APPLYING THE WIENER–HOPF MONTE CARLO SIMULATION TECHNIQUE FOR LÉVY PROCESSES TO PATH FUNCTIONALS

ALBERT FERREIRO-CASTILLA,* *University of Bath*

KEES VAN SCHAIK,** *University of Manchester*

Abstract

In this paper we apply the recently established Wiener–Hopf Monte Carlo simulation technique for Lévy processes from Kuznetsov *et al.* (2011) to path functionals; in particular, first passage times, overshoots, undershoots, and the last maximum before the passage time. Such functionals have many applications, for instance, in finance (the pricing of exotic options in a Lévy model) and insurance (ruin time, debt at ruin, and related quantities for a Lévy insurance risk process). The technique works for any Lévy process whose running infimum and supremum evaluated at an independent exponential time can be sampled from. This includes classic examples such as stable processes, subclasses of spectrally one-sided Lévy processes, and large new families such as meromorphic Lévy processes. Finally, we present some examples. A particular aspect that is illustrated is that the Wiener–Hopf Monte Carlo simulation technique (provided that it applies) performs much better at approximating first passage times than a 'plain' Monte Carlo simulation technique based on sampling increments of the Lévy process.

Keywords: Wiener–Hopf decomposition; Monte Carlo simulation; multilevel Monte Carlo; Lévy process; exotic option pricing; first passage time; overshoot; insurance risk process

2010 Mathematics Subject Classification: Primary 65C05; 68U20; 60G51

1. Introduction

Let $X := (X_t)_{t \geq 0}$ be a Lévy process, i.e. a (real valued) stochastic process starting from 0 with càdlàg paths (right continuous and with left limits) and stationary, independent increments whose law we denote by $\mathbb{P}$. A Lévy process may be thought of as a Brownian motion with drift to which an (infinite) sequence of independent compound Poisson processes are added; infinite to the extent that its small jumps may not be summable. The Lévy–Khintchine formula entails that the characteristic exponent Ψ, defined as $\mathbb{E}[e^{izX_t}] = e^{-t\Psi(z)}$ for all $t \geq 0$ and $z \in \mathbb{R}$, can be expressed as

$$\Psi(z) = \frac{\sigma^2}{2}z^2 + iaz + \int_{\mathbb{R}\setminus\{0\}} (1 - e^{izx} + \mathbf{1}_{\{|x|<1\}}izx)\Pi(dx), \tag{1}$$

where $\sigma, a \in \mathbb{R}$ and Π is a measure on $\mathbb{R} \setminus \{0\}$ satisfying $\int_{\mathbb{R}\setminus\{0\}}(x^2 \wedge 1)\Pi(dx) < \infty$. See, for example, [4], [22], or [26] for a detailed introduction.

Of interest in several fields are quantities of the form

$$\mathbb{E}[f(\tau_u, X_{\tau_u} - u, u - X_{\tau_u-}, u - \bar{X}_{\tau_u-})], \tag{2}$$

where τ_u is the first passage time of X over a level $u > 0$, i.e.

$$\tau_u := \inf\{t > 0 \mid X_t > u\},$$

$X_{\tau_u} - u$ is referred to as the overshoot, $u - X_{\tau_u-}$ the undershoot, and finally $u - \bar{X}_{\tau_u-}$ the last maximum before first passage. Here and throughout we employ the usual notation of $\bar{X}_t := \sup_{s \leq t} X_s$ and $\underline{X}_t := \inf_{s \leq t} X_s$ for all $t \geq 0$.

Some examples of applications of (2) are as follows. In mathematical finance, models driven by Lévy processes are popular extensions of the classic Black–Scholes model used for pricing financial products (see, e.g. [8], [27], or [28]). In such models it is assumed that the financial index on which the payoff of the option is based evolves as $S_t = \exp(X_t)$ for a suitably chosen Lévy process X. The 'fair' price of a perpetual American option in such a model is typically determined by the joint law of the first hitting time and the overshoot. Furthermore, so-called barrier options are popular tools in practice. The 'vanilla' version grants the holder a payoff $g(X_T)$ at a future time T provided that X has not or has, dependent on the variation of the product, crossed some barrier B in the meantime. Hence, the payoff is a function of $(X_T, \bar{X}_T)$. The original Wiener–Hopf Monte Carlo (WHMC) simulation technique deals with this pair, see Section 2 below. A large variety of more 'exotic' versions of such options are popular as well. For instance, a certain rebate may be paid at the moment the crossing of the barrier happens. Discrete barrier options exist where the barrier crossing event is only observed at certain subperiods of $[0, T]$, and Parisian barrier options where the barrier condition kicks in only once X has spent at least a given period of time on the 'wrong' side of the barrier.

Furthermore, in actuarial science a so-called Lévy insurance risk process is a popular extension of the classic Cramér–Lundberg model (see [24]) to study the evolution of the cumulative net of premiums minus claims generated by a homogeneous portfolio of insurance products. See, e.g. [1], [17], or [30]. If the Lévy insurance risk process is $-X$ and the initial capital is u then τ_u corresponds to the time ruin occurs, i.e. the first time the cumulative net becomes negative. Furthermore, $-(X_{\tau_u} - u)$ corresponds to the debt at ruin, $X_{\tau_u} - u$ and $u - X_{\tau_u-}$ give information about the nature of the ruin event (whether the direct cause is a single large claim or rather the accumulation of many small claims), and finally $u - \bar{X}_{\tau_u-}$ gives information about how close ruin has been before the actual event.

These examples illustrate several uses of (2). To the best of our knowledge, evaluating a quantity like (2) can currently be done using one of two other approaches. We will discuss these alternatives and how they compare with the WHMC simulation method below.

The first one is a 'plain' Monte Carlo simulation method, i.e. simulate paths of a random walk whose increments have the same law as X_h for some small $h > 0$ as an approximation of the paths of X. However, there are only a few examples of Lévy processes X for which the law of X_h is known, in other cases it has to be approximated, typically by a numerical Fourier inversion. This introduces numerical inaccuracy and an extra potentially expensive computation step. Another downside is the well-known problem that the empirical law of the simulated first passage time suffers from a very significant bias which vanishes only very slowly as h vanishes, due to the fact that the random walk approach misses excursions over the level u between the grid points; see [5] and the references therein. The WHMC simulation method does not suffer from this bias, due to the fact that the method simulates the path of the pair $(X, \bar{X})$ rather than just X.

See also Section 5.2. Even though this bias, to our best knowledge is only documented for Brownian motion, since any Lévy process can be decomposed as an independent sum where one of the components is a Brownian motion, there is no reason to expect this issue to be any less significant for a more general Lévy process or the rest of the quantities involved in (2).

It should be mentioned that some prominent examples of Lévy processes used in finance do not allow us to apply our method directly. For example, Merton's jump-diffusion model (see [25]) where the driving Lévy process X is a drifted Brownian motion plus a compound Poisson process with normally distributed jumps. As the Wiener–Hopf factors are not explicitly known in this case, the WHMC cannot be applied while plain Monte Carlo can be. The same applies to the popular Carr–Geman–Madan–Yor (CGMY) (see [7]) and normal inverse Gaussian (NIG) (see [3]) models; however, due to the unavailability of the exact law of X_h for a given $h > 0$, additional simulation techniques are necessary (see, e.g. [9] and [15]) and an extra error is incorporated to the plain Monte Carlo method. It is worth mentioning with respect to the latest examples that a parametrisation of the β-family (a particular subclass of Lévy processes for which our methodology is straightforward; see Section 5.1) is such that the CGMY and NIG processes can be obtained as a limit of β-processes (see [18, Section 4]). For such cases a dedicated study would be helpful to decide whether or not the advantages of the WHMC simulation method over plain Monte Carlo, as described in the previous paragraph, outweigh the disadavantages that the WHMC method alone applies to an approximation of the actual driving Lévy process (see [12] or [29] for some results in this direction).

The second approach concerns using the quintuple law from [11], which can be written in the form

$$\mathbb{E}[e^{-q\tau_u}\mathbf{1}_{\{X_{\tau_u}-u\in dx, u-X_{\tau_u-}\in dy, u-\bar{X}_{\tau_u-}\in dz\}}]$$
$$= \frac{1}{q}\mathbb{P}(\bar{X}_{e(q)} \in u - dz)\mathbb{P}(-\underline{X}_{e(q)} \in dy - z)\Pi(dx + y) \tag{3}$$

for $q, x, y > 0$ and $z \in [0, u \vee y]$. Here and throughout, $e(q)$ denotes an exponentially distributed random variable with mean $1/q$, independent of X. Hence, if we know the laws of $\bar{X}_{e(q)}$ and $\underline{X}_{e(q)}$, which is exactly the condition under which our simulation method can be implemented, we might use this result to compute (2). However, to obtain the law of τ_u from the above expression we would need to invert the right-hand side over q, which is in general not a very straightforward operation; see, for instance, Section 5 for how the laws of $\bar{X}_{e(q)}$ and $\underline{X}_{e(q)}$ depend on q when X is a meromorphic Lévy process. Furthermore, obtaining (2) still requires computing a potentially four-dimensional integral, for which in general a numerical method will be required. Since the simulation method we propose in this paper is a straightforward, easy to implement, and efficient method for approximating (2) directly, with only the laws of $\bar{X}_{e(q)}$ and $\underline{X}_{e(q)}$ as input, in general there seems no reason not to prefer it over the quintuple law alternative.

