsilon/\sqrt{2}\}.$$

(d) *Fix* $s \in (0,T)$ *and let* $\delta \in (0,2]$ *satisfy Assumption (O-δ). Then there exists a constant* $C > 0$ *such that for* $f \in \mathrm{BT}_2(s,K,M)$, $K, M > 0$, *we have*

$$\kappa = \sup\{\kappa' \in (0,1] : C\sqrt{K\mu_*^\tau(\kappa', T)} < \epsilon/\sqrt{2}\}.$$

MLMC estimator

Let $(\kappa_j)_{j\in\mathbb{N}}$ (resp. $(n_j)_{j\in\mathbb{N}\cup\{0\}}$) be a decreasing (resp. increasing) sequence in $(0,1]$ (resp. $\mathbb{N}$) such that $\lim_{j\to\infty}\kappa_j = 0$. Let $\underline{\chi}^{0,i} \overset{d}{=} \underline{\chi}_T^{(\kappa_1)}$ and $(\underline{\chi}_1^{j,i}, \underline{\chi}_2^{j,i}) \sim \underline{\Pi}_{n_j,T}^{\kappa_j,\kappa_{j+1}}$, $i,j \in \mathbb{N}$, be independent draws constructed by SBG-Alg. Then, for the parameters $m, N_0, \ldots, N_m \in \mathbb{N}$, the MLMC estimator takes the form

$$\Upsilon_{\mathrm{ML}} = \sum_{j=0}^{m} \frac{1}{N_j} \sum_{i=1}^{N_j} D_j^i, \quad \text{where} \quad D_j^i = \begin{cases} f(\underline{\chi}_2^{j,i}) - f(\underline{\chi}_1^{j,i}), & j \geq 1, \\ f(\underline{\chi}^{0,i}), & j = 0. \end{cases} \tag{4.53}$$

The bias of the MLMC estimator is equal to that of the MC estimator in (4.52) with $\kappa = \kappa_m$. Given the sequences $(n_j)_{j\in\mathbb{N}\cup\{0\}}$ and $(\kappa_j)_{j\in\mathbb{N}}$, which determine the simulation algorithms used in estimator (4.53), §3.5.2 derives the asymptotically optimal (as $\epsilon \searrow 0$) values for the integers m and $(N_j)_{j=0}^m$ minimising the expected computational complexity of (4.53) under the constraint that the L^2-accuracy of Υ_{ML} is of level ϵ. The key quantities are the bounds $B(j)$, $V(j)$ and $C(j)$ on the bias, level variance and the computational complexity of SBG-Alg at level j (i.e. run with parameters κ_j and n_j). The number of levels m in (4.53) is determined by the bound on the bias $B(j)$, while the number of samples N_j used at level j is given by the bounds on the complexity and level variances, see the formulae in (3.44)–(3.45). Proposition 4.27, which is a consequence of Theorem 4.3 and Propositions 4.6, 4.7 & 4.9 (for bias), Theorem 4.22 (for level variance) and Corollary 4.13 (for complexity), summarises the relevant bounds $B(j)$, $V(j)$ and $C(j)$ established in this chapter (suppressing the unknown constants as we are only interested in the asymptotic behaviour as $\epsilon \searrow 0$).

Proposition 4.27. *Given sequences* $(\kappa_j)_{j\in\mathbb{N}}$ *and* $(n_j)_{j\in\mathbb{N}\cup\{0\}}$ *as above, set* $C(j) = n_j + \overline{\nu}(\kappa_{j+1})T$. *The following choices of functions* B *and* V *ensure that, for any* $\epsilon > 0$, *the MLMC estimator* Υ_{ML}, *with integers* m *and* $\{N_j\}_{j=0}^m$ *given by* (3.44)-(3.45), *has* L^2*-accuracy of level* ϵ *with complexity asymptotically proportional to*

$\mathcal{C}_{\mathrm{ML}}(\epsilon) = 2\epsilon^{-2}\big(\sum_{j=0}^m \sqrt{C(j)V(j)}\big)^2$.

(a) *If* $K>0$, $g \in \mathrm{Lip}_K(\mathbb{R}^2)$ *(resp.* $g \in \mathrm{locLip}_K(\mathbb{R}^2)$*) and* $f:(x,z,t)\mapsto g(x,z)$*, then for any* $j\in\mathbb{N}$,

$$B(j) = \mu_1(\kappa_j, T) \quad \textit{and} \quad V(j) = \sigma^2 2^{-n_j} + \overline{\sigma}^2_{\kappa_j},$$
$$(\textit{resp. } B(j) = \mu_2(\kappa_j, T) \quad \textit{and} \quad V(j) = (2/3)^{n_j/2}\cdot \mathbb{1}_{\mathbb{R}\setminus\{0\}}(\sigma) + \overline{\sigma}^2_{\kappa_j} + \overline{\sigma}_{\kappa_j}\kappa_j).$$

(b) *Pick* $y<0$ *and let Assumption (H) hold for some* $C,\gamma>0$*. If* $f\in \mathrm{BT}_1(y,K,M)$, $K,M>0$*, then for any* $j\in\mathbb{N}$,

$$B(j) = \min\{\mu_1(\kappa_j,T)^{\gamma/(1+\gamma)}, \mu_2(\kappa_j,T)^{2\gamma/(2+\gamma)}\} \quad \textit{and}$$
$$V(j) = \sigma^{2\gamma/(2+\gamma)} 2^{-n_j\gamma/(2+\gamma)} + \overline{\sigma}^{2\gamma/(2+\gamma)}_{\kappa_j}.$$

(c) *Let Assumption (O-*δ*) hold for some* $\delta\in(0,2]$ *and* $f:(x,z,t)\mapsto g(t)$ *for some* $g\in\mathrm{Lip}_K(\mathbb{R})$, $K>0$*, then for any* $j\in\mathbb{N}$,

$$B(j) = \mu^\tau_*(\kappa_j,T) \quad \textit{and} \quad V(j) = 2^{-n_j} + \overline{\sigma}^{\min\{1/2, 2\delta/(2-\delta)\}}_{\kappa_j}(1+|\log\kappa_j|\cdot \mathbb{1}_{\{\delta=\frac{2}{5}\}}).$$

