arXiv:2112.13874v1 [stat.CO] 27 Dec 2021

Unbiased Parameter Inference for a Class of Partially Observed Lévy-Process Models

BY HAMZA RUZAYQAT & AJAY JASRA

Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, 23955-6900, KSA. E-Mail: hamza.ruzayqat@kaust.edu.sa, ajay.jasra@kaust.edu.sa

Abstract

We consider the problem of static Bayesian inference for partially observed Lévy-process models. We develop a methodology which allows one to infer static parameters and some states of the process, without a bias from the time-discretization of the afore-mentioned Lévy process. The unbiased method is exceptionally amenable to parallel implementation and can be computationally efficient relative to competing approaches. We implement the method on S &P 500 log-return daily data and compare it to some Markov chain Monte Carlo (MCMC) algorithms.

Keywords: Unbiased inference, Lévy process, multilevel/sequential Monte Carlo methods.

1 Introduction

Lévy processes are one of the most important classes of stochastic processes. They have been widely used in applications including economics, finance, differential geometry, meteorology, turbulence, laser cooling and quantum field theory, to mention just a few. A comprehensive overview of applications of Lévy processes can be found in e.g. [3, 26]. For instance in economics and finance, Lévy processes are used to model insurance risks, stock prices, exchange and interest rates. They provide more representational models that can capture extreme behaviors like the sudden and discontinuous moves in the price and market volatility, which other classical linear and nonlinear diffusion processes, e.g., Black-Scholes model and Cox-Ingersoll-Ross, usually fail to capture.

In this article, we are concerned with statistical (parameter) inference associated to partially observed Lévy processes that are solutions to stochastic differential equations (SDEs) and observations are made at discrete and regular times. The particular model structure that we follow is that of a hidden Markov model (HMM) e.g. [7] with unknown, static (fixed in time) and finite dimensional parameters. In particular, we are concerned with Bayesian static parameter inference associated to one fixed data set. One of the main challenges with such Bayesian inference is that the transition density of the Lévy process is often intractable; that is, it does not admit an analytical form, nor does there exist a non-negative and unbiased estimator of it. In such scenarios, one often has to resort to time-discretization, leading to biased inference. In this paper we consider the Euler-discretization for Lévy driven SDEs, see [32, 29].

Many works in literature have considered Bayesian inference for diffusions and diffusions with jumps. For example, the articles [17, 31] consider fully observed (jump) diffusions, while the works [4, 14, 18, 22, 24] consider jump diffusions that are partially observed at discrete times. The main objective of this article is to perform Bayesian inference of partially observed Lévy driven SDEs with no time-discretisation error, i.e. *unbiased Bayesian inference*. The main benefits for doing so include being able to provide an exact benchmark for inference and an embarrassingly parallel estimator that is based upon independent realizations. Our work is mainly inspired by the unbiased inference approach introduced in [8] for partially observed diffusions driven by Brownian motion. In particular, given some function φ defined on the space of parameters and hidden states, the objective is to unbiasedly estimate the expectation of φ with respect to (w.r.t.) the posterior distribution over the parameters and hidden states, given that it is well-defined, where the latter are diffusions driven by a Lévy process.

When considering inference that relies on time-discretizations, a high computational cost is often required to achieve smaller bias errors. For instance, as seen in Section 6, running the particle marginal Metropolis-Hastings (PMMH) algorithm in [1] at a moderate discretization size to infer partially observed diffusions driven by Lévy process will be computationally expensive compared to the method we propose. The scheme we propose consists of two main tasks; the first is to perform PMMH at a lower discretization level, which can be done cheaply, whereas the second is a correction step. A debiasing scheme as in [30] combined with the particle filter (PF) coupling of [21] is implemented in the correction step. The unbiasedness scheme is based on similar ideas to multilevel Monte Carlo (MLMC) (see [15]), but instead of optimal allocation of computational resources for minimizing the error, the main focus is to provide unbiased estimators. Ultimately, the approach presented here will result in an unbiased estimator of the expectation of φ at a relatively small computational cost, in comparison to some competing methods.

This article is structured as follows. In Section 2, we provide the needed notations, the HMM model of interest and the methodology we follow. A brief introduction to Lévy process is also provided in this section. Section 3 describes a method to numerically approximate a Lévy process, and also explains the Euler scheme of [10] for SDEs driven by Lévy processes. We start Section 4 by reviewing the PF approach and MLMC in the context of SDEs driven by Brownian motions and Lévy processes. In the same section, we illustrate the process of generating coupled samples of the diffusion process which then used to compute an unbiased estimator of level differences that appear in the debiasing scheme. The overall methodology is detailed in Section 5 where we provide the Bayesian model and main algorithm for unbiased Bayesian inference. Finally, in Section 6, we implement the algorithm on real daily S &P 500 log-return data set and compare it to PMMH method.

2 Partially Observed Lévy-Process Model

In this section we start by giving notations that will be used throughout the paper. Then, we describe the model and methodology adopted and give a brief introduction to Lévy process.

2.1 Notations

Let $(\mathsf{X}, \mathcal{X})$ be a measurable space. We denote the empty set by ϕ. For $\varphi : \mathsf{X} \to \mathbb{R}$, we write $\mathcal{B}_b(\mathsf{X})$ as the collection of bounded measurable functions. Let $\varphi : \mathbb{R}^d \to \mathbb{R}$, $\text{Lip}_{\|\cdot\|_2}(\mathbb{R}^d)$ denote the collection of real-valued functions that are Lipschitz w.r.t. $\|\cdot\|_2$ (the $\mathbb{L}_2-$norm of a vector $x \in \mathbb{R}^d$.) That is, $\varphi \in \text{Lip}_{\|\cdot\|_2}(\mathbb{R}^d)$ if there exists a $C < +\infty$ such that for any $(x, y) \in \mathbb{R}^{2d}$, $|\varphi(x) - \varphi(y)| \leq C\|x - y\|_2$. $\mathcal{P}(\mathsf{X})$ denotes the collection of probability measures on $(\mathsf{X}, \mathcal{X})$. For a measure μ on $(\mathsf{X}, \mathcal{X})$ and a $\varphi \in \mathcal{B}_b(\mathsf{X})$, the notation $\mu(\varphi) = \int_{\mathsf{X}} \varphi(x)\mu(dx)$ is used. $\mathsf{B}(\mathbb{R}^d)$ denote the Borel sets on $\mathbb{R}^d$. dx is used to denote the Lebesgue measure. Let $M : \mathsf{X} \times \mathcal{X} \to [0, \infty)$ be a non-negative operator and μ a measure then we use the notations $\mu M(dy) = \int_{\mathsf{X}} \mu(dx)M(x, dy)$ and for $\varphi \in \mathcal{B}_b(\mathsf{X})$, $M(\varphi)(x) = \int_{\mathsf{X}} \varphi(y)M(x, dy)$. For $A \in \mathcal{X}$ the indicator is written $\mathbb{I}_A(x)$. $\mathcal{N}_s(\mu, \Sigma)$ denotes an $s-$dimensional Gaussian distribution of mean μ and covariance Σ. For a vector X, we use the notation $X_{u:q}$ to denote the vector's elements $\{X_u, X_{u+1}, \cdots, X_q\}$. For X & Y vectors in $\mathbb{R}^d$, $\langle X, Y\rangle := \sum_{i=1}^d X_i\ Y_i$ is the inner product on $\mathbb{R}^d$. We denote by I_d the identity matrix of size

$d \times d$. For $\mu \in \mathcal{P}(\mathsf{X})$ and X a random variable on X with distribution associated to μ we use the notation $X \sim \mu(\cdot)$. We denote by $a \wedge b$ the minimum of $\{a, b\}$.