To return to the focus of this paper, in [21] the WHMC simulating technique is introduced which allows us to sample from a law that is a good approximation of the law of $(X_T, \bar{X}_T)$, provided that samples can be produced from $\bar{X}_{e(q)}$ and $\underline{X}_{e(q)}$. This method was extended to a multilevel version and a theoretical analysis was given in [13]. In this paper we pursue the observation that the main idea behind the WHMC simulating technique can also be used to generate samples from (an approximation of)

$$(\tau_u, X_{\tau_u} - u, u - X_{\tau_u-}, u - \bar{X}_{\tau_u-}), \tag{4}$$

rather than only (an approximation of) $(X_T, \bar{X}_T)$. In fact, not only (4) but any functional of

the pair $(X, \bar{X})$ could be handled by the method (see Remark 1, below). Once this observation has been established it is simply a matter of applying the usual setup: generate a large number of such samples, apply the function f to each of them, and compute the resulting average to obtain an approximation of (2).

For simplicity, in the remainder of the paper we will refer to (4) as the *4-tuple*.

The rest of this paper is organized as follows. In Section 2 we recall the original WHMC simulating technique and discuss how the underlying idea is useful for path functionals as well. Section 3 is devoted to the main result, describing how to obtain an approximation of (4) in terms of $\bar{X}_{e(q)}$ and $\underline{X}_{e(q)}$, and the convergence in distribution of this approximation to the exact law of (4). In Section 4 we explore the convergence rate of the approximation of the first passage time, as this is the key quantity involved in (2). We will show that $\mathbb{E}[(\hat{\tau}_u^n - \tau_u)^2] = \mathcal{O}(n^{-1})$, where $\hat{\tau}_u^n$ is the approximation of τ_u as given by the extended WHMC simulating technique. We will also show that the extended WHMC simulating technique admits a multilevel version to approximate τ_u and that such enhancement makes the algorithm optimal. Finally, the implementation, some examples and some numerical results supporting the theoretical claims and exposing the practical side of the extended WHMC simulating technique are collected in Section 5.

2. The WHMC simulating technique

Let us recall the WHMC simulating technique as introduced in [21] and discuss how that setup can be used for path functionals as well. Fix some $t > 0$. The idea in [21] is to construct an approximation of the joint law of $(X_t, \bar{X}_t)$ making use of a 'stochastic grid' on the time axis together with the Wiener–Hopf factorisation as follows. Recall that $e(\lambda)$ denotes an exponentially distributed random variable with mean $1/\lambda$ independent of X. For any $n \geq 1$, enlarge the probability space on which X lives with an independent and identically distributed (i.i.d.) sequence $\{e_i(n/t)\}_{i \geq 1}$ and define a set of grid points as

$$g\left(0, \frac{n}{t}\right) := 0, \qquad g\left(k, \frac{n}{t}\right) := \sum_{i=1}^{k} e_i\left(\frac{n}{t}\right) \quad \text{for } k \geq 1. \tag{5}$$

For any n the set of random points $\{0 = g(0, n/t) < g(1, n/t) < \cdots\}$ forms a grid on the time axis, the distance between the grid points forming a sequence of i.i.d. exponentially distributed random variables; equivalently, the grid points can be seen as the arrival times of a Poisson process with rate n/t. For convenience we still denote the law of X and the grid by $\mathbb{P}$. The idea of using such a 'stochastic grid' was first coined in [6] in the context of finding the value of American put options in a Black–Scholes model.

The usefulness of this setup to sample from an approximation of $(X_t, \bar{X}_t)$ relies on the following three facts.

Firstly, the celebrated Wiener–Hopf factorisation tells us that for any Lévy process X and $q > 0$ we have

$$(X_{e(q)}, \bar{X}_{e(q)}) \stackrel{\mathrm{D}}{=} (\bar{X}_{e(q)} + \underline{X}_{e(q)}, \bar{X}_{e(q)}), \tag{6}$$

where '$\stackrel{\mathrm{D}}{=}$' denotes equality in distribution.

Secondly, making use of stationary independent increments of X, it can be shown that the above equality can be extended in the sense that we have, for any n,

$$(X_{g(n,n/t)}, \bar{X}_{g(n,n/t)}) \stackrel{\mathrm{D}}{=} (V_n^{(n/t)}, J_n^{(n/t)}),$$

where $V_n^{(n/t)}$ and $J_n^{(n/t)}$ are random variables on a new probability space whose law can be expressed in terms of the laws of $\bar{X}_{e(n/t)}$ and $\underline{X}_{e(n/t)}$ in a straightforward way. See Theorem 1, below.

Thirdly and finally, especially in recent years many families of Lévy processes have emerged for which the laws of both $\bar{X}_{e(n/t)}$ and $\underline{X}_{e(n/t)}$ are known in explicit enough form to allow sampling from. Besides classical examples such as stable processes, we can mention the class of spectrally one-sided Lévy processes for which the so-called scale functions are explicit enough (see, e.g. [16], [22], or [23]), meromorphic Lévy processes (see [20]), and Vigon's technique (see [31]) for constructing new Lévy processes from prespecified ladder height processes (and, hence, through the Wiener–Hopf factorization, with prespecified laws for $\bar{X}_{e(n/t)}$ and $\underline{X}_{e(n/t)}$).

Theorem 1. (Kuznetsov *et al.* [21, Theorem 1].) *Suppose that $\lambda > 0$. Let $S_0^{(\lambda)} = I_0^{(\lambda)} := 0$ and let $(S_i^{(\lambda)})_{i \geq 1}$ (respectively $(I_i^{(\lambda)})_{i \geq 1}$) be a sequence of i.i.d. random variables with common law equal to the law of $\bar{X}_{e(\lambda)}$ (respectively $\underline{X}_{e(\lambda)}$). Then we have, for any $n \in \mathbb{N}$,*

$$(X_{g(n,\lambda)}, \bar{X}_{g(n,\lambda)}) \stackrel{\mathrm{D}}{=} (V_n^{(\lambda)}, J_n^{(\lambda)}),$$

where $V_n^{(\lambda)}$ and $J_n^{(\lambda)}$ are iteratively defined by $V_0^{(\lambda)} = J_0^{(\lambda)} = 0$ and, for $i \geq 1$,

$$V_i^{(\lambda)} = V_{i-1}^{(\lambda)} + S_i^{(\lambda)} + I_i^{(\lambda)} \quad and \quad J_i^{(\lambda)} = \max\{J_{i-1}^{(\lambda)}, V_{i-1}^{(\lambda)} + S_i^{(\lambda)}\}.$$

Since $g(n, n/t) \to t$ almost surely (a.s.) as $n \to \infty$ by the law of large numbers and since X does not jump at fixed times, we have $(X_{g(n,n/t)}, \bar{X}_{g(n,n/t)}) \to (X_t, \bar{X}_t)$ a.s. as $n \to \infty$; hence, for large n the law of $(V_n^{(n/t)}, J_n^{(n/t)})$ provides an approximation of the joint law of $(X_t, \bar{X}_t)$.

Now let us discuss how the above setup is useful for producing samples from the 4-tuple as well. The idea is quite straightforward: the 'stochastic grid' as defined in (5) does not only satisfy $g(n, n/t) \to t$ a.s. as $n \to \infty$, but for any sequence $k(n)$ such that $k(n) \in \{0, \ldots, n\}$ and $k(n)t/n \to s \in [0, t]$ a.s. as $n \to \infty$, we have again, by the law of large numbers, $g(k(n), n/t) \to s$ a.s. Consequently, as above, for large n the law of $(V_{k(n)}^{(n/t)}, J_{k(n)}^{(n/t)})$ provides an approximation of the law of $(X_s, \bar{X}_s)$. In this sense the 'stochastic grid' becomes dense in the interval $[0, t]$. Furthermore, as is obvious from Theorem 1, due to the iterative nature of the definitions of V and J, obtaining a sample from the pair $(V_n^{(n/t)}, J_n^{(n/t)})$ requires producing a sample from the vector $((V_0^{(n/t)}, J_0^{(n/t)}), \ldots, (V_n^{(n/t)}, J_n^{(n/t)}))$. Hence, constructing the approximative law of $(X_t, \bar{X}_t)$ automatically yields an approximative law of the vector $((X_0, \bar{X}_0), (X_{1/n}, \bar{X}_{1/n}), \ldots, (X_t, \bar{X}_t))$ (see Proposition 1, below) and it is therefore at least intuitively clear that we should also be able to approximate a quantity like (2). This is made rigorous in Theorem 2, below. The difficulty to be overcome is that convergence on the 'stochastic grid' is less obvious than on a traditional deterministic grid.