(d) *Let* $f\in\mathrm{BT}_2(s,K,M)$ *for some* $s\in(0,T)$ *and* $K,M\geq 0$*. If* $\delta\in(0,2]$ *satisfies Assumption (O-*δ*), then for any* $j\in\mathbb{N}$,

$$B(j) = \sqrt{\mu^\tau_*(\kappa_j,T)} \quad \textit{and} \quad V(j) = 2^{-n_j/2} + \overline{\sigma}^{\min\{1/4,\delta/(2-\delta)\}}_{\kappa_j}\big(1+\sqrt{|\log\kappa_j|}\cdot\mathbb{1}_{\{\delta=\frac{2}{5}\}}\big).$$

Remark 4.28. By (4.12) and (3.45) we note that κ_m in Proposition 4.27(a) is bounded by (and typically proportional to) $C_0\epsilon/|\log\epsilon|$. Moreover, if $\kappa_m = \mathrm{e}^{-r(m-1)}$ for some $r>0$, then the constant C_0 does not depend on the rate r. A similar statement holds for (b), (c) and (d), see Table 4.2 above.

It remains to choose the parameters $(n_j)_{j\in\mathbb{N}\cup\{0\}}$ and $(\kappa_j)_{j\in\mathbb{N}}$ for the estimator in (4.53). Since we require the bias to vanish geometrically fast, we set $\kappa_j = \mathrm{e}^{-r(j-1)}$ for $j\in\mathbb{N}$ and some $r>0$. The value of the rate r in Theorem 4.29 below is obtained by minimising the multiplicative constant in the complexity $\mathcal{C}_{\mathrm{ML}}(\epsilon)$. Note that n_j does not affect the bias (nor the bound $B(j)$) of Υ_{ML}. By Proposition 4.27, n_j may be as small as a multiple of $\log(1/\overline{\sigma}^2_{\kappa_j})$ without affecting the asymptotic behaviour of the level variances $V(j)$ and as large as $\overline{\nu}(\kappa_{j+1})$ without increasing the asymptotic behaviour of the cost of each level $C(j)$. Moreover, to ensure that the term $\sigma^2 2^{-n_j}$ in the level variances (see Theorem 4.22 above) decays geometrically, it suffices to let n_j grow at least linearly in j. In short, there is large interval within which we may choose n_j without it having any effect on the asymptotic performance of the MLMC estimation (see Theorem 4.29 below). The choice $n_j = n_0 + \lceil \max\{j, \log^2(1+\overline{\nu}(\kappa_{j+1})T)\}\rceil$, for $j \in \mathbb{N}$, in the numerical examples of §4.5 fall within this interval (recall $\lceil x \rceil = \inf\{j \in \mathbb{Z} : j \geq x\}$ for $x \in \mathbb{R}$).

Theorem 4.29. *Suppose $q \in (0,2]$ and $c > 0$ satisfy $\overline{\nu}(\kappa) \leq c\kappa^{-q}$ and $\overline{\sigma}_\kappa^2 \leq c\kappa^{2-q}$ for all $\kappa \in (0,1]$. Pick $r > 0$, set $\kappa_j = \mathrm{e}^{-r(j-1)}$ and assume that $\max\{j, \log_{2/3}(\overline{\sigma}_{\kappa_j}^4)\} \leq n_j \leq C\overline{\nu}(\kappa_{j+1})$ for some $C > 0$ and all sufficiently large $j \in \mathbb{N}$. Then, in cases (a)–(d) below, there exists a constant $C_r > 0$ such that, for all $\epsilon \in (0,1)$, the MLMC estimator Υ_{ML} defined in (4.53), with parameters given by (3.44)-(3.45), is L^2-accurate at level ϵ with the stated expected computational complexity $\mathcal{C}_{\mathrm{ML}}(\epsilon)$. Moreover, C_r is minimal for $r = (2/|a|)\log(1+|a|/q) \cdot \mathbb{1}_{\mathbb{R}\setminus\{0\}}(a) + (2/q)\cdot\mathbb{1}_{\{0\}}(a)$, with $a \in \mathbb{R}$ given explicitly in each case (a)–(d).*

(a) *Let $g \in \mathrm{Lip}_K(\mathbb{R}^2) \cup \mathrm{locLip}_K(\mathbb{R}^2)$ for $K > 0$ and $f : (x,z,t) \mapsto g(x,z)$. Define $a = 2(q-1)$ and $b = \mathbb{1}_{\{\sigma=0\}} + \mathbb{1}_{\{\sigma\neq 0\}}(\mathbb{1}_{\{g\in\mathrm{Lip}_K(\mathbb{R}^2)\}}\cdot\frac{1}{3-q} + \mathbb{1}_{\{g\notin\mathrm{Lip}_K(\mathbb{R}^2)\}}\cdot\frac{2}{4-q})$, then*

$$\mathcal{C}_{\mathrm{ML}}(\epsilon) \leq \frac{C_r}{\epsilon^{2+a^+b}}\big(1+\log^2\epsilon\cdot\mathbb{1}_{\{1\}}(q) + |\log\epsilon|^{(a/2)(1+\mathbb{1}\{g\in\mathrm{Lip}_K(\mathbb{R}^2)\})}\cdot\mathbb{1}_{(1,2]}(q)\big). \quad (4.54)$$

(b) *Let $f : (x,z,t) \mapsto g(x,z)$ where $g \in \mathrm{BT}_1(y,K,M)$ for some $y < 0$ and $K, M \geq 0$, such that (H) is satisfied by y and some $C, \gamma > 0$. Define $a = 2\frac{q(1+\gamma)-\gamma}{2+\gamma} \in (-\frac{2\gamma}{2+\gamma}, 2]$ and $b = (1/2+1/\gamma)(\mathbb{1}_{\{\sigma=0\}} + \mathbb{1}_{\{\sigma\neq0,q<1\}}\cdot\frac{4}{9-3q} + \mathbb{1}_{\{\sigma\neq0,q\geq1\}}\cdot\frac{2}{4-q})$, then*

$$\begin{aligned}\mathcal{C}_{\mathrm{ML}}(\epsilon) \leq \frac{C_r}{\epsilon^{2+a^+b}}\big(&1+\log^2\epsilon\cdot\mathbb{1}_{\{q=\frac{\gamma}{1+\gamma}\}} + |\log\epsilon|^a\cdot\mathbb{1}_{(\frac{\gamma}{1+\gamma},1)}(q)\\ &+ |\log\epsilon|^{a/2}\cdot\mathbb{1}_{[1,2]}(q)\big),\end{aligned} \quad (4.55)$$