2.2 Model

Let $T_f \in \mathbb{N}$ be given. We consider a probability space $(\mathsf{X}, \mathcal{X}, \{\mathcal{X}_t\}_{t\in[0,T_f]}, \mathbb{P})$, where $\{\mathcal{X}_t\}_{t\in[0,T_f]}$ is a right-continuous filtration on $\mathcal{X}$ and a stochastic process $\{Y_t\}_{t\in[0,T_f]}$, defined through the following SDE

$$dY_t = f_\theta(Y_{t^-})\ dX_t, \tag{1}$$

where $Y_0 = y_0 \in \mathbb{R}^d$ is known, $\theta \in \Theta \subseteq \mathbb{R}^{d_\theta}$ is a static model parameter, $f_\theta : \mathbb{R}^d \to \mathbb{R}^{d\times r}$ is a function, and $\{X_t\}_{t\in[0,T_f]}$ is an r-dimensional Lévy process. For background on Lévy processes and analysis of SDEs driven by the processes, we refer readers to the books [2, 3, 5, 26, 28, 34]. We assume that there are partial noisy observations, $\{Z_i = z_i\}_{i=1}^{T_f}$, $z_i \in \mathbb{R}^m$, of the diffusion process at discrete times $\{1, \ldots, T_f\}$; to connect with the standard HMM literature we will set $n = T_f$ and use the notation exchangeably. It is assumed that conditional on Y_i, Z_i is independent of the random variables $\{Y_j\}_{j\neq j}$ and that

$$Z_i|Y_i = y_i \sim g_\theta(y_i, \cdot) =: G_i^{(\theta)}(y_i), \tag{2}$$

where $G_i^{(\theta)}$ is a non-negative, bounded and measurable function that can depend on the unknown parameters θ; we suppress the dependence on the observations. The model given by (1)-(2) is an HMM and is often referred to in the literature as the state-space model. To complete the specification, one requires a prior on θ, $\overline{\Pi}(\cdot)$, but for now we shall leave the precise description for later on in the article. We note, however, that we will abuse the notation and use $\overline{\Pi}$ to denote measure and density simultaneously; this applies to all distributions/densities included in this paper as well. Therefore, the model of interest can be written as

$$\Pi(d\theta, dy_{1:n}|z_{1:n}) \propto \left(\prod_{i=1}^{n} G_i^{(\theta)}(y_i) M_i^{(\theta,\infty)}(y_{i-1}, dy_i)\right) \overline{\Pi}(d\theta)$$

where $M_i^{(\theta,\infty)}$ is the transition density induced by (1) that is assumed to exist; the significance of the superscript ∞ will become clear below. In practice, one can seldom work with the aforementioned probability and so one often resorts to a time-discretization of the Lévy process, in-between observation times. That is, let $l \in \mathbb{N} \cup \{0\}$ be a parameter that controls the number of discretization points (e.g. 2^l), then time discretization induces a collection of auxiliary random variables U_i^l ($i \in \{1, \ldots, n\}$) on a space U^l say, whose dimension grows (in-definitely with l), so that one has

$$M_i^{(\theta,l)}(y_{i-1}, dy_i) := \int_{\mathsf{U}^l} \overline{M}_i^{(\theta,l)}(y_{i-1}, dy_i, du_i^l) \tag{3}$$

for some positive transition kernel $\overline{M}_i^{(\theta,l)}(y_{i-1}, dy_i, du_i^l)$. Note that $M_i^{(\theta,l)}(y_{i-1}, dy_i)$ is constructed so that for each $i \in \{1, \ldots, n\}$ and $(y_{i-1}, A) \in \mathbb{R}^d \times \mathsf{B}(\mathbb{R}^d)$

$$\lim_{l\to\infty} \int_A M_i^{(\theta,l)}(y_{i-1}, dy_i) = \int_A M_i^{(\theta,\infty)}(y_{i-1}, dy_i).$$

In general, one has to resort to working with

$$\Pi^{(l)}(d\theta, dy_{1:n}, du^l_{1:n}|z_{1:n}) \propto \left(\prod_{i=1}^{n} G_i^{(\theta)}(y_i)\overline{M}_i^{(\theta,l)}(y_{i-1}, dy_i, du^l_i)\right)\overline{\Pi}(d\theta).$$

The exact nature of the random variables U^l_i will be discussed later on in the article. We refer to l as the level of discretization.

In this paper, we develop an unbiased static parameter inference for the HMM (1)-(2). The main algorithm consists of two steps:

Step 1. A PMMH algorithm [1] targeting a coarse-level (low l) model that can be implemented inexpensively.

Step 2. An importance sampling (IS) type correction (see [35]) that comprises

(a) a PF coupling as in [21],

(b) a single-term randomized multilevel Monte Carlo (MLMC) type estimator (see e.g [27, 30]), and

(c) a multilevel (ML) estimation of the Lévy driven SDE (see [10]).

The PMMH algorithm in Step 1 is basically a Metropolis-Hastings (MH) MCMC algorithm accompined with an Euler type discretization of the given SDE. PMMH targets the full joint posterior distribution, $\Pi(d\theta, dy_{1:n}|z_{1:n})$, for certain state-space models by jointly updating the parameter θ and the states $y_{1:n}$.

The posterior $\Pi(d\theta, dy_{1:n}|z_{1:n})$ can be decomposed as

$$\Pi(d\theta, dy_{1:n}|z_{1:n}) \propto \overline{\Pi}(d\theta)\ \eta_n^{(\theta)}(dy_{1:n}|z_{1:n})\ \mathcal{L}_n^{(\theta)}(z_{1:n}),$$

where $\eta_n^{(\theta)}(dy_{1:n}|z_{1:n})$ is the joint probability measure of the observations and the diffusion process defined through (1) at unit times $1, \cdots, n \in \mathbb{N}$, given the parameter θ, and $\mathcal{L}_n^{(\theta)}(z_{1:n})$ is the likelihood of the observations, $Z_{1:n}$, given the parameter θ. The proposal density

$$q_n(\widetilde{\theta}, \widetilde{y}_{1:n}|\theta, y_{1:n}) := q(\widetilde{\theta}|\theta)\ \eta_n^{(\widetilde{\theta})}(\widetilde{y}_{1:n}|z_{1:n})$$

is used in the MH update step to propose $\widetilde{\theta} \sim q(\cdot|\theta)$ and $\widetilde{y}_{1:n} \sim \eta_n^{(\widetilde{\theta})}(\cdot|z_{1:n})$, which then will lead to acceptance probability given by

$$1 \wedge \frac{\Pi(\widetilde{\theta}, \widetilde{y}_{1:n}|z_{1:n})\ q_n(\theta, y_{1:n}|\widetilde{\theta}, \widetilde{y}_{1:n})}{\Pi(\theta, y_{1:n}|z_{1:n})\ q_n(\widetilde{\theta}, \widetilde{y}_{1:n}|\theta, y_{1:n})} = 1 \wedge \frac{\overline{\Pi}(\widetilde{\theta})\ \mathcal{L}_n^{(\widetilde{\theta})}(z_{1:n})\ q(\theta|\widetilde{\theta})}{\overline{\Pi}(\theta)\ \mathcal{L}_n^{(\theta)}(z_{1:n})\ q(\widetilde{\theta}|\theta)}.$$

In this article, we use PMMH algorithm to generate samples from the approximate posterior

$$\Pi^{(l)}(d\theta, dy_{1:n}|z_{1:n}) = \int_{\mathsf{U}^l} \Pi^{(l)}(d\theta, dy_{1:n}, du^l_{1:n}|z_{1:n}),$$

which requires sampling from an approximation of $\eta_n^{(\theta)}(dy_{1:n}|z_{1:n})$ (namely $\eta_n^{(\theta,l)}$ defined in Subsection 4.1) that is done through the PF algorithm. The samples resulting from the PF are also used to estimate the likelihood term that appears in the formula of MH acceptance probability.

The bias associated to running PMMH at a low discretization level is dealt with through the correction process in Step 2. The idea is to construct an approximate coupling of the posterior Π over the parameter and hidden states at two consecutive discretization levels (with a fixed parameter), namely l and $l-1$, for some $l \in \mathbb{N}$, then correcting by an IS method. This specific coupling will result in an unbiased estimation of the difference of two unnormalised integrals over the hidden states (see Subsection 4.3 for more details), a difference that appears in a ML collapsing sum identity. As mentioned in the introduction, the main focus of this article is to unbiasedly estimate the expectation of some given function φ that is defined over $\Theta \times \mathbb{R}^{nd}$ w.r.t. the joint posterior distribution Π at some time $n \in \mathbb{N}$. The unbiasedness arises from running the coupled PF along with the IS correction at random discretization levels, $l \in \mathbb{N}$, sampled from a given probability mass function (PMF) P_l.

This approach was presented in [8] for partially observed diffusions driven by a Wiener process. In this article, we apply this method to partially observed models driven by a Lévy process. In order to properly describe our methodology, we will have to detail several aspects, and this is the topic of the following sections. We start by giving a brief background of Lévy processes and their properties.