Proposition 1. *Let X be a Lévy process, $\lambda > 0$, and recall V and J as defined in Theorem 1 and the definition of the stochastic grid in (5). Then we have*

$$((X_{g(0,\lambda)}, \bar{X}_{g(0,\lambda)}), \ldots, (X_{g(k,\lambda)}, \bar{X}_{g(k,\lambda)})) \stackrel{\mathrm{D}}{=} ((V_0^{(\lambda)}, J_0^{(\lambda)}), \ldots, (V_k^{(\lambda)}, J_k^{(\lambda)})). \tag{7}$$

Proof. This is a straightforward adaptation of the proof of Theorem 1 in [21]; we include it here for completeness. The proof is by induction over k. It is trivially true for $k = 1$ on account

of (6). Let $k \geq 2$. We have, where Y is an independent copy of X and we use the notation $\bar{X}_{s_1,s_2} := \sup_{s_1 \leq u \leq s_2} X_u$,

$$
\begin{aligned}
&((X_{g(0,\lambda)}, \bar{X}_{g(0,\lambda)}), \ldots, (X_{g(k,\lambda)}, \bar{X}_{g(k,\lambda)})) \\
&= ((X_{g(0,\lambda)}, \bar{X}_{g(0,\lambda)}), \ldots, (X_{g(k-1,\lambda)}, \bar{X}_{g(k-1,\lambda)}), \\
&\quad (X_{g(k,\lambda)}, \max\{\bar{X}_{0,g(k-1,\lambda)}, \bar{X}_{g(k-1,\lambda),g(k,\lambda)}\})) \\
&\stackrel{\mathrm{D}}{=} ((X_{g(0,\lambda)}, \bar{X}_{g(0,\lambda)}), \ldots, (X_{g(k-1,\lambda)}, \bar{X}_{g(k-1,\lambda)}), \\
&\quad (X_{g(k-1,\lambda)} + Y_{\mathrm{e}(\lambda)}, \max\{\bar{X}_{0,g(k-1,\lambda)}, X_{g(k-1,\lambda)} + \bar{Y}_{\mathrm{e}(\lambda)}\})) \\
&\stackrel{\mathrm{D}}{=} ((V_0^{(\lambda)}, J_0^{(\lambda)}), \ldots, (V_{k-1}^{(\lambda)}, J_{k-1}^{(\lambda)}), (V_{k-1}^{(\lambda)} + S_k^{(\lambda)} + I_k^{(\lambda)}, \max\{J_{k-1}^{(\lambda)}, V_{k-1}^{(\lambda)} + S_k^{(\lambda)}\})) \\
&= ((V_0^{(\lambda)}, J_0^{(\lambda)}), \ldots, (V_{k-1}^{(\lambda)}, J_{k-1}^{(\lambda)}), (V_k^{(\lambda)}, J_k^{(\lambda)})),
\end{aligned}
$$

where the second equality uses that X has stationary, independent increments and the third equality uses (6) together with the induction hypothesis and the definition of the sequences $(S_i^{(\lambda)})_{i \geq 1}$ and $(I_i^{(\lambda)})_{i \geq 1}$.

There is yet another heuristic justification to support the skeleton $\{X_{g(k,n/t)}\}_{k \geq 0}$ as a good random walk approximation of the Lévy process to compute pathwise quantities. It can be inferred from [10] that, for all $k > 0$, the random variables

$$
M_k := \sup_{g(k,n/t) \leq t < g(k+1,n/t)} X_t \quad \text{and} \quad m_k := \inf_{g(k,n/t) \leq t < g(k+1,n/t)} X_t
$$

can be written as

$$
M_k = S_0^{(n/t)} + Y_k^{(+)} \quad \text{and} \quad m_k = I_0^{(n/t)} + Y_k^{(-)},
$$

where $\{Y_k^{(+)}\}_{k \geq 0}$ and $\{Y_k^{(-)}\}_{k \geq 0}$ are random walks with the same distribution as $\{X_{g(k,n/t)}\}_{k \geq 0}$ and independent of $S_0^{(n/t)}$ and $I_0^{(n/t)}$, respectively. Since it is clear that

$$
m_k \leq X_t \leq M_k \quad \text{for } g\left(k, \frac{n}{t}\right) \leq t < g\left(k+1, \frac{n}{t}\right),
$$

the derivations in [10] assert that it is possible to 'stochastically' bound the path of X from above and below by two random walks which are equal in distribution to the skeleton constructed in Proposition 1, but with different random starting points. Heuristically, the random walk produced in Proposition 1 should be particularly useful when used to approximate pathwise quantities of X.

3. Approximate distribution of the 4-tuple

Let us now show how the setup introduced in Section 2 can be used to generate an approximate distribution of the 4-tuple as well. The idea is to approximate τ_u by finding points on our 'stochastic grid' (5) enclosing it, i.e. $k(n) \in \mathbb{N}$ such that $g(k(n) - 1, n/t) \leq \tau_u \leq g(k(n), n/t)$ for all $n \in \mathbb{N}$, and evaluate the functionals involving overshoots and undershoots using these grid points.

Theorem 2. *Let X be any Lévy process. Fix some $t > 0$ and $u > 0$. Recall V and J as defined in Theorem 1. Set, for all $n \in \mathbb{N}$,*

$$
\kappa_u^{(n)} := \inf\{k \in \{0, \ldots, n\} \mid J_k^{(n/t)} > u\}
$$

(where as usual we understand $\inf \varnothing = \infty$*). Then we have, as* $n \to \infty$,

$$\left(\frac{t}{n} (\kappa_u^{(n)} \wedge n), V_{\kappa_u^{(n)} \wedge n}^{(n/t)} - u, u - V_{(\kappa_u^{(n)} - 1) \wedge n}^{(n/t)}, u - J_{(\kappa_u^{(n)} - 1) \wedge n}^{(n/t)} \right)$$
$$\stackrel{\mathrm{D}}{=} (\tau_u \wedge t, X_{\tau_u \wedge t} - u, u - X_{(\tau_u \wedge t)-}, u - \bar{X}_{(\tau_u \wedge t)-}). \tag{8}$$

Before we prove this result let us prove two technical lemmas, the first of which is a well-known result reproduced here for the sake of completeness and the second one is a computation which will appear in the proof of Theorem 2.

Lemma 1. *Suppose that* Z *is a compound Poisson process with only positive jumps following an exponential distribution with mean* $1/\theta$. *Then, for any* $u > 0$ *and* $\varepsilon \in [0, u)$,

$$\mathbb{P}(Z_{\tau_u} - u > \varepsilon) = \mathbb{P}(u - Z_{\tau_u-} > \varepsilon) = e^{-\theta \varepsilon}.$$

Proof. Condition on the event that first passage over u happens at the nth jump, this jump is now from an independent random level below u over u and follows an exponential distribution. Hence, using the lack of memory property the result for the overshoot follows. For the undershoot $u - Z_{\tau_u-}$, we can condition on the same event and use that the undershoot is the difference between u and the sum of $n - 1$ independent exponentially distributed random variables to conclude the proof.

Lemma 2. *Let* X *be a Lévy process,* T *a random time,* $(A^{(n)})_{n \geq 1}$ *a sequence of events, and* $(T^{(n)})_{n \geq 1}$ *a sequence of random times such that*

$$\mathbf{1}_{A^{(n)}} (T^{(n)} - T) \stackrel{\mathbb{P}}{\to} 0 \quad \text{as } n \to \infty.$$

(i) *If, for all* n, *we have* $T^{(n)} \geq T$ *on* $A^{(n)}$ *a.s., then*

$$\mathbf{1}_{A^{(n)}} (X_{T^{(n)}} - X_T) \stackrel{\mathbb{P}}{\to} 0 \quad \text{as } n \to \infty. \tag{9}$$

(ii) *If, for all* n, *we have* $T^{(n)} < T$ *on* $A^{(n)}$ *a.s., then*

$$\mathbf{1}_{A^{(n)}} (X_{T^{(n)}} - X_{T-}) \stackrel{\mathbb{P}}{\to} 0 \quad \text{and} \quad \mathbf{1}_{A^{(n)}} (\bar{X}_{T^{(n)}} - \bar{X}_{T-}) \stackrel{\mathbb{P}}{\to} 0 \quad \text{as } n \to \infty. \tag{10}$$

Proof. For (9), fix some $\varepsilon > 0$ and let $\varepsilon' > 0$, then we have

$$\mathbb{P}(\mathbf{1}_{A^{(n)}} | X_{T^{(n)}} - X_T | > \varepsilon) = \mathbb{P}(\mathbf{1}_{A^{(n)}} (\mathbf{1}_{\{T^{(n)} \in [T, T+\varepsilon']\}} + \mathbf{1}_{\{T^{(n)} > T+\varepsilon'\}}) | X_{T^{(n)}} - X_T | > \varepsilon)$$
$$\leq \mathbb{P}\left(\sup_{s \in [0, \varepsilon']} |X_{T+s} - X_T| > \varepsilon \right) + \mathbb{P}(\mathbf{1}_{A^{(n)}} (T^{(n)} > T + \varepsilon')),$$

where the second term vanishes as $n \to \infty$ by assumption, while the first can be made arbitrarily small by choosing ε' small enough on account of right-continuity and path regularity. Clearly, (10) follows by similar means and left-continuity of $(X_{s-})_{s \geq 0}$.

Proof of Theorem 2. The proof goes by defining an auxiliary random vector which would be equal in distribution to the left-hand side of (8) and which is such that converges in probability to the right-hand side of (8), thus, obtaining the claim of the statement. In order to define such

an auxiliary random vector let us first introduce the following quantities:

$$k_X^{(n)} := \inf\{k \in \{0, \ldots, n\} \mid \bar{X}_{g(k, n/t)} > u\},$$

$$k_g^{(n)} := \inf\left\{g\left(k, \frac{n}{t}\right) \;\middle|\; g\left(k, \frac{n}{t}\right) > \tau_u\right\},$$

$$\sigma_+^{(n)} := \mathbf{1}_{\{k_X^{(n)} < \infty\}} g\left(k_X^{(n)}, \frac{n}{t}\right) + \mathbf{1}_{\{k_X^{(n)} = \infty\}} g\left(n, \frac{n}{t}\right),$$

$$\sigma_-^{(n)} := \mathbf{1}_{\{k_X^{(n)} < \infty\}} g\left(k_X^{(n)} - 1, \frac{n}{t}\right) + \mathbf{1}_{\{k_X^{(n)} = \infty\}} g\left(n, \frac{n}{t}\right).$$

We first observe some relationships between the above variables. Note that we may write

$$\mathbf{1}_{\{k_X^{(n)} < \infty\}} \sigma_+^{(n)} = \mathbf{1}_{\{k_X^{(n)} < \infty\}} k_g^{(n)}. \tag{11}$$

Indeed, on the event $\{\bar{X}_{g(k, n/t)} \leq u\}$ we have, on the one hand, $\tau_u \geq g(k, n/t)$; hence, $k_g^{(n)} > g(k, n/t)$. On the other hand, we have $k_X^{(n)} > k$; hence, $\sigma_+^{(n)} > g(k, n/t)$. On the event $\{\bar{X}_{g(k, n/t)} > u\}$, since X does not jump at fixed times it also does not at $g(k, n/t)$; thus, $\tau_u < g(k, n/t)$, so $k_g^{(n)} \leq g(k, n/t)$; moreover, $k_X^{(n)} \leq k$; hence, $\sigma_+^{(n)} \leq g(k, n/t)$. Summarizing, on the event $\{\bar{X}_{g(k, n/t)} \leq u\}$ we have that $k_g^{(n)}, \sigma_+^{(n)} > g(k, n/t)$ and on the event $\{\bar{X}_{g(k, n/t)} > u\}$ we have $k_g^{(n)}, \sigma_+^{(n)} \leq g(k, n/t)$, which proves that (11) holds since $k_g^{(n)}$ and $\sigma_+^{(n)}$ can only take values on the stochastic grid. Furthermore, we have

$$\mathbf{1}_{\{k_X^{(n)} < \infty\}} \sigma_-^{(n)} < \mathbf{1}_{\{k_X^{(n)} < \infty\}} \tau_u, \tag{12}$$

since by construction $\sigma_-^{(n)} \leq \tau_u$ on the event $\{k_X^{(n)} < \infty\}$, and $\sigma_-^{(n)} = \tau_u$ cannot happen a.s. due to the fact that X and $\{g(k, n/t)\}_{k=0}^{n}$ are independent.