(c) *Let $f : (x,z,t) \mapsto g(t)$ where $g \in \mathrm{Lip}_K(\mathbb{R})$, $K > 0$, and let (O-δ) hold for some $\delta \in (0,2]$. Define $a = q - (1-\frac{q}{2})\min\{\frac{1}{2}, \frac{2\delta}{2-\delta}\}$ and $b = \min\{\frac{2}{\delta}, \max\{\frac{3}{2}, \frac{1}{\delta}\}\}$, then*

$$\mathcal{C}_{\mathrm{ML}}(\epsilon) \leq \frac{C_r}{\epsilon^{2+a^+b}}\begin{cases} 1+\log^2\epsilon\cdot\mathbb{1}_{\{q=\delta\in(0,\frac{2}{5})\}\cup\{\delta=2,q=\frac{2}{5}\}}, & \delta\in(0,2]\setminus\{\frac{2}{5},\frac{2}{3}\},\\ |\log\epsilon|\cdot\mathbb{1}_{(2/5,2]}(q) + |\log\epsilon|^3\cdot\mathbb{1}_{\{q=\frac{2}{5}\}}, & \delta=\frac{2}{5},\\ |\log\epsilon|^a, & \delta=\frac{2}{3}.\end{cases} \quad (4.56)$$

(d) *Fix $s \in (0,T)$ and let $\delta \in (0,2]$ satisfy (O-δ). Define $a = q-(1-\frac{q_2}{2})\min\{\frac{1}{4},\frac{\delta}{2-\delta}\}$ and $b = \min\{\frac{4}{\delta}, \max\{3, \frac{2}{\delta}\}\}$, then for any $K, M \geq 0$ and $f \in \mathrm{BT}_2(s,K,M)$, we have*

$$\mathcal{C}_{\mathrm{ML}}(\epsilon) \leq \frac{C_r}{\epsilon^{2+a^+b}}\begin{cases} 1+\log^2\epsilon\cdot\mathbb{1}_{\{q=\frac{2}{9}\}}, & \delta=2,\\ 1+\sqrt{|\log\epsilon|}\cdot\mathbb{1}_{\{\delta=\frac{2}{5}\}} + |\log\epsilon|^{\frac{a}{2}}\cdot\mathbb{1}_{\{\delta=\frac{2}{3}\}}, & \delta\in(0,2).\end{cases} \quad (4.57)$$

Remark 4.30. For most models either $\beta = \delta$ or $\sigma > 0$, implying $a^+b \in [0,2]$ in parts (a) and (c), $a^+b \in [0, 2(1/2 + 1/\gamma)]$ in part (b) (with γ typically equal to 1) and $a^+b \in [0,4]$ in part (d).

Proof of Theorem 4.29. Note that $\kappa_1 = 1$ by definition independently or $r > 0$, thus making both the variance $\mathbb{V}[D_0^i]$ and the cost of sampling of D_0^i independent of r. We may thus ignore the 0-th term in the bound $\epsilon^{-2}(\sum_{j=0}^m \sqrt{V(j)C(j)})^2$ on the complexity $\mathcal{C}_{\mathrm{ML}}(\epsilon)$ derived in §3.5.2. Since m is given by (3.44), by Table 4.2 and Remark 4.28, the function $\overline{m} : (0,1) \mapsto (0,\infty)$ given by

$$\overline{m}(\epsilon) = \begin{cases} (b|\log\epsilon| + c\log|\log\epsilon|)/r, & \text{in (a) \& (b) and, if } \delta = \frac{2}{3}, \text{ in (c) \& (d)}, \\ b|\log\epsilon|/r & \text{in (c) \& (d) if } \delta \neq \frac{2}{3}, \end{cases}$$

$$\text{where} \quad c = \begin{cases} 1, & \text{in (a) \& (c)}, \\ 1/2, & \text{in (b) \& (d)}, \end{cases}$$

satisfies $m \le \overline{m}(\epsilon) + C'/r$ for all $\epsilon \in (0,1)$ and $r > 0$, where the constant $C' > 0$ is independent of $r > 0$. Thus, we need only study the growth rate of

$$\phi(\epsilon) = \sum_{j=1}^{\lceil \overline{m}(\epsilon) \rceil} \sqrt{C(j)V(j)} = \sum_{j=1}^{\lceil \overline{m}(\epsilon) \rceil} \sqrt{(n_j + \overline{\nu}(\kappa_{j+1})T)V(j)}, \quad \text{as } \epsilon \to 0,$$

because $\mathcal{C}_{\mathrm{ML}}(\epsilon)$ is bounded by a constant multiple of $\epsilon^{-2}\phi(\epsilon)^2$. In the cases where $V(j)$ contains a term of the form e^{-sn_j} for some $s > 0$ (only possible if $\sigma \neq 0$), the product $n_j\mathrm{e}^{-sn_j} \le \mathrm{e}^{-sn_j/2}$ vanishes geometrically fast since $n_j \ge j$ for all large j. Thus, the corresponding component in $\phi(\epsilon)$ is bounded as $\epsilon \to 0$ and may thus be ignored. By Proposition 4.27, in all cases we may assume that $V(j)$ is bounded by a multiple of a power of $\overline{\sigma}^2_{\kappa_j}$ and $C(j)$ is dominated by a multiple of $\overline{\nu}(\kappa_{j+1})$.