2.3 Lévy Processes

Lévy processes are stochastic processes that are almost surely zero at time zero, continuous in probability with independent and stationary increments. It can be thought of as random walks in continuous time. Examples of Lévy processes include Wiener, Poisson and Gamma processes. The distribution of a Lévy process $\{X_t\}_{t\in[0,1]}$ is determined by its characteristic function given by the celebrated Lévy-Khintchine formula

$$\mathbb{E}[\exp\{i\langle w, X_t\rangle\}] = \psi_1(t,w) + \psi_2(t,w) + \psi_3(t,w),$$

for all $w \in \mathbb{R}^r$, such that

$$\psi_1(t,w) = \exp\{it\langle b, w\rangle\}, \quad \psi_2(t,w) = \exp\left\{-\frac{t}{2}\langle w, \Sigma w\rangle\right\}, \text{ and}$$
$$\psi_3(t,w) = \exp\left\{t\int_{\mathbb{R}^r\setminus\{0\}} \left[\exp\{i\langle w,x\rangle\} - 1 - i\langle w,x\rangle \mathbb{I}_{B_1}(x)\right]\nu(dx)\right\},$$

where $b \in \mathbb{R}^r$, $0 \leq \Sigma \in \mathbb{R}^{r\times r}$ is symmetric, $B_\delta = \{x \in \mathbb{R}^r \;:\; \|x\|_2 < \delta\}$, and ν is called the Lévy measure of X_t satisfying $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}^r}(1 \wedge \|x\|_2^2)\nu(dx) < \infty$. Note that this condition on ν implies that $\nu(B_\delta) < \infty$ for all $\delta > 0$, see [25]. Since probability distributions are uniquely determined by characteristic functions, the Lévy process X_t is uniquely determined by the so-called Lévy triplet (ν, Σ, b). The distribution of a Lévy process is infinitely divisible, that is, the law of a Lévy process at time t is the sum of the n laws of the Lévy process increments over time intervals of length t/n, where n is some given positive integer. Conversely, if μ is an infinitely divisible distribution then there exists a Lévy process $\{X_t\}_{t\in[0,1]}$ such that $X_1 \sim \mu$. Moreover, Lévy-Khintchine and Lévy-Itô decomposition theorems propose that every Lévy process is the sum of three Lévy processes, which is again a Lévy process (see [34] for more details.) The terms are a linear Brownian motion (which corresponds to ψ_1 and ψ_2) and a Lévy jump process (which corresponds to ψ_3.) For a fixed $\delta > 0$, note that ψ_3 can be written as

$$\psi_3(t,w) = \exp\left\{t\nu(B_\delta^c)\int_{\mathbb{R}^r}\left[\exp\{i\langle w,x\rangle\} - 1\right]\mu(dx) + t\int_{\mathbb{R}^r}\left[\exp\{i\langle w,x\rangle\} - 1 - i\langle w,x\rangle\right]\zeta(dx)\right\},$$

where $\mu(dx) = \mathbb{I}_{B_\delta^c}(x)\nu(dx)/\nu(B_\delta^c)$ and $\zeta(dx) = \mathbb{I}_{B_\delta \setminus \{0\}}(x)\nu(dx)$. By Lévy-Itô decomposition theorem, the first integral in the above identity corresponds to a compound Poisson process with intensity $\nu(B_\delta^c)$ and distribution μ, while the second integral corresponds to a compensated Poisson process that is a square-integrable martingale [25]. Then, any Lévy process $\{X_t\}_{t\in[0,1]}$ can be written as

$$X_t = tb + \Sigma^{1/2} W_t + L_t, \tag{4}$$

where the process L_t is independent of the linear Brownian motion. An important parameter that helps with understanding the behavior of a Lévy process is Blumenthal-Getoor (BG) index [6] defined as

$$\Gamma := \inf \left\{ p > 0 \ : \ \int_{B_1} |x|^p \ \nu(dx) < \infty \right\} \in [0, 2].$$

The BG index measures the frequency of small jumps. Large index corresponds to a Lévy process that has small jumps at high frequencies.

We assume that the Lévy process $\{X_t\}_{t\in[0,T_f]}$ defined in (4) and the Lévy-driven process $\{Y_t\}_{t\in[0,T_f]}$ given in (1) satisfy the following assumptions:

Assumption 2.1. *There exists $C_1, C_2, C_3 > 0$ such that*

(i) $\|f_\theta(x) - f_\theta(y)\|_2 \le C_1 \|x - y\|_2$ *and* $\|f_\theta(x)\|_2 \le C_2$ *for all* $x \in \mathbb{R}^d$*,*

(ii) $0 < \int_{\mathbb{R}^d} |x|^2 \ \nu(dx) \le C_3$.

Note that Assumption 2.1 (i) is different from [10, Assumption A] where the authors only assume that $\|f_\theta(y_0)\|_2 \le C$, with y_0 as in (1).

It is generally impossible to exactly simulate the increments of a Lévy process. Therefore, in addition to an Euler scheme that is used to approximate the process $\{Y_t\}_{t\in[0,T_f]}$, one first has to approximately simulate the Lévy increments. In this paper, we assume that one cannot exactly sample from the law of X_t, hence of Y_t, but rather numerical approximation of such processes is adopted. In the following section, we show how one can approximately simulate the increments of a Lévy process which are then used within an Euler scheme to give approximate samples of the process $\{Y_t\}_{t\ge 0}$.

3 Numerical Approximations

Here we describe an Euler scheme in which one can produce approximate solutions to the Lévy-driven SDE in (1). We adopt the approach in [29] (see also [19, 32, 10]) which is summarized in Subsection 3.1 and Subsection 3.2 below.

3.1 A Numerical Approximation to Lévy Process

We first consider an approximation scheme to simulate the Lévy process over the time interval $[0, 1]$. Let $\Delta_l = 2^{-l}$, $l \in \mathbb{N} \cup \{0\}$, be given and denote by $\delta_l > 0$ the jump threshold parameter so that jumps with height smaller than δ_l are ignored. We will refer to l as the level of discretization. Define $\lambda_l := \nu(B_{\delta_l}^c)$, then given $l \in \mathbb{N} \cup \{0\}$, we choose δ_l such that $\lambda_l = \Delta_l^{-1}$, which assures that $\lambda_l > 0$ for sufficiently small $\delta_l > 0$. Let $\{N(t, A)\}_{t\ge 0, A\in \mathsf{B}(\mathbb{R}^r\setminus\{0\})}$ denote a Poisson process with rate $\nu(A)$. Then, the number of jumps with heights larger than δ_l before time t is a Poisson process

$N(t, B^c_{\delta_l})$ with rate λ_l. We denote by $\Delta L_t := L_t - \lim_{s \uparrow t} L_s$ the jump height at discontinuity time t and $\Delta L_{\widehat{T}^l_i} := L_{\widehat{T}^l_i} - L_{\widehat{T}^l_{i-1}}$ the jump height at discontinuity time $\widehat{T}^l_i$, where l is the discretization level and $i \in \{1, \cdots, \widehat{K}\}$ for some $\widehat{K} \in \mathbb{N}$ such that $\widehat{T}^l_0 = 0$ and $\widehat{T}^l_{\widehat{K}} = 1$. Set

$$F^l_0 := \int_{B^c_{\delta_l}} x \; \nu(dx)$$

and

$$\mu^l(dx) := \frac{1}{\lambda_l} \mathbb{I}_{B^c_{\delta_l}}(x) \nu(dx),$$

then $\mu^l(dx)$ defines a probability measure on $\mathbb{R}^r \setminus \{0\}$ such that $\int_{B^c_{\delta_l}} x \; N(t, dx)$ is a random finite sum of random variables that are sampled from $\mu^l(dx)$. These random variables are the heights of all jumps larger than δ_l before time t. The compensated compound Poisson process defined by

$$L^{\delta_l}_t := \int_{B^c_{\delta_l}} x \; N(t, dx) - tF^l_0 = \sum_{i=1}^{N(t, B^c_{\delta_l})} \Delta L_{\widehat{T}^l_i} - tF^l_0$$