Now, with these definitions we have

$$\left(\frac{t}{n}(k_X^{(n)} \wedge n), X_{\sigma_+^{(n)}} - u, u - X_{\sigma_-^{(n)}}, u - \bar{X}_{\sigma_-^{(n)}}\right)$$

$$\stackrel{\mathrm{D}}{=} \left(\frac{t}{n}(\kappa_u^{(n)} \wedge n), V_{\kappa_u^{(n)} \wedge n}^{(n/t)} - u, u - V_{(\kappa_u^{(n)} - 1) \wedge n}^{(n/t)}, u - J_{(\kappa_u^{(n)} - 1) \wedge n}^{(n/t)}\right),$$

which follows from applying the same functional to the left- and right-hand side of (7) (with $\lambda = n/t$). Hence, the main result follows if we show that

$$\left(\frac{t}{n}(k_X^{(n)} \wedge n), X_{\sigma_+^{(n)}} - u, u - X_{\sigma_-^{(n)}}, u - \bar{X}_{\sigma_-^{(n)}}\right)$$

$$\stackrel{\mathbb{P}}{\to} (\tau_u \wedge t, X_{\tau_u \wedge t} - u, u - X_{(\tau_u \wedge t)-}, u - \bar{X}_{(\tau_u \wedge t)-}) \quad \text{as } n \to \infty, \tag{13}$$

for which in turn it is enough to show that each component on the left-hand side converges in probability to its counterpart on the right-hand side. This takes up the remainder of the proof.

To prove the convergence in probability of the first component in (13), note that, since $g(n, n/t) \to t$ a.s. as $n \to \infty$, we have

$$\mathbf{1}_{\{k_X^{(n)} = \infty\}}(t - \tau_u \wedge t) \leq \mathbf{1}_{\{k_X^{(n)} = \infty, \tau_u < t\}} t = \mathbf{1}_{\{g(n, n/t) < \tau_u, \tau_u < t\}} t \to 0 \text{ a.s.} \quad \text{as } n \to \infty.$$

Similarly,

$$\mathbf{1}_{\{k_X^{(n)}<\infty,\tau_u\geq t\}}\left|\frac{t}{n}k_X^{(n)}-t\right|\leq\mathbf{1}_{\{g(n,n/t)>\tau_u,\tau_u>t\}}t\to 0\quad\text{a.s.}\quad\text{as }n\to\infty;$$

hence, it is enough to show that

$$\mathbf{1}_{\{k_X^{(n)}<\infty,\tau_u<t\}}\left(\frac{t}{n}k_X^{(n)}-\tau_u\right)\xrightarrow{\mathbb{P}}0\quad\text{as }n\to\infty.\tag{14}$$

To prove that (14) holds, let $\varepsilon>0$ and take $0=t_0<t_1<\cdots<t_N=t$ such that $t_i-t_{i-1}<\frac{1}{2}\varepsilon$ for all i. Then we have

$$\mathbb{P}\left(\mathbf{1}_{\{k_X^{(n)}<\infty,\tau_u<t\}}\left|\frac{t}{n}k_X^{(n)}-\tau_u\right|>\varepsilon\right)$$

$$=\sum_{i=1}^{N}\mathbb{P}\left(\mathbf{1}_{\{k_X^{(n)}<\infty,\tau_u\in[t_{i-1},t_i)\}}\left|\frac{t}{n}k_X^{(n)}-\tau_u\right|>\varepsilon\right)$$

$$\leq\sum_{i=1}^{N}\mathbb{P}\left(t_{i-1}-\frac{t}{n}\chi^{(n)}(t_{i-1})>\frac{\varepsilon}{2}\right)+\mathbb{P}\left(\frac{t}{n}\chi^{(n)}(t_i)-t_i>\frac{\varepsilon}{2}\right),\tag{15}$$

where, for any $x\geq 0$,

$$\chi^{(n)}(x):=\inf\left\{k\geq 0\,\Big|\,g\left(k,\frac{n}{t}\right)>x\right\}.$$

Fix $1\leq i\leq N$ and let $\bar{k}(n)=n(t_i+\frac{1}{2}\varepsilon)/t$ and $\underline{k}(n)=n(t_{i-1}-\frac{1}{2}\varepsilon)/t$ for $n\geq 1$, then we see

$$\mathbb{P}\left(\frac{t}{n}\chi^{(n)}(t_i)-t_i>\frac{\varepsilon}{2}\right)\leq\mathbb{P}(\chi^{(n)}(t_i)>\bar{k}(n))=\mathbb{P}\left(g\left(\bar{k}(n),\frac{n}{t}\right)\leq t_i\right)\to 0,\tag{16}$$

$$\mathbb{P}\left(t_{i-1}-\frac{t}{n}\chi^{(n)}(t_{i-1})>\frac{\varepsilon}{2}\right)\leq\mathbb{P}(\underline{k}(n)>\chi^{(n)}(t_{i-1}))=\mathbb{P}\left(t_{i-1}\leq g\left(\underline{k}(n),\frac{n}{t}\right)\right)\to 0\tag{17}$$

as $n\to\infty$, since the law of large numbers ensures that

$$g\left(\bar{k}(n),\frac{n}{t}\right)\to t_i+\frac{\varepsilon}{2}>t_i\quad\text{a.s.}\quad\text{and}\quad g\left(\underline{k}(n),\frac{n}{t}\right)\to t_{i-1}-\frac{\varepsilon}{2}<t_{i-1}\quad\text{a.s..}$$

Finally, we use (16) and (17) in (15) to conclude that (14) holds.

For the second component of (13), we again use the convergence of $g(n,n/t)\to t$ and the fact that X does not jump at fixed times to note that

$$\mathbf{1}_{\{k_X^{(n)}=\infty\}}X_{\sigma_+^{(n)}}=\mathbf{1}_{\{\bar{X}_{g(n,n/t)}\leq u\}}X_{g(n,n/t)}\to\mathbf{1}_{\{\bar{X}_t\leq u\}}X_t=\mathbf{1}_{\{\tau_u\geq t\}}X_t\quad\text{a.s.}$$

Using this with $\mathbf{1}_{\{k_X^{(n)}<\infty\}}\to\mathbf{1}_{\{\tau_u<t\}}$ a.s., it remains to show that

$$\mathbf{1}_{\{k_X^{(n)}<\infty,\tau_u<t\}}(X_{\sigma_+^{(n)}}-X_{\tau_u})\xrightarrow{\mathbb{P}}0\quad\text{as }n\to\infty.$$

By virtue of (9) and recalling (11) to check that $\sigma_+^{(n)} = k_g^{(n)} > \tau_u$ on $\{k_X^{(n)} < \infty\}$, the above limit will hold if we show that

$$\mathbf{1}_{\{k_X^{(n)} < \infty, \tau_u < t\}}(\sigma_+^{(n)} - \tau_u) \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \to \infty.$$

For any $\varepsilon > 0$ recall the partition of $[0, t]$ in (15). Then we have

$$
\begin{aligned}
\mathbb{P}(\mathbf{1}_{\{k_X^{(n)} < \infty, \tau_u < t\}}(\sigma_+^{(n)} - \tau_u) > \varepsilon) &= \sum_{i=1}^{N} \mathbb{P}(\mathbf{1}_{\{k_X^{(n)} < \infty, \tau_u \in [t_{i-1}, t_i)\}}(\sigma_+^{(n)} - \tau_u) > \varepsilon) \\
&\leq \sum_{i=1}^{N} \mathbb{P}(\upsilon^{(n)}(t_i) - t_{i-1} > \varepsilon) \\
&= \sum_{i=1}^{N} \mathbb{P}\left(\upsilon^{(n)}(t_i) - t_i > \frac{\varepsilon}{2}\right),
\end{aligned}
\tag{18}
$$

where, for any $x \geq 0$,

$$\upsilon^{(n)}(x) := \inf\left\{ g\left(k, \frac{n}{t}\right) \geq 0 \,\middle|\, g\left(k, \frac{n}{t}\right) > x \right\}.$$

Since $\upsilon^{(n)}$ is nothing but the first passage time over x by a compound Poisson process with exponential jumps with mean t/n, we see from Lemma 1 that the right-hand side of (18) indeed vanishes as $n \to \infty$.