Since $\overline{\nu}(\kappa) \le c\kappa^{-q}$ and $\overline{\sigma}^2_\kappa \le c\kappa^{2-q}$ for $\kappa \in (0,1]$, Proposition 4.27 implies

$$\phi(\epsilon) \le K_* \begin{cases} \sum_{j=1}^{\lceil \overline{m}(\epsilon) \rceil} \sqrt{\kappa_{j+1}^{-q}\kappa_j^{2-q}}, & \text{in (a)}, \\ \sum_{j=1}^{\lceil \overline{m}(\epsilon) \rceil} \sqrt{\kappa_{j+1}^{-q}\kappa_j^{(2-q)\gamma/(2+\gamma)}}, & \text{in (b)}, \\ \sum_{j=1}^{\lceil \overline{m}(\epsilon) \rceil} \sqrt{\kappa_{j+1}^{-q}\kappa_j^{(2-q)\min\{1/2, 2\delta/(2-\delta)\}}(1 + |\log\kappa_j|\mathbb{1}_{\{2/5\}}(\delta))}, & \text{in (c)}, \\ \sum_{j=1}^{\lceil \overline{m}(\epsilon) \rceil} \sqrt{\kappa_{j+1}^{-q}\kappa_j^{(2-q)\min\{1/4, \delta/(2-\delta)\}}(1 + \sqrt{|\log\kappa_j|}\mathbb{1}_{\{2/5\}}(\delta))}, & \text{in (d)}, \end{cases}$$

for some constant $K_* > 0$ independent of r and all $\epsilon \in (0,1)$, where in part (a) we used the fact that $\overline{\sigma}_\kappa \kappa \le \sqrt{c}\kappa^{2-q/2}$ for all $\kappa \in (0,1]$.

(a) Recall that $\kappa_j = \mathrm{e}^{-r(j-1)}$ and $\kappa_{j+1} = \mathrm{e}^{-r(j-1)-r}$, implying

$$\kappa_{j+1}^{-q}\kappa_j^{2-q} = \mathrm{e}^{rq}\mathrm{e}^{ar(j-1)}, \qquad \text{for all } j \in \mathbb{N}, \text{ where } a = 2(q-1), \tag{4.58}$$

Suppose $a < 0$, implying $q \in (0,1)$. By (4.58), the sequence $(\kappa_{j+1}^{-q}\kappa_j^{2-q})_{j\in\mathbb{N}}$ decays geometrically fast. This implies that $\lim_{\epsilon\downarrow 0}\phi(\epsilon) < \infty$ and gives the desired result. Moreover, the leading constant C_r, as a function of r, is proportional to $\mathrm{e}^{rq}/(1-\mathrm{e}^{ar/2})^2$ as $\epsilon \downarrow 0$. Since $a \neq 0$ for $q \in (0,1)$, the minimal value of C_r is attained when $r = (2/|a|)\log(1+|a|/q)$.

Suppose $a = 0$, implying $q = 1$. By (4.58) and the definition of $\overline{m}(\epsilon)$, $\phi(\epsilon) \leq K_*\mathrm{e}^{r/2}(b|\log\epsilon| + \log|\log\epsilon|)/r$, giving the desired result. As before, the leading constant C_r, as a function of r is proportional to e^r/r^2 as $\epsilon \to 0$, attaining its minimum at $r = 2$.

Suppose $a > 0$, implying $q \in (1,2]$. By (4.58) and the definition of $\overline{m}(\epsilon)$, it similarly follows that

$$\phi(\epsilon)^2 \leq \frac{K_*^2\mathrm{e}^{rq}}{(\mathrm{e}^{ar/2}-1)^2}\mathrm{e}^{a(b|\log\epsilon|+\log|\log\epsilon|)} = \frac{K_*^2\mathrm{e}^{rq}}{(\mathrm{e}^{ar/2}-1)^2}\epsilon^{-ab}|\log\epsilon|^a.$$

The corresponding result follows easily, where the leading constant C_r, as a function of r, is proportional to $\mathrm{e}^{rq}/(\mathrm{e}^{ar/2}-1)^2$ as $\epsilon \downarrow 0$ and attains its minimum at $r = (2/a)\log(1+a/q)$, concluding the proof of (a).

(b) As before, we have

$$\kappa_{j+1}^{-q}\kappa_j^{(2-q)\gamma/(2+\gamma)} = \mathrm{e}^{rq}\mathrm{e}^{ar(j-1)}, \qquad \text{for all } j \in \mathbb{N}, \text{ where } a = 2\frac{q(1+\gamma)-\gamma}{2+\gamma}. \tag{4.59}$$

Suppose $a < 0$, implying $q < \gamma/(1+\gamma)$. Then $\lim_{\epsilon\downarrow 0}\phi(\epsilon) < \infty$ by (4.59), implying the claim. Moreover, $r = (2/|a|)\log(1+|a|/q)$ minimises C_r as in part (a).

Suppose $a = 0$, implying $q = \gamma/(1+\gamma)$. Then we have the bound $\phi(\epsilon)^2 \leq K_*^2 r^{-2}\mathrm{e}^{rq}(b|\log\epsilon| + \log|\log\epsilon|/2)^2$, and $r = 2/q = 2 + 2/\gamma$ minimises the leading constant.

Suppose $a > 0$, implying $q > \gamma/(1+\gamma)$. By (4.59), we have

$$\phi(\epsilon)^2 \leq \frac{K_*^2\mathrm{e}^{rq}}{(\mathrm{e}^{ar/2}-1)^2}\mathrm{e}^{a(b|\log\epsilon|+\log|\log\epsilon|/2)} = \frac{K_*^2\mathrm{e}^{rq}}{(\mathrm{e}^{ar/2}-1)^2}\epsilon^{-ab}|\log\epsilon|^{a/2},$$

and the leading constant is minimal for $r = (2/a)\log(1+a/q)$.

In parts (c) and (d), note that $a < 0$ if and only if $\delta = 2$ (i.e. $\sigma \neq 0$). Analogous arguments as in (a) and (b), complete the proof of the theorem. $\square$

Bibliography

[1] Josh Abramson and Jim Pitman. Concave majorants of random walks and related Poisson processes. *Combin. Probab. Comput.*, 20(5):651–682, 2011. ISSN 0963-5483. doi: 10.1017/S0963548311000307. URL http://dx.doi.org/10.1017/S0963548311000307.

[2] Josh Abramson, Jim Pitman, Nathan Ross, and Gerónimo Uribe Bravo. Convex minorants of random walks and Lévy processes. *Electron. Commun. Probab.*, 16:423–434, 2011. ISSN 1083-589X. URL https://doi.org/10.1214/ECP.v16-1648.