is an L_2 martingale which converges in L_2 as $\delta_l \to 0$ to the Lévy process L_t in (4) [2, 10]. The term tF^l_0 is the expected total length of all jumps before time t given by $\mathbb{E}\Big[\int_{B^c_{\delta_l}} x \; N(t, dx) \Big]$. The time increments $\{\widehat{T}^l_i - \widehat{T}^l_{i-1}\}_{i=1}^{\widehat{K}}$ of the jumps with heights larger than δ_l are sampled from an exponential distribution with parameter λ_l. However, since we want the Brownian increments $W_{\widehat{T}^l_i} - W_{\widehat{T}^l_{i-1}}$ to be sampled from a normal distribution with covariance matrix $\alpha_i \, I_r$ such that α_i is at most Δ_l, a refinement of the jump times is required. Set $T^l_0 = 0$, then the new jump times are

$$T^l_i = \min \Big\{ T^l_{i-1} + \Delta_l, \; \min\{\widehat{T}^l_j > T^l_{i-1} \; : \; j \in \{1, \cdots, \widehat{K}\}\Big\},$$

for $i \in \{1, \cdots, K\}$, where $K \in \mathbb{N}$ and $T^l_K = 1$. To summarize, increments of the process $\{X_t\}_{t \in [0,1]}$ are approximated at discretization times $\{T^l_i\}_{i=1}^K$ by

$$\Delta X^l_{T^l_i} = (b - F^l_0)(T^l_i - T^l_{i-1}) + \Sigma^{1/2}(W_{T^l_i} - W_{T^l_{i-1}}) + \Delta L_{T^l_i}. \qquad (5)$$

If $T^l_i = \widehat{T}^l_j$ for some $j \in \{1, \cdots, \widehat{K}\}$, then $\Delta L_{T^l_i} = \Delta L_{\widehat{T}^l_j} \sim \mu^l$, otherwise $\Delta L_{T^l_i} = 0$. One can use the inversion sampling method to sample from μ^l if the cumulative distribution function can be computed analytically or else other methods can be used (e.g. rejection sampling.) Below, we state a needed assumption for the choice of δ_l.

Assumption 3.1. *There exists a $\kappa > 0$ such that for all $l \in \mathbb{N}$, there exists a $\delta_l > 0$ such that $\delta_l = \mathcal{O}(\Delta^\kappa_l)$ and $\lambda_l = \Delta_l^{-1}$.*

In Algorithm 1 we summarize the process of obtaining a single-level approximation of increments of the Lévy process on the interval $[0, 1]$.

Algorithm 1 : Single-level increments of the Lévy process on $[0, 1]$

```
1. Input: $l$, $\lambda_l$ and $\mu^l(dx)$.
```

```
2. Initialization: Set $\widehat{T}_0^l = 0$, $T_0^l = 0$ and $i = 1$.
3. Jump times:
    • Set $\widehat{T}_i^l = \min\{1, \widehat{T}_{i-1}^l + \zeta_i\}$, where $\zeta_i \sim \mathrm{Exp}(\lambda_l)$.
    • If $\widehat{T}_i^l = 1$, set $\widehat{K} = i$ and go to Step 4..
    • Otherwise, $i = i + 1$ and go to start of Step 3..
4. Jump heights: For $i = 1, \cdots, \widehat{K}$, sample $\Delta L_{\widehat{T}_i^l} \sim \mu^l$. Set $i = 1$.
5. Refinement of time increments:
    • $T_i^l = \min\left\{T_{i-1}^l + \Delta_l,\ \min\left\{\widehat{T}_j^l > T_{i-1}^l \ : \ j \in \{1, \cdots, \widehat{K}\}\right\}\right\}$.
    • If $T_i^l = \widehat{T}_j^l$ for some $j \in \{1, \cdots, \widehat{K}\}$, set $\Delta L_{T_i^l} = \Delta L_{\widehat{T}_j^l}$; otherwise $\Delta L_{T_i^l} = 0$.
    • If $T_i^l = 1$, set $K = i$ and go to output Step 6.
    • Otherwise, $i = i + 1$ and go to start of Step 5..
6. **Output**: Return $K$, $\{T_i^l\}_{i=0}^K$ and $\{\Delta L_{T_i^l}\}_{i=1}^K$.
```

3.2 An Euler Scheme to Lévy-driven SDEs

Assume that $\theta \in \Theta$ is fixed. The Euler approximation of the Lévy-driven process $\{Y_t\}_{t\in[0,1]}$ at discretization times $\{T_i^l\}_{i=1}^K$ is given by

$$Y_{T_i^l}^l = Y_{T_{i-1}^l}^l + f_\theta(Y_{T_{i-1}^l}^l)\Delta X_{T_i^l}^l, \tag{6}$$

with $Y_{T_0^l}^l = Y_0$ and the Lévy increments $\{\Delta X_{T_i^l}^l\}_{i=1}^K$ are computed through (5). Algorithm 2 describes the process of generating samples Y_1^l, an Euler approximation of the process $\{Y_t\}_{t\geq 0}$ at time 1 and a discretization level l. In this paper, we shall consider the discrete sequence $\{Y_0, Y_1, \cdots, Y_{T_f}\}$, where for each $n = 1, \cdots, T_f$ we approximate Y_n by Y_n^l, where $Y_n^l = Y_{T_K^l}^l$ and in the input of Algorithm 2 Y_0^l is set to Y_{n-1}^l.

Theorem 2 of [10] provides asymptotic bounds for the strong error of the Euler scheme presented in Algorithm 2. The theorem states that under Assumption 2.1, for any $l \in \mathbb{N}$ and $\delta_l > 0$ such that $\nu(B_{\delta_l}^c) = \Delta_l^{-1}$, there exists a $C < \infty$ such that

$$\begin{aligned} \mathbb{E}\left[\sup_{t\in[0,1]} \left|Y_t - Y_t^l\right|^2\right] &\leq C(\sigma_{\delta_l}^2 + \Delta_l|\log(\Delta_l)|), \qquad \Sigma \neq 0, \\ \mathbb{E}\left[\sup_{t\in[0,1]} \left|Y_t - Y_t^l\right|^2\right] &\leq C(\sigma_{\delta_l}^2 + |b - F_0^l|^2\Delta_l^2), \qquad \Sigma = 0, \end{aligned} \tag{7}$$

where $\sigma_{\delta_l}^2 := \int_{B_{\delta_l}} |x|^2 \, \nu(dx)$. As we can observe from (7) the strong error rate of the Euler scheme will only depend upon the choice of ν and whether the Lévy process has a Brownian motion component. We will show in Section 6 that for a certain choice of ν, the strong error rate β is 3.

Algorithm 2 : Single-level sampling for Lévy-driven process

```
1. Input: $Y_0$, $l$, $\lambda_l$, $F_0^l$, $\mu^l(dx)$ and $\theta$.
2. Call Algorithm 1 with input $l$, $\lambda_l$ and $\mu^l(dx)$ to return $\{T_i^l\}_{i=0}^K$ and $\{\Delta L_{T_i^l}\}_{i=1}^K$.
3. For $i = 1, \cdots, K$, sample the Brownian increments $W_{T_i^l} - W_{T_{i-1}^l}$ from $\mathcal{N}_r(0, (T_i^l - T_{i-1}^l)I_r)$.
4. With $\{T_i^l\}_{i=0}^K$, $\{\Delta L_{T_i^l}\}_{i=1}^K$, $\{W_{T_i^l} - W_{T_{i-1}^l}\}_{i=1}^K$ and $F_0^l$, generate the increments $\{\Delta X_{T_i^l}\}_{i=1}^K$
   through (5). Then, compute $Y_1^l$ via the recursion in (6).
5. Output: Return $Y_1^l$.
```

Remark 3.1. *Algorithms 1 and 2 are in fact a single-level version of a more general coupled discretization scheme given in [10] which will be described shortly in Subsection 4.2.*

The Markov transition between Y_{n-1} and Y_n, $n = 1, \cdots, T_f$, is given by the transition kernel $M^{(\theta,\infty)}(y_{n-1}, dy_n)$ of the process in (1) over unit time. Set $M_n^{(\theta,\infty)}(y_{n-1}, dy_n) = M^{(\theta,\infty)}(y_{n-1}, dy_n)$, $n = 1, \cdots, T_f$, then the pair $(M_n^{(\theta,\infty)}, G_n^{(\theta)})$ defines a Feynman-Kac type model (see [9]). In many cases, $M^{(\theta,\infty)}$ exists, but one cannot exactly sample from it or evaluate a non-negative unbiased estimator of it, and therefore, one has to work with a discretization of the model as in (6).