For the third and forth components of (13), a similar argument to the one in the beginning of the previous paragraph will tell us that it is enough to show

$$\mathbf{1}_{\{k_X^{(n)} < \infty, \tau_u < t\}}(X_{\sigma_-^{(n)}} - X_{\tau_u-}) \xrightarrow{\mathbb{P}} 0 \quad \text{and} \quad \mathbf{1}_{\{k_X^{(n)} < \infty, \tau_u < t\}}(\bar{X}_{\sigma_-^{(n)}} - \bar{X}_{\tau_u-}) \xrightarrow{\mathbb{P}} 0,$$

as $n \to \infty$. Recalling (12) to check the assumptions of (10), both limits above will hold if we prove

$$\mathbf{1}_{\{k_X^{(n)} < \infty, \tau_u < t\}}(\sigma_-^{(n)} - \tau_u) \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \to \infty.$$

For this we may apply the obvious analogue to the argument in (18), with the understanding that rather than $\upsilon^{(n)}(x) - x$ we now need to make use of

$$x - \sup\left\{ g\left(k, \frac{n}{t}\right) \geq 0 \,\middle|\, g\left(k, \frac{n}{t}\right) < x \right\},$$

as it is apparent from the definition of $\sigma_-^{(n)}$ and (12). However, this expression is nothing but the undershoot of a compound Poisson process with exponential jumps at first passage over x; hence, Lemma 1 again applies to yield the result.

Remark 1. As will be clear from the setup and the proof of Theorem 2, there is nothing special about the entries in the 4-tuple other than that they can be expressed as path functionals of the pair $(X, \bar{X})$. Any other functional of this type can be handled by the method as well, and by symmetry so can path functionals of the pair $(X, \underline{X})$. It is worth noting though that path functionals depending on both $\bar{X}$ and $\underline{X}$ cannot (in general) be handled by the current method as there is no analogue of (6) for the pair $(\bar{X}_{e(q)}, \underline{X}_{e(q)})$.

Remark 2. It could be argued that a weakness of the WHMC simulation method is the fact that we replace the fixed time t by a random variable $g(n, n/t)$ and the *a priori* error this causes. In [13] a comprehensive error analysis is carried out for the bivariate distribution $(X_{g(n,n/t)}, \bar{X}_{g(n,n/t)})$ from where convergence rates are derived in terms of the moments of $g(n, n/t)$. For a Lévy process with finite second moment it is deduced that $\mathbb{E}[(X_{g(n,n/t)} - X_t)^2] = \mathcal{O}(n^{-1/2})$. Reassuringly, the WHMC simulation technique clearly outperforms 'plain' Monte Carlo when the running supremum gets involved. (Also see Section 5.2).

4. Convergence rate of the first passage time

All quantities involved in (2) ultimately depend on the first passage time. We now derive a convergence rate for the approximation of the first passage time in order to gain some insight into the efficiency of the Monte Carlo scheme based on the construction in Theorem 2. Recall from the proof of Theorem 2 that $\kappa_u^{(n)} \overset{\mathrm{D}}{=} k_X^{(n)}$, where $k_X^{(n)}$ lives on the same probability space as X.

Theorem 3. *Using the same notation as in Theorem 2, we have*

$$\mathbb{E}\left[\left(\frac{t}{n}(k_X^{(n)} \wedge n) - \tau \wedge t\right)^2\right] \leq \frac{2t^2}{n}.$$

Proof. Note that an alternative definition of $k_X^{(n)}$ in the proof of Theorem 2 is

$$k_X^{(n)} := \inf\left\{k \in \{0, \ldots, n\} \ \middle| \ g\left(k, \frac{n}{t}\right) > \tau_u\right\};$$

hence, conditioned on τ_u, $k_X^{(n)}$ follows a truncated Poisson distribution. Let us write $\alpha = \tau/t$ to ease the notation and denote by $N := \{N_t\}_{t \geq 0}$ a Poisson process with rate n/t. We then write

$$\frac{1}{t^2}\mathbb{E}\left[\left(\frac{t}{n}(k_X^{(n)} \wedge n) - \tau_u \wedge t\right)^2 \ \middle| \ \tau_u\right]$$

$$= \underbrace{\mathbb{E}[(1 - \alpha \wedge 1)^2 \mid \tau_u]\mathbb{P}(k_X^{(n)} = \infty \mid \tau_u)}_{(I)} + \underbrace{\mathbb{E}\left[\left(\frac{k_X^{(n)}}{n} - \alpha \wedge 1\right)^2 \ \middle| \ \tau_u\right]\mathbb{P}(k_X^{(n)} \leq n \mid \tau_u)}_{(II)}.$$

$$(19)$$

We note that, when $\tau_u \geq t$ ($\alpha \geq 1$), the term (I) vanishes; hence,

$$(I) \leq \mathbf{1}_{\{\alpha < 1\}}(1 - \alpha)^2\mathbb{P}(k_X^{(n)} = \infty \mid \tau_u)$$

$$= \mathbf{1}_{\{\alpha < 1\}}(1 - \alpha)^2\mathbb{P}(N_{\tau_u} > n \mid \tau_u)$$

$$\leq \mathbf{1}_{\{\alpha < 1\}}(1 - \alpha)^2 \frac{\alpha n}{\alpha n + (1 - \alpha)^2 n^2}$$

$$\leq \frac{1}{n},$$

$$(20)$$

where the last equality follows from the observation that N_{τ_u} has a Poisson distribution with parameter $\tau_u n/t = \alpha n$ and the one-sided Chebyshev inequality, i.e.

$$\mathbb{P}(Z - \mu \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2},$$

for a real random variable Z with mean μ and variance σ^2. We now bound the probability factor in term (II) by 1 and note that

$$
\begin{aligned}
(II) &\leq \mathbf{1}_{\{\alpha \geq 1\}} \sum_{i=1}^{n} \left(\frac{i}{n} - 1\right)^2 e^{-\alpha n} \frac{(\alpha n)^i}{i!} + \mathbf{1}_{\{\alpha < 1\}} \sum_{i=1}^{n} \left(\frac{i}{n} - \alpha\right)^2 e^{-\alpha n} \frac{(\alpha n)^i}{i!} \\
&\leq \mathbf{1}_{\{\alpha \geq 1\}} e^{(1-\alpha)n} \alpha^n \sum_{i=1}^{n} \left(\frac{i}{n} - 1\right)^2 e^{-n} \frac{n^i}{i!} + \mathbf{1}_{\{\alpha < 1\}} \sum_{i=1}^{\infty} \left(\frac{i}{n} - \alpha\right)^2 e^{-\alpha n} \frac{(\alpha n)^i}{i!} \\
&\leq \mathbf{1}_{\{\alpha \geq 1\}} e^{(1-\alpha)n} \alpha^n \frac{1}{n} + \mathbf{1}_{\{\alpha < 1\}} \frac{\alpha}{n} \\
&\leq \frac{1}{n},
\end{aligned}
\tag{21}
$$

where the last equality follows from $(1 + x) \leq e^x$ for $x \geq 0$.

The claim of the statement follows by noting that (20) and (21) are upper bounds independent from τ_u and applying the tower property in (19).

The convergence rates for the rest of the tuple seem less straightforward to derive, which is due to the fact that the error for the other entries in the tuple (4) depend more heavily on the path of X. To briefly illustrate that this is a somewhat subtle problem, it seems that if X is any Lévy process with jumps it holds that

$$
\mathbb{E}\left[\max_{k \leq n}(X_{g(k,n/t)} - X_{kt/n})^2\right]
\tag{22}
$$

does not vanish as $n \to \infty$. (This claim is currently supported by numerical evidence only.) Indeed, if a joint realisation of the stochastic grid and of the process X is such that at least one jump of X occurs in the set

$$
J_n := \bigcup_{k \leq n} \left(\min\left\{ g\left(k, \frac{n}{t}\right), \frac{kt}{n} \right\}, \max\left\{ g\left(k, \frac{n}{t}\right), \frac{kt}{n} \right\} \right]
$$

then the random variable inside the expectation in (22) is bounded below by the size of that jump squared. Hence, (22) could only vanish if in the limit no jumps occur in J_n. However, some numerical experiments conducted by the authors suggest that the Lebesgue measure of J_n tends to t as $n \to \infty$ which suggests that this requirement is not satisfied.

Nevertheless, even though the vanishing of (22) would facilitate deriving error bounds for the other entries of the tuple in the spirit of Theorem 3, this is not a necessary condition. Indeed, it could very well be the case that (22) does not vanish while the expected squared error for the remaining entries in the tuple does vanish. In Section 5 some numerical results can be found, suggesting this is indeed the case and yielding empirical numerical convergence rates.

4.1. Multilevel Monte Carlo schemes for the first passage time

In [13] the original WHMC simulation technique as introduced in Section 2, i.e. to approximate quantities of the form $\mathbb{E}[f(X_t, \bar{X}_t)]$, was extended to a multilevel Monte Carlo algorithm under the assumption that f is uniformly Lipschitz. In principle, the same multilevel scheme presented in that paper can be used to approximate (2) as well, provided that X has a finite second moment (and provided that the random tuple approximation converges in mean square error). It is out of the scope of this paper to give the full details of such a scheme, we refer

the reader to [14] for the general theory of multilevel Monte Carlo methods and to [13] for the specifics in the adaptation to the Wiener–Hopf scheme for Lévy processes. However, since we have derived the convergence rate for the first passage time approximation in Theorem 3 we are able to derive the convergence rate for its multilevel version as well. Its numerical performance is checked in Section 5.