[3] Leif Andersen and Alexander Lipton. Asymptotics for exponential Lévy processes and their volatility smile: survey and new results. *Int. J. Theor. Appl. Finance*, 16(1):1350001, 98, 2013. ISSN 0219-0249. doi: 10.1142/S0219024913500015. URL https://doi.org/10.1142/S0219024913500015.

[4] Søren Asmussen. *Applied probability and queues*, volume 51 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. ISBN 0-387-00211-1. Stochastic Modelling and Applied Probability.

[5] Søren Asmussen and Hansjörg Albrecher. *Ruin probabilities*, volume 14 of *Advanced Series on Statistical Science & Applied Probability*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, second edition, 2010. ISBN 978-981-4282-52-9; 981-4282-52-9. doi: 10.1142/9789814282536. URL https://doi.org/10.1142/9789814282536.

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

[7] Søren Asmussen, Peter Glynn, and Jim Pitman. Discretization error in simulation of one-dimensional reflecting Brownian motion. *Ann. Appl. Probab.*,

5(4):875–896, 1995. ISSN 1050-5164. URL http://links.jstor.org/sici?sici=1050-5164(199511)5:4<875:DEISOO>2.0.CO;2-D&origin=MSN.

[8] Florin Avram, Terence Chan, and Miguel Usabel. On the valuation of constant barrier options under spectrally one-sided exponential Lévy models and Carr's approximation for American puts. *Stochastic Process. Appl.*, 100:75–107, 2002. ISSN 0304-4149. doi: 10.1016/S0304-4149(02)00104-7. URL https://doi.org/10.1016/S0304-4149(02)00104-7.

[9] David Bang, Jorge I. González Cázares, and Aleksandar Mijatović. Asymptotic length of the concave majorant of a Lévy process. 2021. URL https://arxiv.org/abs/2106.09066.

[10] E.J. Baurdoux, Z. Palmowski, and M.R. Pistorius. On future drawdowns of Lévy processes. *Stochastic Processes and their Applications*, 127(8):2679–2698, 2017. doi: 10.1016/j.spa.2016.12.008. URL https://ideas.repec.org/a/eee/spapps/v127y2017i8p2679-2698.html.

[11] Erik J. Baurdoux and Kees van Schaik. Predicting the time at which a Lévy process attains its ultimate supremum. *Acta Appl. Math.*, 134:21–44, 2014. ISSN 0167-8019. doi: 10.1007/s10440-014-9867-2. URL https://doi.org/10.1007/s10440-014-9867-2.

[12] Violetta Bernyk, Robert C. Dalang, and Goran Peskir. Predicting the ultimate supremum of a stable Lévy process with no negative jumps. *Ann. Probab.*, 39(6):2385–2423, 2011. ISSN 0091-1798. URL https://doi.org/10.1214/10-AOP598.

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

[14] Patrick 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. doi: 10.1002/9780470316962. URL https://doi.org/10.1002/9780470316962. A Wiley-Interscience Publication.

[15] N. H. Bingham. Maxima of sums of random variables and suprema of stable processes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 26(4):273–296, 1973. doi: 10.1007/BF00534892. URL http://dx.doi.org/10.1007/BF00534892.

[16] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1989. ISBN 0-521-37943-1.

[17] Krzysztof Bisewski and Jevgenijs Ivanovs. Zooming-in on a lévy process: failure to observe threshold exceedance over a dense grid. *Electron. J. Probab.*, 25: 33 pp., 2020. doi: 10.1214/20-EJP513. URL https://doi.org/10.1214/20-EJP513.

[18] R. M. Blumenthal and R. K. Getoor. Sample functions of stochastic processes with stationary independent increments. *J. Math. Mech.*, 10:493–516, 1961.

[19] Sergey Bobkov and Michel Ledoux. One-dimensional empirical measures, order statistics, and Kantorovich transport distances. *Mem. Amer. Math. Soc.*, 261 (1259):v+126, 2019. ISSN 0065-9266. doi: 10.1090/memo/1259. URL https://doi.org/10.1090/memo/1259.

[20] Svetlana Boyarchenko and Sergei Levendorskiĭ. Barrier options and touch-and-out options under regular Lévy processes of exponential type. *Ann. Appl. Probab.*, 12(4):1261–1298, 2002. ISSN 1050-5164. doi: 10.1214/aoap/1037125863. URL https://doi.org/10.1214/aoap/1037125863.

[21] Mark Broadie, Paul Glasserman, and Steven Kou. A continuity correction for discrete barrier options. *Math. Finance*, 7(4):325–349, 1997. ISSN 0960-1627. doi: 10.1111/1467-9965.00035. URL https://doi.org/10.1111/1467-9965.00035.

[22] Mark Broadie, Paul Glasserman, and S. G. Kou. Connecting discrete and continuous path-dependent options. *Finance Stoch.*, 3(1):55–82, 1999. ISSN 0949-2984. doi: 10.1007/s007800050052. URL https://doi.org/10.1007/s007800050052.

[23] Chunhao Cai, Junyi Guo, and Honglong You. Non-parametric estimation for a pure-jump Lévy process. *Annals of Actuarial Science*, 12(2):338–349, 2018. doi: 10.1017/S1748499517000100.

[24] Alexandra Carpentier, Céline Duval, and Ester Mariucci. Total variation distance for discretely observed Lévy processes: a Gaussian approximation of the small jumps. *Ann. Inst. Henri Poincaré Probab. Stat.*, page to appear, 2020.

[25] Peter Carr, Hélyette Geman, Dilip B. Madan, and Marc Yor. Stochastic volatility for Lévy processes. *Math. Finance*, 13(3):345–382, 2003. ISSN 0960-1627. doi: 10.1111/1467-9965.00020. URL https://doi.org/10.1111/1467-9965.00020.

[26] Peter Carr, Hongzhong Zhang, and Olympia Hadjiliadis. Maximum drawdown insurance. *International Journal of Theoretical and Applied Finance*, 14(08): 1195–1230, 2011. doi: 10.1142/S0219024911006826. URL https://doi.org/10.1142/S0219024911006826.

[27] Jorge González Cázares, Arturo Kohatsu-Higa, and Aleksandar Mijatović. Joint density of a stable process and its supremum: regularity and upper bounds, 2020. URL https://arxiv.org/abs/2008.01894.