4 Particle Filter and Multilevel Monte Carlo Methods

We mentioned earlier in Subsection 2.2 that Step 2. of the main algorithm requires a coupling of the PF. Before discussing that, we give a brief review of the PF algorithm in Subsection 4.1. In Subsection 4.2, we summarize the MLMC method for diffusions driven by Wiener processes and that of [10] for diffusions driven by Lévy process. Then in Subsection 4.3, we explain the PF coupling of [21], which will be used in our main algorithm to produce an unbiased scheme for estimating the level-difference in (12). In all coming subsections we assume that $l \in \mathbb{N} \cup \{0\}$ and $\theta \in \Theta$ are both fixed.

4.1 Particle Filter

The PF (see for example [9, 12]) is a method that generates a collection of N samples in parallel associated with a set of weights and combines IS and resampling techniques to numerically approximate expectations of functionals w.r.t the normalized or unnormalized measures. Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a function in $\mathcal{B}_b(\mathbb{R}^d)$. Suppose that one is interested in approximating the expected value $\eta_n^{(\theta)}(\varphi) =: \mathbb{E}[\varphi(Y_{1:n})|Z_{1:n} = z_{1:n}]$, where Y_t is the solution to the Lévy-driven SDE in (1) at discrete time $t \in \mathbb{N}$, and $\eta_n^{(\theta)}(dy_{1:n})$ is the joint probability measure of the observations and the diffusion process defined through (1) at discrete times, given by

$$\eta_n^{(\theta)}(dy_{1:n}) \propto \prod_{i=1}^{n} G_i^{(\theta)}(y_i) M_i^{(\theta,\infty)}(y_{i-1}, dy_i) := \gamma_n^{(\theta)}(dy_{1:n}), \tag{8}$$

where the dependence on the observations is suppressed. We refer to $\eta_n^{(\theta)}$ and $\gamma_n^{(\theta)}$ as *normalized* and *unnormalized* smoothing measures, respectively. In practice, $M_n^{(\theta,\infty)}(y_{n-1}, dy_n)$ is typically intractable as discussed earlier on, and as a result, one considers targets associated to a time discretization given by

$$\eta_n^{(\theta,l)}(dy_{1:n}^l) \propto \prod_{i=1}^{n} G_i^{(\theta)}(y_i^l) M_i^{(\theta,l)}(y_{i-1}^l, dy_i^l) := \gamma_n^{(\theta,l)}(dy_{1:n}^l), \tag{9}$$

where $M_n^{(\theta,l)}(y_{n-1}^l, dy_n^l)$ is the induced transition kernel from the Euler approximation in (6) over unit time, and $\gamma_n^{(\theta,l)}(dy_{1:n}^l)$ refers to the unnormalized measure. Note that here we include a superscript l on $y_{1:n}$ to emphasize that they are samples at unit times from the discretized SDE in (6). Let

$$\eta_n^{(\theta,l)}(\varphi) := \int_{\mathbb{R}^{nd}} \varphi(y_n^l)\ \eta_n^{(\theta,l)}(dy_{1:n}^l), \tag{10}$$

denote the expectation of a bounded and measurable function φ w.r.t. the probability measure $\eta_n^{(\theta,l)}(dy_{1:n}^l)$ at discrete time n. We also define

$$\gamma_n^{(\theta,l)}(\varphi) := \int_{\mathbb{R}^{nd}} \varphi(\tilde{y}_n)\ \gamma_n^{(\theta,l)}(dy_{1:n}^l), \tag{11}$$

the expectation of φ w.r.t. the unnormalized measure. Note that $\eta_n^{(\theta,l)}(\varphi) = \gamma_n^{(\theta,l)}(\varphi)/\gamma_n^{(\theta,l)}(1)$. Both $\eta_n^{(\theta,l)}(\varphi)$ and $\gamma_n^{(\theta,l)}(\varphi)$ can be estimated using the PF method described in Algorithm 3.

Algorithm 3 : Particle filter

```
1. Input: Level $l$, number of particles $N$, $Y_0$, terminal time $n$, $\theta$ and potential functions $\{G_i^{(\theta)}\}_{i=1}^n$.
2. Initialization: For $i = 1, \cdots, N$ set $\mathbf{Y}_{\mathbf{0}}^{\boldsymbol{l,i}} = Y_0^{l,i} = Y_0$
3. For $k = 1, \cdots, n$, do:
    (i) For $i = 1, \cdots, N$ simulate $Y_1^{l,i}$ via (6) with $Y_0^{l,i} = Y_{k-1}^{l,i}$ and set $\widehat{\mathbf{Y}}_{\boldsymbol{k}}^{\boldsymbol{l,i}} = (\mathbf{Y}_{\boldsymbol{k-1}}^{\boldsymbol{l,i}}, Y_1^{l,i})$.
    (ii) For $i = 1, \cdots, N$ compute $w_k^i = G_k^{(\theta)}(y_1^{l,i})$ and set $W_k^i = w_k^i / \sum_{j=1}^N w_k^j$.
    (iii) Resample the particles $\{\widehat{\mathbf{Y}}_{\boldsymbol{k}}^{\boldsymbol{l,i}}\}_{i=1}^N$ according to the normalized weights $\{W_k^i\}_{i=1}^N$.
          Denote the resampled particles $\{\mathbf{Y}_{\boldsymbol{k}}^{\boldsymbol{l,i}}\}_{i=1}^N = \{(Y_0^{l,i}, \cdots, Y_k^{l,i})\}_{i=1}^N$.
4. For $i = 1, \cdots, N$ set $V^i = \left[\prod_{k=1}^{n-1} \frac{1}{N} \sum_{j=1}^N w_k^j\right] \frac{1}{N} w_n^i$.
5. Output: Return $\{\mathbf{Y}_{\boldsymbol{n}}^{\boldsymbol{l,i}}, V^i\}_{i=1}^N$.
```

We remark that Step 3(i) of Algorithm 3 corresponds to the resampling step, which can be multinomial, residual, stratified or systematic (see [11] for a comparison.) In this paper, we use multinomial resampling. It is well-known that $\sum_{i=1}^N V^i\ \varphi(\mathbf{Y}_{\boldsymbol{n}}^{\boldsymbol{l,i}})$ and $\sum_{i=1}^N W_n^i\ \varphi(\mathbf{Y}_{\boldsymbol{n}}^{\boldsymbol{l,i}})$ will converge almost surely as $N \to \infty$ to $\gamma_n^{(\theta,l)}(\varphi)$ and $\eta_n^{(\theta,l)}(\varphi)$, respectively. Moreover, the sum $\sum_{i=1}^N V^i\ \varphi(\mathbf{Y}_{\boldsymbol{n}}^{\boldsymbol{l,i}})$ is an unbiased estimator of the unnormalized smoother $\gamma_n^{(\theta,l)}(\varphi)$ (see e.g. [9, 33]).

A quantity of interest that one aims to estimate is the level difference $\eta_n^{(\theta,l)}(\varphi) - \eta_n^{(\theta,l-1)}(\varphi)$ (or $\gamma_n^{(\theta,l)}(\varphi) - \gamma_n^{(\theta,l-1)}(\varphi)$) for some $l \in \mathbb{N}$. This quantity appears in multiple algorithms like those found in the ML context (e.g. [10, 15, 20]) and in some debiasing schemes (e.g. [27, 30]). In particular, it appears in the debiasing scheme within the correction step of the main algorithm in Section 5. In Subsection 4.2, we explain how to obtain coupled (possibly biased) samples from the law of the discretized Lévy-driven process in (6). Then in Subsection 4.3, we present an unbiased estimator of the level difference

$$\gamma_n^{(\theta,l)}(\varphi) - \gamma_n^{(\theta,l-1)}(\varphi), \tag{12}$$

for given $(\theta, l) \in \Theta \times \mathbb{N}$ and function $\varphi \in \mathcal{B}_b(\mathbb{R}^d)$. Before moving forward, we require the following assumptions on the Feynman-Kac model defined by $(M_n^{(\theta,l)}, G_n^{(\theta)})$.

Assumption 4.1. *There are $c > 1$ and $C > 0$, such that for all $n, l \in \mathbb{N}$, $\varphi \in Lip_{\|\cdot\|_2}(\mathbb{R}^d) \cap \mathcal{B}_b(\mathbb{R}^d)$, and $(\theta, x, x') \in \Theta \times \mathbb{R}^d \times \mathbb{R}^d$, we have*

(i) $c^{-1} \leq G_n^{(\theta)}(x) \leq c$.