For the following discussion it is enough to consider $f : [0, t] \to \mathbb{R}$ Lipschitz with constant 1, fix $n_0, L \in \mathbb{N}$, and set $n_\ell = 2^\ell n_0$ for $\ell = 0, \dots, L$. Let us write

$$f^{n_\ell} := f\left(\frac{t}{n_\ell}(\kappa_u^{(n_\ell)} \wedge n_\ell)\right).$$

The multilevel Monte Carlo algorithm proposes to estimate $\mathbb{E}[f(\tau_u \wedge t)]$ by $\mathbb{E}[f^{n_L}]$ according to the right-hand side of

$$\mathbb{E}[f^{n_L}] = \mathbb{E}[f^{n_0}] + \sum_{\ell=1}^{L} \mathbb{E}[f^{n_\ell} - f^{n_{\ell-1}}]. \tag{23}$$

In other words, the multilevel version of a Monte Carlo estimator proposes to perform regular Monte Carlo estimators in each of the expectations on the right-hand side of (23) as an approximation to $\mathbb{E}[f(\tau_u \wedge t)]$, i.e.

$$\mathbb{E}[f(\tau_u \wedge t)] \approx \hat{f}_{\mathrm{ML}}^{\mathcal{M}(n_0, L)} := \frac{1}{M_0} \sum_{i=1}^{M_0} f^{n_0,(i)} + \sum_{\ell=1}^{L} \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} (f^{n_\ell,(i)} - f^{n_{\ell-1},(i)}),$$

where $\mathcal{M}(n_0, L) := \{M_\ell\}_{\ell=0}^{L}$ are the Monte Carlo trials in each level and $f^{n_\ell,(i)}$ denotes the ith sample from f^{n_ℓ}. The following theorem describes the gain of this approach.

Theorem 4. ([13, Theorem 4.2].) *Let $t > 0$ and $n_\ell = n_0 2^\ell$, for some $\ell, n_0 \in \mathbb{N}$; suppose that there are positive constants $\alpha, \beta > 0$ with $\alpha \geq \frac{1}{2}(\beta \wedge 1)$ such that*

(i) $|\mathbb{E}[f^{n_\ell} - f(\tau_u \wedge t)]| \lesssim n_\ell^{-\alpha}$,

(ii) $\mathrm{var}(f^{n_\ell} - f^{n_{\ell-1}}) \lesssim n_\ell^{-\beta}$,

(iii) $\mathbb{E}[\mathcal{C}_{n_\ell}] \lesssim n_\ell$,

where $\mathcal{C}_{n_\ell}$ represents the cost of computing a single sample of f^{n_ℓ}. Then, for every $v \in \mathbb{R}_{>0}$, there exists a value L and a sequence $\mathcal{M}(n_0, L) = \{M_\ell\}_{\ell=0}^{L}$ such that

$$\mathbb{E}[\mathcal{C}(\hat{f}_{\mathrm{ML}}^{\mathcal{M}(n_0, L)})] \lesssim v$$

and

$$\mathbb{E}[(\hat{f}_{\mathrm{ML}}^{\mathcal{M}(n_0, L)} - f(\tau_u \wedge t))^2] \lesssim \begin{cases} v^{-1/2} & \text{if } \beta > 1, \\ v^{-1/2} \log v & \text{if } \beta = 1, \\ v^{-1/(2+(1-\beta)/\alpha)} & \text{if } \beta < 1. \end{cases}$$

In Section 5 we perform some numerical simulations for a Lévy process X belonging to the class of so-called meromorphic Lévy processes (see Section 5.1). It is inferred from [13] that

any functional applied to the random walk generated in (7) satisfies condition (iii) of Theorem 4. The Lipschitz assumption on f and the triangle inequality together with Theorem 3 ensure that $\beta = 1$ and, therefore, up to logarithms, the multilevel Monte Carlo estimator of $f(\tau_u \wedge t)$ provided by Theorem 4 is optimal. Note that in the regime $\beta = 1$ the bias does not play a role in the convergence of the algorithm. That is, the scheme proposed here used in a multilevel Monte Carlo estimator of the quantity $\mathbb{E}[f(\tau_u \wedge t)]$ is optimal.

5. Numerical implementation

In this section we discuss some examples of implementations of the method. Our aim is to show that the method is easily implemented, has a clear advantage over a 'plain' Monte Carlo approach (cf. Section 5.2), and give some intuition about the convergence rates of the method. There are already other studies available which highlight the numerical performance of the WHMC simulation technique in terms of computation time (see [13] or [29]), at least for the original setup to approximate quantities depending on $(X_t, \bar{X}_t)$ for some $t > 0$ (see Section 2). For this reason we focus here on a more qualitative numerical analysis in order to gain further insight into how the WHMC simulation technique performs and we omit speed comparisons which ultimately depend on the particular problem in hand.

5.1. The β-family

With the exception of Section 5.2 we choose X to be a member of the β-family of Lévy processes to perform the numerical experiments. This is a subclass, like the θ-processes (see [19]) and general hypergeometric Lévy processes (see [21]), of the family of so-called meromorphic Lévy processes as recently introduced in [20]. The class of meromorphic Lévy processes is very rich: paths of bounded and unbounded variation and both finite and infinite activity jumps can be generated. In terms of the triplet (see (1)) a meromorphic Lévy process can be characterized by any σ, any a, and any Lévy measure that can be written as a (possibly infinite) mixture of exponential distributions on $\mathbb{R}_{<0}$ and $\mathbb{R}_{>0}$. The name comes from the fact that their characteristic exponent Ψ can be extended to a meromorphic function on $\mathbb{C}$.

In particular, a member of the β-family has a Lévy measure with a density π given by

$$\pi(x) = \mathbf{1}_{\{x>0\}} c_1 \frac{e^{-\alpha_1 \beta_1 x}}{(1 - e^{-\beta_1 x})^{\lambda_1}} + \mathbf{1}_{\{x<0\}} c_2 \frac{e^{\alpha_2 \beta_2 x}}{(1 - e^{\beta_2 x})^{\lambda_2}}, \tag{24}$$

where $\alpha_i, \beta_i > 0$, $c_i \geq 0$, and $\lambda_i \in (0, 3)$. For our simulation method we need to know the laws of $\bar{X}_{e(q)}$ and $\underline{X}_{e(q)}$; they are easily obtained as follows (see [18] for more details). With the usual notation, $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x + y)$ and $\psi(x) = (d/dx)\log\Gamma(x)$, the characteristic exponent (1) of the Lévy–Khintchine representation can be written as

$$\Psi(z) = \frac{\sigma^2}{2} z^2 + i\rho z - \frac{c_1}{\beta_1} B\left(\alpha_1 - \frac{iz}{\beta_1}, 1 - \lambda_1\right) - \frac{c_2}{\beta_2} B\left(\alpha_2 + \frac{iz}{\beta_2}, 1 - \lambda_2\right) + \gamma,$$

where

$$\gamma = \frac{c_1}{\beta_1} B(\alpha_1, 1 - \lambda_1) + \frac{c_2}{\beta_2} B(\alpha_2, 1 - \lambda_2),$$

$$\rho = \frac{c_1}{\beta_1^2} B(\alpha_1, 1 - \lambda_1)(\psi(1 + \alpha_1 - \lambda_1) - \psi(\alpha_1))$$

$$- \frac{c_2}{\beta_2^2} B(\alpha_2, 1 - \lambda_2)(\psi(1 + \alpha_2 - \lambda_2) - \psi(\alpha_2)) - a.$$

For any $q > 0$, the equation $\zeta \mapsto q + \Psi(i\zeta)$ has infinitely many zeros, all real and simple, located as follows:

$$\zeta_0^- \in (-\beta_1\alpha_1, 0), \qquad \zeta_0^+ \in (0, \alpha_2\beta_2),$$
$$\zeta_k^- \in (\beta_1(k - \alpha_1), \beta_1(k + 1 - \alpha_1)) \quad \text{for } k \leq -1,$$
$$\zeta_k^+ \in (\beta_2(\alpha_2 + k - 1), \beta_2(\alpha_2 + k)) \quad \text{for } k \geq 1.$$

Finally, under the above notation the characteristic functions for $\bar{X}_{e(q)}$ and $\underline{X}_{e(q)}$ are given by

$$
\begin{aligned}
\varphi_q^+(z) &= \mathbb{E}[e^{iz\bar{X}_{e(q)}}] = \prod_{n \leq 0} \frac{1 + iz/\beta_1(n - \alpha_1)}{1 + iz/\zeta_n^-}, \\
\varphi_q^-(z) &= \mathbb{E}[e^{iz\underline{X}_{e(q)}}] = \prod_{n \geq 0} \frac{1 + iz/\beta_2(n + \alpha_2)}{1 + iz/\zeta_n^+}.
\end{aligned}
\tag{25}
$$

In order to implement the WHMC simulation technique we need to sample from the random variables $\bar{X}_{e(q)}$ and $\underline{X}_{e(q)}$. This requires truncation of the infinite product representation in (25). (Note that this applies to any meromorphic Lévy process, since in that case φ_q^+ and φ_q^- take the same infinite product form albeit with different roots and poles generated by the function $\zeta \mapsto q + \Psi(i\zeta)$.) For example, if we truncate φ_q^- after N factors then we effectively sample from the random variable $\underline{X}_{e(q)}^N$ with characteristic function

$$
\varphi_{q,N}^-(z) = \mathbb{E}[e^{iz\underline{X}_{e(q)}^N}] = \prod_{n \geq 0}^N \frac{1 + iz/\beta_2(n + \alpha_2)}{1 + iz/\zeta_n^+}.
\tag{26}
$$

Samples from $\underline{X}_{e(q)}^N$ are easily obtained using the observation that each factor of the finite product in (26) can be rewritten as

$$
\frac{1 + iz/\beta_2(n + \alpha_2)}{1 + iz/\zeta_n^+} = \frac{\zeta_n^+}{\beta_2(n + \alpha_2)} + \left(1 - \frac{\zeta_n^+}{\beta_2(n + \alpha_2)}\right)\left(1 + \frac{iz}{\zeta_n^+}\right)^{-1}.
$$

The above is nothing but the characteristic function of a measure consisting of an atom in zero plus a defective negative exponential distribution, i.e.

$$
\frac{\zeta_n^+}{\beta_2(n + \alpha_2)}\delta_0 + \left(1 - \frac{\zeta_n^+}{\beta_2(n + \alpha_2)}\right)e(-\zeta_n^+),
\tag{27}
$$

and it is straightforward to obtain samples according to such a measure. Hence, the law of $\underline{X}_{e(q)}^N$ can be expressed as a finite sum of independent random variables with a probability measure as in (27). A similar construction is valid for the supremum. The above observation provides a very straightforward way to simulate the supremum and the infimum for general meromorphic Lévy processes and, as far as we know, it is novel. Furthermore, in [12] the mean squared error due to the truncation of the infinite product was derived as follows:

$$
\mathbb{E}[(\underline{X}_{e(q)} - \underline{X}_{e(q)}^N)^2] \leq \frac{3}{(\zeta_{N+1}^+)^2} \leq \frac{3}{\beta_2^2(\alpha_2 + N)^2} \lesssim \mathcal{O}(N^{-2}).
$$

The recent literature working with meromorphic Lévy processes typically inverts the distribution functions of $\underline{X}_{e(q)}$ and $\bar{X}_{e(q)}$ which are also available (see [18]) for producing samples.