[28] Loïc Chaumont. On the law of the supremum of Lévy processes. *Ann. Probab.*, 41(3A):1191–1217, 2013. ISSN 0091-1798. doi: 10.1214/11-AOP708. URL https://doi.org/10.1214/11-AOP708.

[29] Loïc Chaumont and Jacek Małecki. On the asymptotic behavior of the density of the supremum of Lévy processes. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(3):1178–1195, 2016. ISSN 0246-0203. doi: 10.1214/15-AIHP674. URL https://doi.org/10.1214/15-AIHP674.

[30] Ao Chen. *Sampling error of the supremum of a Levy process*. ProQuest LLC, Ann Arbor, MI, 2011. ISBN 978-1267-15502-3. URL http://gateway.proquest.com/openurl?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&res_dat=xri:pqm&rft_dat=xri:pqdiss:3496281. Thesis (Ph.D.)–University of Illinois at Urbana-Champaign.

[31] Song X. Chen, Aurore Delaigle, and Peter Hall. Nonparametric estimation for a class of Lévy processes. *J. Econometrics*, 157(2):257–271, 2010. ISSN 0304-4076. doi: 10.1016/j.jeconom.2009.12.005. URL https://doi.org/10.1016/j.jeconom.2009.12.005.

[32] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup. Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Comput. Vis. Sci.*, 14(1):3–15, 2011. ISSN 1432-9360. doi: 10.1007/s00791-011-0160-x. URL https://doi.org/10.1007/s00791-011-0160-x.

[33] Donald L. Cohn. *Measure theory*. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks].

Birkhäuser/Springer, New York, second edition, 2013. ISBN 978-1-4614-6955-1; 978-1-4614-6956-8. doi: 10.1007/978-1-4614-6956-8. URL https://doi.org/10.1007/978-1-4614-6956-8.

[34] Fabienne Comte and Valentine Genon-Catalot. Estimation for Lévy processes from high frequency data within a long time interval. *Ann. Statist.*, 39(2):803–837, 2011. ISSN 0090-5364. doi: 10.1214/10-AOS856. URL https://doi.org/10.1214/10-AOS856.

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

[36] Rama Cont and Peter Tankov. *Financial Modelling with Jump Processes, Second Edition*. Chapman and Hall/CRC Financial Mathematics Series. Taylor & Francis, 2015. ISBN 9781420082197. URL https://books.google.co.uk/books?id=-fZtKgAACAAJ.

[37] José Manuel Corcuera, Dilip B. Madan, Jan De Spiegeleer, Andreas E. Kyprianou, and Wim Schoutens. Pricing of contingent convertibles under smile conform models. 9(3):121–140, 2013.

[38] Steffen Dereich. Multilevel Monte Carlo algorithms for Lévy-driven SDEs with Gaussian correction. *Ann. Appl. Probab.*, 21(1):283–311, 2011. ISSN 1050-5164. doi: 10.1214/10-AAP695. URL https://doi.org/10.1214/10-AAP695.

[39] Steffen Dereich and Felix Heidenreich. A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations. *Stochastic Process. Appl.*, 121 (7):1565–1587, 2011. ISSN 0304-4149. doi: 10.1016/j.spa.2011.03.015. URL https://doi.org/10.1016/j.spa.2011.03.015.

[40] Luc Devroye. On exact simulation algorithms for some distributions related to Brownian motion and Brownian meanders. In *Recent developments in applied probability and statistics*, pages 1–35. Physica, Heidelberg, 2010. doi: 10.1007/978-3-7908-2598-5_1. URL https://doi.org/10.1007/978-3-7908-2598-5_1.

[41] E. H. A. Dia. Error bounds for small jumps of Lévy processes. *Adv. in Appl. Probab.*, 45(1):86–105, 2013. ISSN 0001-8678. URL http://projecteuclid.org/euclid.aap/1363354104.

[42] E. H. A. Dia and D. Lamberton. Connecting discrete and continuous lookback or hindsight options in exponential Lévy models. *Adv. in Appl. Probab.*, 43 (4):1136–1165, 2011. ISSN 0001-8678. doi: 10.1239/aap/1324045702. URL https://doi.org/10.1239/aap/1324045702.

[43] El Hadj Aly Dia and Damien Lamberton. Continuity correction for barrier options in jump-diffusion models. *SIAM J. Financial Math.*, 2(1):866–900, 2011. ISSN 1945-497X. doi: 10.1137/100817553. URL https://doi.org/10.1137/100817553.

[44] Lucas Downey. Ulcer index (UI) definition. https://www.investopedia.com/terms/u/ulcerindex.asp. Investopedia (2009).

[45] A. Ferreiro-Castilla, A. E. Kyprianou, R. Scheichl, and G. Suryanarayana. Multilevel Monte Carlo simulation for Lévy processes based on the Wiener-Hopf factorisation. *Stochastic Process. Appl.*, 124(2):985–1010, 2014. ISSN 0304-4149. doi: 10.1016/j.spa.2013.09.015. URL https://doi.org/10.1016/j.spa.2013.09.015.

[46] José E. Figueroa-López. Small-time moment asymptotics for Lévy processes. *Statist. Probab. Lett.*, 78(18):3355–3365, 2008. ISSN 0167-7152. doi: 10.1016/j.spl.2008.07.012. URL https://doi.org/10.1016/j.spl.2008.07.012.

[47] Nicolas Fournier. Simulation and approximation of Lévy-driven stochastic differential equations. *ESAIM Probab. Stat.*, 15:233–248, 2011. ISSN 1292-8100. doi: 10.1051/ps/2009017. URL https://doi.org/10.1051/ps/2009017.

[48] Michael B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3): 607–617, 2008. ISSN 0030-364X. doi: 10.1287/opre.1070.0496. URL https://doi.org/10.1287/opre.1070.0496.

[49] Michael B. Giles and Yuan Xia. Multilevel Monte Carlo for exponential Lévy models. *Finance Stoch.*, 21(4):995–1026, 2017. ISSN 0949-2984. doi: 10.1007/s00780-017-0341-7. URL https://doi.org/10.1007/s00780-017-0341-7.