(ii) $G_n^{(\theta)}$ *is Lipschitz, i.e.,* $\|G_n^{(\theta)}(x) - G_n^{(\theta)}(x')\|_2 \leq C\|x - x'\|_2$.

(iii) $M_n^{(\theta,l)}(\varphi)$ *is Lipschitz, i.e.,* $\|M_n^{(\theta,l)}(\varphi)(x) - M_n^{(\theta,l)}(\varphi)(x')\|_2 \leq C\|x - x'\|_2$.

4.2 A Multilevel Monte Carlo Method

First, let us assume that $\{Y_t\}_{t\in[0,1]}$ is a diffusion process that is driven by a Wiener process. Suppose that we are interested in estimating the expectation $\mathbb{E}[\varphi(Y_1)] := \zeta_1(\varphi)$, where ζ_1 is the law of the process $\{Y_t\}_{t\in[0,1]}$ at time 1. Furthermore, assume that one has to work with a discretization of the the diffusion process. Let $L \in \mathbb{N}$ be the level of discretization and define $\zeta_1^L(\varphi) := \mathbb{E}[\varphi(Y_1^L)]$ the expectation w.r.t. the law ζ_1^L associated with the discretization scheme at time 1. The standard Monte Carlo (MC) provides an unbiased estimator of this expectation through

$$\zeta_1^{L,MC}(\varphi) := \frac{1}{N}\sum_{i=1}^{N} \varphi(Y_1^{L,i}),$$

where $\{Y_1^{L,i}\}_{i=1}^N$ are i.i.d. samples from ζ_1^L. Then the mean square error (MSE) of the estimator is

$$\begin{aligned}\mathbb{E}\left[\left(\zeta_1^{L,MC}(\varphi) - \zeta_1(\varphi)\right)^2\right] &= \mathbb{E}\left[\left(\zeta_1^{L,MC}(\varphi) - \zeta_1^L(\varphi) + \zeta_1^L(\varphi) - \zeta_1(\varphi)\right)^2\right] \\ &= \frac{\mathbb{V}ar[\varphi(Y_1^L)]}{N} + \left(\zeta_1^L(\varphi) - \zeta_1(\varphi)\right)^2,\end{aligned}$$

where the first term on the right is the variance and the second term is the bias squared which is of order $\mathcal{O}(\Delta_L^2)$. To get an MSE of order $\mathcal{O}(\epsilon^2)$, for a given $\epsilon > 0$, one needs to choose $N = \mathcal{O}(\epsilon^{-2})$ and $\Delta_L = \mathcal{O}(\epsilon)$. If we assume that the cost of sampling Y_1^L from ζ_1^L is $C_L = \mathcal{O}(\Delta_L^{-1})$, then to achieve an MSE of order $\mathcal{O}(\epsilon^2)$, the cost of the MC estimator will be $\mathcal{O}(N\Delta_L^{-1}) = \mathcal{O}(\epsilon^{-3})$. On the other hand, MLMC, which is proposed in [15] (see also [20]) for diffusions driven by a Wiener process, achieves this MSE at a lower cost. The idea of MLMC method is to write $\zeta_1^L(\varphi)$ as a telescoping sum

$$\zeta_1^L(\varphi) = \zeta_1^0(\varphi) + \sum_{l=1}^{L} [\zeta_1^l(\varphi) - \zeta_1^{l-1}(\varphi)].$$

The sum is estimated by approximating $\zeta_1^0(\varphi)$ through a standard MC method with N_0 independent samples, and approximating the differences $\zeta_1^l(\varphi) - \zeta_1^{l-1}(\varphi)$ by independently simulating N_l samples $\{Y_1^{l,i}, Y_1^{l-1,i}\}_{i=1}^{N_l}$ from an appropriate coupling of the diffusion process. Let $\{Y_t\}_{t\in[0,1]}$ be the solution of an SDE driven by a Wiener process and approximated with an Euler discretization, then the coupling emerges from concatenating the fine increments of the Wiener process to generate the coarse increments. The MLMC estimator of $\zeta_1^L(\varphi)$ is then

$$\zeta_1^{L,MLMC} := \frac{1}{N_0} \sum_{i=1}^{N_0} \varphi(Y_1^{0,i}) + \sum_{l=1}^{L} \frac{1}{N_l} \sum_{i=1}^{N_l} [\varphi(Y_1^{l,i}) - \varphi(Y_1^{l-1,i})].$$

Similar to before, the overall MSE for the ML estimator $\zeta_1^{L,MLMC}$ can be expressed as

$$\mathbb{E}\left[\left(\zeta_1^{L,MLMC}(\varphi) - \zeta_1(\varphi)\right)^2\right] = \sum_{l=0}^{L} \frac{\mathbb{V}ar[\varphi(Y_1^l) - \varphi(Y_1^{l-1})]}{N_l} + \left(\zeta_1^L(\varphi) - \zeta_1(\varphi)\right)^2,$$

with the convention $\varphi(Y_1^{-1}) = 0$. The sum on the right hand side is the variances and the other term is the bias squared. Let $V_l = \mathbb{V}ar[\varphi(Y_1^l) - \varphi(Y_1^{l-1})]$, then the dependent coupling aforementioned will induce the variance to be a function of Δ_l. MLMC aims at reducing the variance term $\sum_{l=0}^{L} V_l/N_l$, leaving unchanged the bias due to the Euler discretization. The level L and the number of samples $\{N_l\}_{l=0}^{L}$ are chosen to balance these terms so that the MSE is of $\mathcal{O}(\epsilon^2)$, for a given $\epsilon > 0$, and that the overall computational complexity is minimized. In particular, L is chosen so that the bias term is $\mathcal{O}(\epsilon^2)$, and given C_l and V_l as functions of Δ_l, the total cost $\sum_{l=0}^{L} C_l N_l$ is minimized with the constrain $\sum_{l=0}^{L} V_l/N_l = \mathcal{O}(\epsilon^2)$ by optimizing $\{N_l\}_{l=0}^{L}$. In [15] the author shows that the cost to achieve such an MSE using an MLMC estimator is

$$\text{Cost} = \begin{cases} \mathcal{O}(\epsilon^{-2}\log(\epsilon)^2) & \text{if} \quad \beta = 1, \\ \mathcal{O}(\epsilon^{-2}) & \text{if} \quad \beta > 1, \\ \mathcal{O}(\epsilon^{-2-(1-\beta)/\alpha}) & \text{if} \quad 0 < \beta < 1, \end{cases}$$

where α and β are respectively the weak and strong error rates of the discretization scheme.

Nevertheless, the situation is more complicated when SDEs are purely driven by general Lévy processes. An approach based on Poisson thinning has been suggested in [16] for pure-jump diffusion and by [13] for general Lévy processes. However, an alternative construction based on the Lévy-Itô decomposition is proposed in [10]. The authors of that article provide a procedure to obtain coupled increments for the Lévy process on $[0, 1]$ which then can be used in (6) to generate samples $\{Y_1^{l,i}, Y_1^{l-1,i}\}_{i=1}^{N_l}$ from a coupled-kernel induced by the Euler scheme. This construction is employed in our paper and we summarize it in Algorithms 4 and 5. We introduce the following necessary notations.