However, these distribution functions are expressed as infinite sums; hence, truncation has to be applied in this approach as well. Furthermore, the truncated distribution function has to be inverted numerically. Since this has to happen very efficiently this inevitably introduces an additional error, and it is not easy to analyse how such an error influences the end result. The above approach does not suffer from this problem.

5.2. First passage time of Brownian motion

In this section we take X to be a standard Brownian motion. Our goal is to present some numerical evidence for the fact that 'plain' Monte Carlo (i.e. a random walk approximation of X by sampling increments of X) yields a significant bias when it comes to approximating the first passage time, as already alluded to in Section 1. As is well known we have $\mathbb{P}(\tau_u \leq s) = 2(1 - \Phi(u/\sqrt{2s}))$ for $s > 0$; furthermore, both $\bar{X}_{e(q)}$ and $-\underline{X}_{e(q)}$ follow an exponential
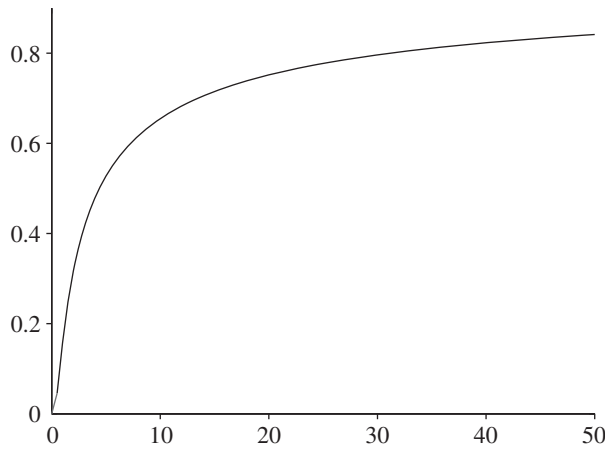

FIGURE 1: A plot of the cumulative distribution function $t \mapsto \mathbb{P}(\tau_2 \leq t)$ for a Brownian motion X.
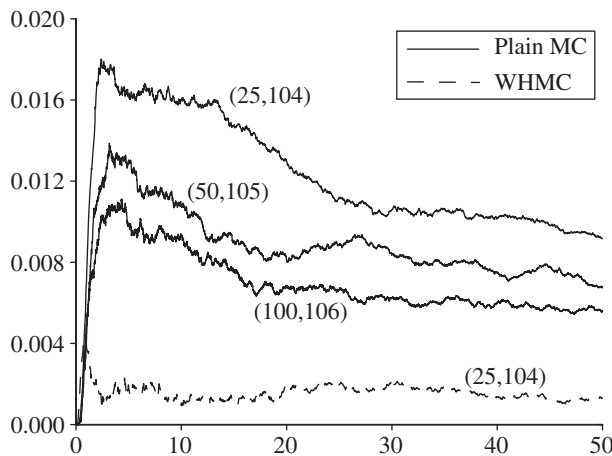

FIGURE 2: Absolute error in the approximation of the cumulative distribution function of τ_2 for plain Monte Carlo and the WHMC simulation technique. Coordinates (n, m) stand for stepsize $1/n$ ($1/2n$ for plain Monte Carlo), and number of samples m.

distribution with mean $1/\sqrt{2q}$. It is hence straightforward to implement the WHMC simulation technique and approximate τ_u. Since in the WHMC simulation technique a time step requires two samples, namely one from $\bar{X}_{e(q)}$ and one from $\underline{X}_{e(q)}$ (see Section 3), we take for the plain Monte Carlo half the step size (i.e. $1/2n$).

Figure 1 shows a plot of $t \mapsto \mathbb{P}(\tau_2 \le t)$ for $t \in [0, 50]$. Figure 2 shows the errors of both plain Monte Carlo and the WHMC simulation technique. Note also that if we decrease the step size and increase the number of samples considerably for the plain Monte Carlo approximation, the corresponding error remains several factors larger than the error the WHMC simulation technique makes.

5.3. Convergence rates

We now consider a driftless pure jump process X belonging to the β-family described in Section 5.1. In particular, we choose the coefficients in the Lévy measure (24) to be $c_i = \beta_i = \lambda_i = 1$ (for $i = 1, 2$), $\alpha_1 = 1$, and $\alpha_2 = 2$. The chosen values for c_i, β_i, and λ_i make X behave in a similar way to the variance gamma process, a popular model in finance (see [29] for the relation of the variance gamma process with the β-family). We choose different quantities for α_1 and α_2 to make the process asymmetric and ensure that a substantial quantity of the simulated paths will cross the barrier level which is set to $u = 1$. We also set the monitoring time $t = 1$. Variations of these parameters do not make significant changes in the following plots.

In Figures 3 and 4 we run a refinement sequence of Monte Carlo estimates for the mean square difference of consecutive levels of approximations for the first passage time, the overshoot, the undershoot, and the last maximum before the first passage. In particular, Figure 3 depicts the expectation

$$\mathbb{E}\left[\left(\frac{t}{n_\ell}(\kappa_u^{(n_\ell)} \wedge n_\ell) - \frac{t}{n_{\ell-1}}(\kappa_u^{(n_{\ell-1})} \wedge n_{\ell-1})\right)^2\right]$$

for $n_\ell = 2^\ell$ with $\ell = 4, \ldots, 10$. The plot uses $\log_2$-scales to show a clear decreasing rate of slope 1. This convergence rate is dictated by the result in Theorem 3 and the triangle inequality; hence, Figure 3 provides numerical evidence for Theorem 3. In Figure 4 we proceed to repeat
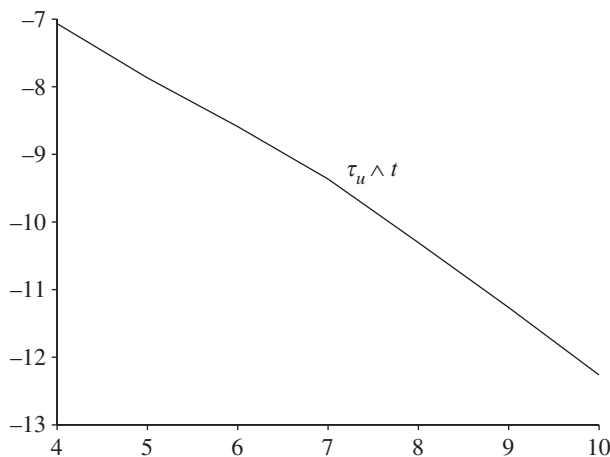

FIGURE 3: Mean square error for the difference in consecutive levels of the approximation for the first passage time (using $\log_2$-scales).
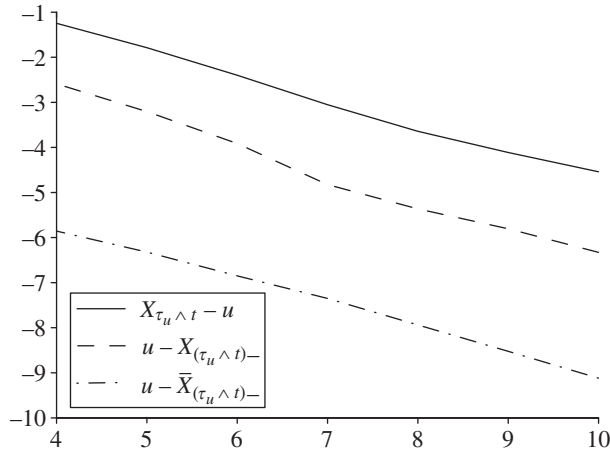
FIGURE 4: Mean square error for the difference in consecutive levels of the approximation for the overshoot, undershoot, and last maximum before passage (using $\log_2$-scales).

the same experiment for the overshoot, the undershoot, and the last maximum before the first passage. We observe numerically that the decreasing rate is very close to $\frac{1}{2}$, i.e.

$$\mathbb{E}[(V^{(n_\ell/t)}_{\kappa_u^{(n_\ell)} \wedge n_\ell} - V^{(n_{\ell-1}/t)}_{\kappa_u^{(n_{\ell-1})} \wedge n_{\ell-1}})^2] \lesssim \frac{1}{n^{1/2}},$$

for the overshoot and analogous statements for undershoot and the last maximum before the passage. The above is a particular choice of a Cauchy sequence and, therefore, it cannot prove the convergence

$$(V^{(n_\ell/t)}_{\kappa_u^{(n_\ell)} \wedge n_\ell} - u) \xrightarrow{L^2} (X_{\tau_u \wedge t} - u).$$

Nevertheless, it suggests that the above convergence is plausible and indicates its possible convergence rate. The same conclusions can be derived for the overshoot and the last maximum before the first passage time according to Figure 4.