[50] Jorge González Cázares, Aleksandar Mijatović, and Gerónimo Uribe Bravo. Code for the simulation of the triplet $(X_T, \overline{X}_T, \tau_T)$. https://github.com/jorgeignaciogc/LevySupSim.jl, 2018. GitHub repository.

[51] Jorge I. González Cázares and Aleksandar Mijatović. SB-ARA. https://github.com/jorgeignaciogc/SBARA.jl, 2020. GitHub repository.

[52] Jorge I. González Cázares and Aleksandar Mijatović. Stick-breaking with small-jump Gaussian approximation for the drawdown of a lévy process. *arXiv e-prints*, art. arXiv:1810.11039, Mar 2020. URL https://arxiv.org/abs/2011.06618.

[53] Jorge I. González Cázares and Aleksandar Mijatović. Convex minorants and the fluctuation theory of Lévy processes. 2021. URL https://arxiv.org/abs/2105.15060.

[54] Jorge I. González Cázares, Aleksandar Mijatović, and Gerónimo Uribe Bravo. Exact simulation of the extrema of stable processes. *Adv. in Appl. Probab.*, 51 (4):967–993, 2019. ISSN 0001-8678. doi: 10.1017/apr.2019.39. URL https://doi.org/10.1017/apr.2019.39.

[55] Jorge I. González Cázares, Aleksandar Mijatović, and Gerónimo Uribe Bravo. ε-strong simulation of the convex minorants of stable processes and meanders. *Electron. J. Probab.*, 25:1–33, 2020. ISSN 1083-6489. doi: 10.1214/20-EJP503. URL https://arxiv.org/abs/1910.13273.

[56] Jorge Ignacio González Cázares, Aleksandar Mijatović, and Gerónimo Uribe Bravo. Geometrically convergent simulation of the extrema of Lévy processes. *Math. Oper. Res.*, 2021. doi: 10.1287/moor.2021.1163. URL https://doi.org/10.1287/moor.2021.1163.

[57] Michael Grabchak. Rejection Sampling for Tempered Levy Processes. *arXiv e-prints*, art. arXiv:1806.00671, Jun 2018.

[58] Michael Grabchak. Rejection sampling for tempered Lévy processes. *Stat. Comput.*, 29(3):549–558, 2019. ISSN 0960-3174. doi: 10.1007/s11222-018-9822-6. URL https://doi.org/10.1007/s11222-018-9822-6.

[59] S. Heinrich. Multilevel monte carlo methods. In *Large-scale scientific computing*, pages 58–67. 2001.

[60] Jevgenijs Ivanovs. Zooming in on a Lévy process at its supremum. *Ann. Appl. Probab.*, 28(2):912–940, 2018. ISSN 1050-5164. doi: 10.1214/17-AAP1320. URL https://doi.org/10.1214/17-AAP1320.

[61] Jean Jacod and Albert 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. doi: 10.1007/978-3-662-05265-5. URL https://doi.org/10.1007/978-3-662-05265-5.

[62] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002. ISBN 0-387-95313-2. doi: 10.1007/978-1-4757-4015-8. URL http://dx.doi.org/10.1007/978-1-4757-4015-8.

[63] Young Shin Kim, Svetlozar T. Rachev, Michele Leonardo Bianchi, and Frank J. Fabozzi. A new tempered stable distribution and its application to finance. In Georg Bol, Svetlozar T. Rachev, and Reinhold Würth, editors, *Risk Assessment*, pages 77–109, Heidelberg, 2009. Physica-Verlag HD. ISBN 978-3-7908-2050-8.

[64] J. F. C. Kingman. *Poisson processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press, Oxford University Press, New York, 1993. ISBN 0-19-853693-3. Oxford Science Publications.

[65] Claudia Klüppelberg, Andreas E. Kyprianou, and Ross A. Maller. Ruin probabilities and overshoots for general Lévy insurance risk processes. *Ann. Appl. Probab.*, 14(4):1766–1801, 2004. ISSN 1050-5164. URL https://doi.org/10.1214/105051604000000927.

[66] Victor Korolev and Irina Shevtsova. An improvement of the Berry-Esseen inequality with applications to Poisson and mixed Poisson random sums. *Scand. Actuar. J.*, (2):81–105, 2012. ISSN 0346-1238. doi: 10.1080/03461238.2010.485370. URL https://doi.org/10.1080/03461238.2010.485370.

[67] S. G. Kou. A jump-diffusion model for option pricing. *Management Science*, pages 1086–1101, 2002.

[68] Steven Kou. *Lévy Processes in Asset Pricing*. American Cancer Society, 2014. ISBN 9781118445112. doi: 10.1002/9781118445112.stat03738. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat03738.

[69] Oleg Kudryavtsev and Sergei Levendorskiĭ. Fast and accurate pricing of barrier options under Lévy processes. *Finance Stoch.*, 13(4):531–562, 2009. ISSN 0949-2984. doi: 10.1007/s00780-009-0103-2. URL https://doi.org/10.1007/s00780-009-0103-2.

[70] A. Kuznetsov, A. E. Kyprianou, J. C. Pardo, and K. van Schaik. A Wiener-Hopf Monte Carlo simulation technique for Lévy processes. *Ann. Appl. Probab.*,

21(6):2171–2190, 2011. ISSN 1050-5164. URL https://doi.org/10.1214/10-AAP746.

[71] Alexey Kuznetsov, Andreas E. Kyprianou, and Victor Rivero. The theory of scale functions for spectrally negative Lévy processes. In *Lévy matters II*, volume 2061 of *Lecture Notes in Math.*, pages 97–186. Springer, Heidelberg, 2012. doi: 10.1007/978-3-642-31407-0_2. URL https://doi.org/10.1007/978-3-642-31407-0_2.

[72] Mateusz Kwaśnicki, Jacek Małecki, and Michał Ryznar. Suprema of Lévy processes. *Ann. Probab.*, 41(3B):2047–2065, 2013. ISSN 0091-1798. doi: 10.1214/11-AOP719. URL https://doi.org/10.1214/11-AOP719.

[73] Andreas 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.

[74] David Landriault, Bin Li, and Hongzhong Zhang. On magnitude, asymptotics and duration of drawdowns for Lévy models. *Bernoulli*, 23(1):432–458, 2017. ISSN 1350-7265. doi: 10.3150/15-BEJ748. URL https://doi.org/10.3150/15-BEJ748.