Definition 4.1. *Let $l \in \mathbb{N}$, $\theta \in \Theta$ and $\check{y}_n^l = (y_n^l, y_n^{l-1})$. Define the kernel $M_n^{(\theta,l,l-1)} : [\mathbb{R}^d \times \mathbb{R}^d] \times [\sigma(\mathbb{R}^d) \otimes \sigma(\mathbb{R}^d)] \to \mathbb{R}_+$ a coupling of the transition kernels $M_n^{(\theta,l)}(y_n^l, \cdot)$ and $M_n^{(\theta,l-1)}(y_n^{l-1}, \cdot)$ such* *that for any $A \in \sigma(\mathbb{R}^d)$, we have*

$$\int_{A\times\mathbb{R}^d} M_n^{(\theta,l,l-1)}(\check{y}_{n-1}^l, d\check{y}_n^l) = M_n^{(\theta,l)}(y_{n-1}^l, A) \quad \textit{and}$$
$$\int_{\mathbb{R}^d\times A} M_n^{(\theta,l,l-1)}(\check{y}_{n-1}^l, d\check{y}_n^l) = M_n^{(\theta,l-1)}(y_{n-1}^{l-1}, A).$$

In addition, for $(y, y') \in \mathbb{R}^d \times \mathbb{R}^d$ and $\varphi \in \mathcal{B}_b(\mathbb{R}^d \times \mathbb{R}^d)$, we introduce the following notation

$$M_n^{(\theta,l,l-1)}(\varphi)(y,y') := \int_{\mathbb{R}^{2d}} \varphi(\check{y}_n^l) M_n^{(\theta,l,l-1)}((y,y'), d\check{y}_n^l).$$

Algorithm 4 : Coupled increments of the Lévy process on $[0, 1]$

```
1. Input: $l$, $\delta_{l-1}$, $\lambda_l$, $\mu^l(dx)$ and $\theta$.
2. Run Algorithm 1 with input $l$, $\lambda_l$, $\mu^l(dx)$ and $\theta$ to return $K^l$, $\{T_i^l\}_{i=0}^{K^l}$ and $\{\Delta L_{T_i^l}^l\}_{i=1}^{K^l}$. Set
   $\widehat{K}^{l-1} = 0$.
3. Generate jump times and heights at level $l-1$: For $i = 1, \cdots, K^l$, if $|\Delta L_{T_i^l}^l| \geq \delta_{l-1}$, set
   $\widehat{T}_i^{l-1} = T_i^l$, $\Delta L_{\widehat{T}_i^{l-1}}^{l-1} = \Delta L_{T_i^l}^l$ and $\widehat{K}^{l-1} = \widehat{K}^{l-1} + 1$.
4. Refine jump times: Set $i = 1$, and $T_0^{l-1} = 0$.
     (i) If $\left\{\widehat{T}_j^{l-1} > T_{i-1}^{l-1} \ : \ j \in \{1, \cdots, \widehat{K}^{l-1}\}\right\} \neq \phi$, then
         $T_i^{l-1} = \min\left\{T_{i-1}^l + \Delta_{l-1},\ \min\left\{\widehat{T}_j^{l-1} > T_{i-1}^{l-1} \ : \ j \in \{1, \cdots, \widehat{K}^{l-1}\}\right\}\right\}$; otherwise
         $T_i^{l-1} = \min\left\{T_{i-1}^l + \Delta_{l-1},\ 1\right\}$.
    (ii) If $T_i^{l-1} = \widehat{T}_j^{l-1}$ for some $j \in \{1, \cdots, \widehat{K}^{l-1}\}$, set $\Delta L_{T_i^{l-1}} = \Delta L_{\widehat{T}_j^{l-1}}$; otherwise
         $\Delta L_{T_i^{l-1}} = 0$.
   (iii) if $T_i^{l-1} = 1$, set $K^{l-1} = i$; otherwise $i = i + 1$ and go to (i).
5. Output: Return $K^l$, $K^{l-1}$, $\{T_i^l\}_{i=1}^{K^l}$, $\{T_i^{l-1}\}_{i=1}^{K^{l-1}}$, $\{\Delta L_{T_i^l}^l\}_{i=1}^{K^l}$ and $\{\Delta L_{T_i^{l-1}}^{l-1}\}_{i=1}^{K^{l-1}}$.
```

Algorithm 5 produces samples from the coupled-kernel $M_n^{(\theta,l,l-1)}$ associated with the discretization in (6). By [10, Theorem 2], there exists a $0 < C < \infty$ such that for any $(y, \theta, l) \in \mathbb{R}^d \times \Theta \times \mathbb{N}$

$$\int_{\mathbb{R}^{2d}} |y_n^l - y_n^{l-1}|^2 M_n^{(\theta,l,l-1)}((y,y), d\check{y}_n^l) \leq C\Delta_l^{\beta}, \tag{13}$$

where β is the strong error rate of the Euler scheme presented in Subsection 3.2 that can be obtained from the bounds in (7). Moreover, in [10, Theorem 1] the author show that if $\int(\frac{|x|^2}{\delta_l} \wedge 1)\ \nu(dx) \leq g(\delta_l)$ for some decreasing and invertible function $g : (0, \infty) \to (0, \infty)$ and all $\delta_l > 0$, then the MSE $:= \mathbb{E}[(\zeta_1^{L,MLMC}(\varphi) - \zeta_1(\varphi))^2]$ satisfies the following bounds:

(i) If $g(\delta_l) \leq 1/(\delta_l \log(1/\delta_l)^{1+\gamma})$ as $\delta_l \to 0$ for some $\gamma > 0$, and $\Sigma = 0$, then the Cost $\leq C$, for some $C > 0$ and

$$\mathrm{MSE} \leq 1/C.$$

(ii) If $g(\delta_l) \leq \log(1/\delta_l)^{\gamma}/\delta_l$ as $\delta_l \to 0$ for some $\gamma \geq 1/2$, then the Cost $\leq C$, for some $C > 0$ and

$$\mathrm{MSE} \leq \frac{1}{C} \log(C)^{2(1+\gamma)}.$$

(iii) If there exists some $\gamma > 0$ such that $g(\delta_l) \leq \frac{1}{2} g(\gamma\delta_l/2)$ for all sufficiently small $\delta_l > 0$, then the Cost $\leq C$, for some $C > 0$ and

$$\mathrm{MSE} \leq C[g^{-1}(C)]^2.$$

```
Algorithm 5 : Coupled sampling for Lévy-driven SDEs
1. Input: $Y_0^l$, $Y_0^{l-1}$, $l$, $\lambda_l$, $F_0^l$, $\delta_{l-1}$, $F_0^{l-1}$, $\mu^l(dx)$ and $\theta$.
2. Call Algorithm 4 to return $K^l$, $K^{l-1}$, $\{T_i^l\}_{i=1}^{K^l}$, $\{T_i^{l-1}\}_{i=1}^{K^{l-1}}$, $\{\Delta L_{T_i^l}^l\}_{i=1}^{K^l}$ and $\{\Delta L_{T_i^{l-1}}^{l-1}\}_{i=1}^{K^{l-1}}$.
3. For $i = 1, \cdots, K^l$, sample $W_{T_i^l}$ from $\mathcal{N}_r(0, T_i^l I_r)$. Notice that $\{W_{T_i^{l-1}}\}_{i=1}^{K^{l-1}} \subset \{W_{T_i^l}\}_{i=1}^{K^l}$.
4. With $\{T_i^l\}_{i=1}^{K^l}$, $\{\Delta L_{T_i^l}^l\}_{i=1}^{K^l}$, $\{W_{T_i^l} - W_{T_{i-1}^l}\}_{i=1}^{K^l}$ and $F_0^l$, generate the increments $\{\Delta X_{T_i^l}^l\}_{i=1}^{K^l}$
   through (5). With $\{T_i^{l-1}\}_{i=1}^{K^{l-1}}$, $\{\Delta L_{T_i^{l-1}}^{l-1}\}_{i=1}^{K^{l-1}}$, $\{W_{T_i^{l-1}} - W_{T_{i-1}^{l-1}}\}_{i=1}^{K^{l-1}}$ and $F_0^{l-1}$, generate the
   increments $\{\Delta X_{T_i^{l-1}}^{l-1}\}_{i=1}^{K^{l-1}}$ via (5).
5. Given the coupled increments $\{\Delta X_{T_i^l}^l\}_{i=1}^{K^l}$ and $\{\Delta X_{T_i^{l-1}}^{l-1}\}_{i=1}^{K^{l-1}}$, compute $Y_1^l$ and $Y_1^{l-1}$ with $Y_0^l$
   and $Y_0^{l-1}$, respectively, via the recursion in (6).
6. Output: Return $Y_1^l$ and $Y_1^{l-1}$.
```

4.3 Coupled PF and Unbiased Level Difference Estimation

As assumed previously, let $\varphi \in \mathcal{B}_b(\mathbb{R}^d)$ and $\theta \in \Theta$ be fixed. Consider the level difference in (12). If one can estimate the difference using a coupling of $(\gamma_n^l, \gamma_n^{l-1})$, it is then possible to obtain a variance reduction. As seen in the previous subsection this is a crucial requirement in the ML estimation, where one wants the variance of the differences to decay as l increases. The authors of [21] introduced an effective general coupling of PFs which we use to unbiasedly estimate the level difference in (12). Coupled particle filter (CPF) is usually used in multilevel PF (MLPF) algorithms (see [20] and [23] for more details on MLPF for models driven by Wiener and Lévy processes, respectively) or debiasing schemes (see e.g. [8, 33]), similar to MLMC, one writes the expectation of some function w.r.t. smoothing distribution as a telescoping sum, then independently run a particle filter at the first level and CPFs for the higher levels. In Algorithm 6 we present the CPF used in [20, 23] which will generate samples from the coupling of $(\gamma_n^l, \gamma_n^{l-1})$.