5.4. The joint law of first passage time and overshoot

We conclude with a final example; namely, consider

$$v(u, y, q) := \mathbb{E}[e^{-q\tau_u} \mathbf{1}_{\{X_{\tau_u} - u \leq y\}}]. \tag{28}$$

This expression depends both on the first passage time and the overshoot. Indeed, it is a simplified version of the quintuple law mentioned in Section 1 (see (3)). Note that the algorithm described in Section 2 will be the same for any functional f in (2) and, therefore, we only provide a simplified example of (2) for the sake of completeness. Note that any comprehensive numerical analysis of (2) will require some assumptions on the smoothness of f; see, for instance, [13] for an error analysis of $\mathbb{E}[f(\bar{X}_1, X_1)]$, where f is uniformly Lipschitz. Indeed, the above quantity is a simplified version of the Gerber–Shiu penalty function as used in the insurance literature (see, e.g. [2] and the references therein).

For meromorphic Lévy processes a closed form expression is available for v (which is a reason for choosing this example) in terms of the roots and poles of the function $\zeta \mapsto q + \Psi(i\zeta)$ (see [20, Theorem 3]). We choose X to be the same as in Section 5.3. Figures 5 and 6 depict the performance of our algorithm as we decrease the exponential step rate $1/n$. We set the
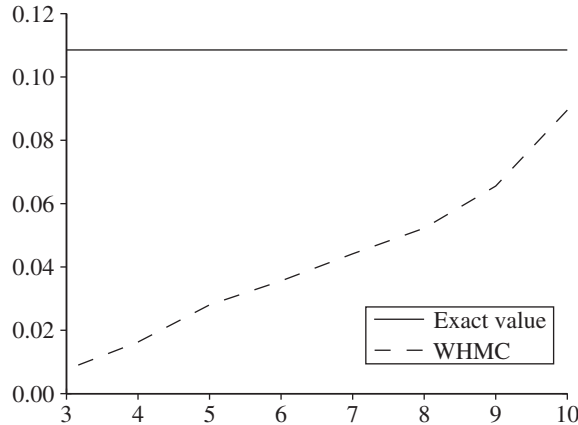
FIGURE 5: Approximation of $v(0.1, 0.05, 1)$ against rate of steps (in $\log_2$-scale) for the WHMC simulation method.
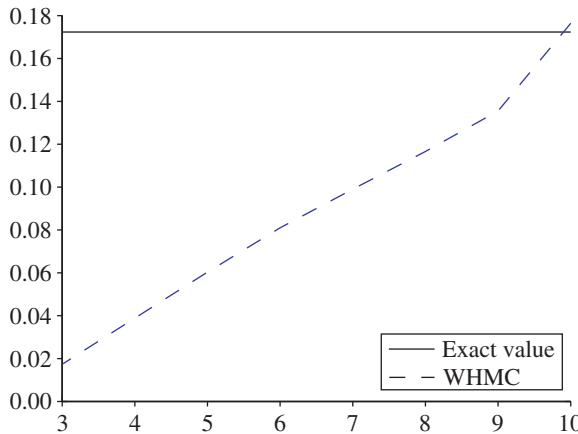

FIGURE 6: Approximation of $v(0.15, 0.15, 1)$ against rate of steps (in $\log_2$-scale) for the WHMC simulation method.

monitoring time $t = 10$ to let the process cross the barrier in most of the samples. (Note that τ_u is defined as the first passage time when the process is left to run indefinitely, in a practical implementation such as this one we have to consider $\tau_u \wedge t$ instead. However, due to the exponential decay, in (28) the effect of replacing τ_u by $\tau_u \wedge t$ can be neglected for t large enough.) Both plots in Figures 5 and 6 show a similar behaviour consistent with the results and convergence rates in [12], [13], and [29] (which only approximated the joint distribution $(X_t, \bar{X}_t)$) and the results in Section 5.3. This is a further indication that the approximation of the 4-tuple exhibits similar behaviour as the approximation of $(X_t, \bar{X}_t)$ derived in earlier work.

References

[1] ASMUSSEN, S. (2000). *Ruin Probabilities*. World Scientific, River Edge, NJ.

[2] AVRAM, F., PALMOWSKI, Z. AND PISTORIUS, M. R. (2014). On Gerber–Shiu functions and optimal dividend distribution for a Lévy risk-process in the presence of a penalty function. *Preprint*. Available at http://uk.arxiv.org/abs/1110.4965.

[3] BARNDORFF-NIELSEN, O. E. (1998). Processes of normal inverse Gaussian type. *Finance Stoch.* **2,** 41–68.

[4] BERTOIN, J. (1996). *Lévy Processes* (Camb. Tracts Math. **121**). Cambridge University Press.

[5] BROADIE, M., GLASSERMAN, P. AND KOU, S. G. (1999). Connecting discrete and continuous path-dependent options. *Finance Stoch.* **3,** 55–82.

[6] CARR, P. (1998). Randomization and the American put. *Rev. Financial Studies* **11,** 597–626.

[7] CARR, P., GEMAN, H., MADAN, D. B. AND YOR, M. (2002). The fine structure of asset returns: an empirical investigation. *J. Business* **75,** 305–333.

[8] CONT, R. AND TANKOV, P. (2004). *Financial Modelling with Jump Processes*. Chapman & Hall/CRC, Boca Raton, FL.

[9] CHEN, Z., FENG, L. AND LIN, X. (2012). Simulating Lévy processes from their characteristic functions and financial applications. *ACM Trans. Model. Comput. Simul.* **22,** 26 pp.

[10] DONEY, R. A. (2004). Stochastic bounds for Lévy processes. *Ann. Prob.* **32,** 1545–1552.

[11] DONEY, R. A. AND KYPRIANOU, A. E. (2006). Overshoots and undershoots of Lévy processes. *Ann. Appl. Prob.* **16,** 91–106.

[12] FERREIRO-CASTILLA, A. AND SCHOUTENS, W. (2012). The β-Meixner model. *J. Comput. Appl. Math.* **236,** 2466–2476.

[13] FERREIRO-CASTILLA, A., KYPRIANOU, A. E., SCHEICHL, R. AND SURYANARAYANA, G. (2014). Multilevel Monte Carlo simulation for Lévy processes based on the Wiener–Hopf factorization. *Stoch. Process. Appl.* **124,** 985–1010.

[14] GILES, M. B. (2008). Multilevel Monte Carlo path simulation. *Operat. Res.* **56,** 607–617.

[15] GLASSERMAN, P. AND LIU, Z. (2010). Sensitivity estimates from characteristic functions. *Operat. Res.* **58,** 1611–1623.

[16] HUBALEK, F. AND KYPRIANOU, A. E. (2011). Old and new examples of scale functions for spectrally negative Lévy processes. In *Seminar on Stochastic Analysis, Random Fields and Applications*, Birkhäuser, Basel, pp. 119-145.

[17] KLÜPPELBERG, C., KYPRIANOU, A. E. AND MALLER, R. A. (2004). Ruin probabilities and overshoots for general Lévy insurance risk processes. *Ann. Appl. Prob.* **14,** 1766–1801.

[18] KUZNETSOV, A. (2010). Wiener–Hopf factorization and distribution of extrema for a family of Lévy processes. *Ann. Appl. Prob.* **20,** 1801–1830.

[19] KUZNETSOV, A. (2010). Wiener–Hopf factorization for a family of Lévy processes related to theta functions. *J. Appl. Prob.* **47,** 1023–1033.

[20] KUZNETSOV, A., KYPRIANOU, A. E. AND PARDO, J. C. (2012). Meromorphic Lévy processes and their fluctuation identities. *Ann. Appl. Prob.* **22,** 1101–1135.

[21] KUZNETSOV, A., KYPRIANOU, A. E., PARDO, J. C. AND VAN SCHAIK, K. (2011). A Wiener–Hopf Monte Carlo simulation technique for Lévy processes. *Ann. Appl. Prob.* **21,** 2171–2190.

[22] KYPRIANOU, A. E. (2006). *Introductory Lectures on Fluctuations of Lévy Processes with Applications*. Springer, Berlin.

[23] KYPRIANOU, A. E., PARDO, J. C. AND RIVERO, V. (2010). Exact and asymptotic n-tuple laws at first and last passage. *Ann. Appl. Prob.* **20,** 522–564.

[24] LUNDBERG, F. (1903). *Approximerad Framställning av Sannolikhetsfunktionen. Återförsäkring av Kollektivrisker*. Almqvist. och Wiksell, Uppsala.

[25] MERTON, R. C. (1976). Option pricing when the underlying stock returns are discontinuous. *J. Financial Econom.* **3,** 125–144.

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

[27] SCHOUTENS, W. (2003). *Lévy Processes in Finance: Pricing Financial Derivatives.* John Wiley, Chichester.

[28] SCHOUTENS, W. AND CARIBONI, J. (2009). *Lévy Processes in Credit Risk.* John Wiley, Chichester.

[29] SCHOUTENS, W. AND VAN DAMME, G. (2011). The β-variance gamma model. *Rev. Deriv. Res.* **14,** 263–282.

[30] SONG, R. AND VONDRAČEK, Z. (2008). On suprema of Lévy processes and application in risk theory. *Ann. Inst. H. Poincaré Prob. Statist.* **44,** 977–986.

[31] VIGON, V. (2002). *Simplifiez vos Lévy en titillant la factorisation de Wiener–Hopf*. Doctoral Thesis, Laboratoire de Mathématiques de L'INSA de Rouen.