[75] Ping Li, Wu Zhao, and Wei Zhou. Ruin probabilities and optimal investment when the stock price follows an exponential Lévy process. *Appl. Math. Comput.*, 259:1030–1045, 2015. ISSN 0096-3003. doi: 10.1016/j.amc.2014.12.042. URL https://doi.org/10.1016/j.amc.2014.12.042.

[76] Ester Mariucci and Markus Reiß. Wasserstein and total variation distance between marginals of Lévy processes. *Electron. J. Stat.*, 12(2):2482–2514, 2018. ISSN 1935-7524. doi: 10.1214/18-EJS1456. URL https://doi.org/10.1214/18-EJS1456.

[77] P.G. Martin and B.B. McCann. *The Investor's Guide to Fidelity Funds*. Wiley, 1989. ISBN 9780471622581. URL https://books.google.co.uk/books?id=nOoJAQAAMAAJ.

[78] Robert C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3:125–144, 1976.

[79] Aleksandar Mijatović and Martijn Pistorius. Buffer-overflows: joint limit laws of undershoots and overshoots of reflected processes. *Stochastic Process. Appl.*,

125(8):2937–2954, 2015. ISSN 0304-4149. doi: 10.1016/j.spa.2015.02.007. URL https://doi.org/10.1016/j.spa.2015.02.007.

[80] Aleksandar Mijatović and Martijn R. Pistorius. On the drawdown of completely asymmetric Lévy processes. *Stochastic Processes and their Applications*, 122 (11):3812 – 3836, 2012. ISSN 0304-4149. doi: https://doi.org/10.1016/j.spa.2012.06.012. URL http://www.sciencedirect.com/science/article/pii/S0304414912001408.

[81] E. Mordecki. Ruin probabilities for Lévy processes with mixed-exponential negative jumps. *Teor. Veroyatnost. i Primenen.*, 48(1):188–194, 2003. ISSN 0040-361X. doi: 10.1137/S0040585X980178. URL https://doi.org/10.1137/S0040585X980178.

[82] Ernesto Mordecki. Optimal stopping and perpetual options for Lévy processes. *Finance Stoch.*, 6(4):473–493, 2002. ISSN 0949-2984. doi: 10.1007/s007800200070. URL https://doi.org/10.1007/s007800200070.

[83] Michael H. Neumann and Markus Reiß. Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15(1):223–248, 2009. ISSN 1350-7265. doi: 10.3150/08-BEJ148. URL https://doi.org/10.3150/08-BEJ148.

[84] V. V. Petrov. *Sums of independent random variables.* Springer-Verlag, New York-Heidelberg, 1975. Translated from the Russian by A. A. Brown, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 82.

[85] Valentin V. Petrov. *Limit theorems of probability theory*, volume 4 of *Oxford Studies in Probability.* The Clarendon Press, Oxford University Press, New York, 1995. ISBN 0-19-853499-X. Sequences of independent random variables, Oxford Science Publications.

[86] Jean Picard. Density in small time at accessible points for jump processes. *Stochastic Process. Appl.*, 67(2):251–279, 1997. ISSN 0304-4149. doi: 10.1016/S0304-4149(97)00008-2. URL https://doi.org/10.1016/S0304-4149(97)00008-2.

[87] Jim Pitman and Gerónimo Uribe Bravo. The convex minorant of a Lévy process. *Ann. Probab.*, 40(4):1636–1674, 2012. ISSN 0091-1798. URL https://doi.org/10.1214/11-AOP658.

[88] Likuan Qin and Viktor Todorov. Nonparametric implied Lévy densities. *Ann. Statist.*, 47(2):1025–1060, 2019. ISSN 0090-5364. doi: 10.1214/18-AOS1703. URL https://doi.org/10.1214/18-AOS1703.

[89] Emmanuel Rio. Upper bounds for minimal distances in the central limit theorem. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(3):802–817, 2009. ISSN 0246-0203. doi: 10.1214/08-AIHP187. URL https://doi.org/10.1214/08-AIHP187.

[90] Jan Rosiński. Series representations of Lévy processes from the perspective of point processes. In *Lévy processes*, pages 401–415. Birkhäuser Boston, Boston, MA, 2001.

[91] Ken-iti 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.

[92] W. Schoutens. *Levy Processes in Finance: Pricing Financial Derivatives*. Wiley Series in Probability and Statistics. Wiley, 2003. ISBN 9780470851562. URL https://books.google.co.jp/books?id=Be4NqIz7h-kC.

[93] W. Schoutens and J. Cariboni. *Levy Processes in Credit Risk*. The Wiley Finance Series. Wiley, 2010. ISBN 9780470685068. URL https://books.google.co.jp/books?id=DjmvZ5CSmxAC.

[94] Wim Schoutens. Exotic options under Lévy models: an overview. *J. Comput. Appl. Math.*, 189(1-2):526–538, 2006. ISSN 0377-0427. doi: 10.1016/j.cam.2005.10.004. URL https://doi.org/10.1016/j.cam.2005.10.004.

[95] Wim Schoutens and Stijn Symens. The pricing of exotic options by Monte-Carlo simulations in a Lévy market with stochastic volatility. *Int. J. Theor. Appl. Finance*, 6(8):839–864, 2003. ISSN 0219-0249. URL https://doi.org/10.1142/S0219024903002249.

[96] D. Sornette. Why stock markets crash: Critical events in complex financial systems, 2003.

[97] Vladimir V. Uchaikin and Vladimir M. Zolotarev. *Chance and stability*. Modern Probability and Statistics. VSP, Utrecht, 1999. ISBN 90-6764-301-7. doi: 10.1515/9783110935974. URL http://dx.doi.org/10.1515/9783110935974.

Stable distributions and their applications, With a foreword by V. Yu. Korolev and Zolotarev.

[98] J. Vecer. Maximum drawdown and directional trading. *Risk*, 19:99–92, 2006.

[99] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. ISBN 978-3-540-71049-3. doi: 10.1007/978-3-540-71050-9. URL https://doi.org/10.1007/978-3-540-71050-9. Old and new.