Definition 4.2. *Let $l \in \mathbb{N}$, $\theta \in \Theta$ and $\check{y}_n^l = (y_n^l, y_n^{l-1})$. Define the potential*

$$\check{G}_n^{(\theta)}(\check{y}^l) := \frac{1}{2}(G_n^{(\theta)}(y_n^l) + G_n^{(\theta)}(y_n^{l-1})).$$

Algorithm 6 : Coupled particle filter

```
1. Input: Level $l$, number of particles $N_l$, $Y_0$, time $n$, $\theta$, $\lambda_l$, $F_0^l$, $\delta_{l-1}$, $F_0^{l-1}$, $\mu^l(dx)$ and potential functions $\{\check{G}_k^{(\theta)}\}_{k=1}^n$.
2. Initialization: For $i = 1, \cdots, N$ set $\boldsymbol{Y_0^{l,i}} = \boldsymbol{Y_0^{l-1,i}} = Y_0^{l,i} = Y_0^{l-1,i} = Y_0$.
3. For $k = 1, \cdots, n$, do:
    (i) For $i = 1, \cdots, N_l$ call Algorithm 5 with $Y_0^{l,i} = Y_{k-1}^{l,i}$ and $Y_0^{l-1,i} = Y_{k-1}^{l-1,i}$ in the input to return $Y_1^{l,i}$ and $Y_1^{l-1,i}$ and set $\widehat{\boldsymbol{Y}}_{\boldsymbol{k}}^{\boldsymbol{l,i}} = (\boldsymbol{Y_{k-1}^{l,i}}, Y_1^{l,i})$ and $\widehat{\boldsymbol{Y}}_{\boldsymbol{k}}^{\boldsymbol{l-1,i}} = (\boldsymbol{Y_{k-1}^{l-1,i}}, Y_1^{l-1,i})$.
    (ii) For $i = 1, \cdots, N_l$ compute $w_k^i = \check{G}_k^{(\theta)}(\check{y}_k^{l,i})$ and set $W_k^i = w_k^i / \sum_{j=1}^{N_l} w_k^j$.
    (iii) Resample the particles $\{\widehat{\boldsymbol{Y}}_{\boldsymbol{k}}^{\boldsymbol{l,i}}\}_{i=1}^N$ and $\widehat{\boldsymbol{Y}}_{\boldsymbol{k}}^{\boldsymbol{l-1,i}}$ according to the normalized weights $\{W_k^i\}_{i=1}^{N_l}$ by sampling the particles indices from a multinomial distribution. Denote the resampled particles $\{\boldsymbol{Y_k^{l,i}}\}_{i=1}^{N_l} = \{(Y_0^{l,i}, \cdots, Y_k^{l,i})\}_{i=1}^{N_l}$ and $\{\boldsymbol{Y_k^{l-1,i}}\}_{i=1}^N = \{(Y_0^{l-1,i}, \cdots, Y_k^{l-1,i})\}_{i=1}^N$.
4. For $i = 1, \cdots, N_l$ set $\check{V}^i = \left[\prod_{k=1}^{n-1} \frac{1}{N_l} \sum_{j=1}^{N_l} w_k^j\right] \frac{1}{N_l} w_n^i$.
5. Output: Return $\{\check{\boldsymbol{Y}}_{\boldsymbol{n}}^{\boldsymbol{i}}\}_{i=1}^{N_l} = \{(\boldsymbol{Y_n^{l,i}}, \boldsymbol{Y_n^{l-1,i}})\}_{i=1}^{N_l}$ and $\{\check{V}^i\}_{i=1}^{N_l}$.
```

We then use the output of Algorithm 6 in Algorithm 7 to unbiasedly estimate the level differences in (12).

Algorithm 7 : Unbiased estimation of level differences

```
1. Input: Level $l$, number of particles $N_l$, $Y_0$, time $n$, $\theta$, $\lambda_l$, $F_0^l$, $\delta_{l-1}$, $F_0^{l-1}$, $\mu^l(dx)$ and potential functions $\{G_k^{(\theta)}, \check{G}_k^{(\theta)}\}_{k=1}^n$.
2. Call Algorithm 6 to return $\{\check{\boldsymbol{Y}}_{\boldsymbol{n}}^{\boldsymbol{i}}\}_{i=1}^{N_l} = \{(\boldsymbol{Y_n^{l,i}}, \boldsymbol{Y_n^{l-1,i}})\}_{i=1}^{N_l}$ and $\{\check{V}^i\}_{i=1}^{N_l}$. Then perform the following
    (i) For $i = 1, \cdots, N_l$, set $\boldsymbol{Y_n^i} = \boldsymbol{Y_n^{l,i}}$ and $V^i = \check{V}^i H^{(\theta,l)}(\check{y}_n^{l,i})$, where
        $$H^{(\theta,l)}(\check{y}_n^{l,i}) = \prod_{k=1}^n \frac{G_k^{(\theta)}(y_k^{l,i})}{\check{G}_k^{(\theta)}(\check{y}_k^{l,i})}. \tag{14}$$
    (ii) For $i = N_l + 1, \cdots, 2N_l$, set $\boldsymbol{Y_n^i} = \boldsymbol{Y_n^{l-1,i-N_l}}$ and $V^i = -\check{V}^{i-N_l} H^{(\theta,l-1)}(\check{y}_n^{l,i-N_l})$, where
        $$H^{(\theta,l-1)}(\check{y}_n^{l,i-N_l}) = \prod_{k=1}^n \frac{G_k^{(\theta)}(y_k^{l-1,i-N_l})}{\check{G}_k^{(\theta)}(\check{y}_k^{l,i-N_l})}. \tag{15}$$
```

3. **Output:** Return $\{\mathbf{Y}_n^{l,i}, V^{l,i}\}_{i=1}^{2N_l}$.

Since the PF method provides an unbiased estimator of the unnormalized smoothing density, one can easily use that to show (see Proposition 2.4 in [8])

$$\mathbb{E}\left[\sum_{i=1}^{2N_l} V^i \, \varphi(\mathbf{Y}_n^{l,i})\right] = \gamma_n^{(\theta,l)}(\varphi) - \gamma_n^{(\theta,l-1)}(\varphi).$$

Remark 4.1. *(i) The choice of $\breve{G}_n^{(\theta)}$ in this paper is similar to that in [8]. However, in the original work [21] where this methodology was employed, the authors take*

$$\breve{G}_n^{(\theta)}(\breve{y}_n^l) = \max\{G_n^{(\theta)}(y_n^l), G_n^{(\theta)}(y_n^{l-1})\}.$$

In fact, it is sufficient for unbiasedness to choose $\breve{G}_n^{(\theta)}$ such that $H^{(\theta,l)}$ and $H^{(\theta,l-1)}$ are uniformly upper bounded; this way the variance of the weights is independent of the length of the observations n, hence, does not grow as $n \to \infty$.

(ii) The unbiasedness of level-difference is achieved regardless of the number of particles.

(iii) We remark that one can combine Algorithms 3 and 7 along with the randomization techniques in [27, 30] to provide an unbiased estimator of $\gamma_n^{(\theta)}(\varphi)$, with $\gamma_n^{(\theta)}$ as in (8). This leads to unbiased inference w.r.t. the normalized smoother $\eta_n^{(\theta)}(\varphi)$ (see [8, Proposition 2.8].)

We state the following Lemma which some of our claims we will rely on.

Lemma 4.1. *Under Assumptions 2.1 and 3.1, there exists a $C < \infty$ such that for any $l \in \mathbb{N}$ and $\theta \in \Theta$,*

$$\sup_{\varphi\in\mathcal{A}} \sup_{y\in\mathbb{R}^d} \left| M_n^{(\theta,l)}(\varphi)(y) - M_n^{(\theta,l-1)}(\varphi)(y)\right| \leq C\Delta_l^{\beta/2}$$

where $\mathcal{A} = \mathcal{B}_b(\mathbb{R}^d) \cap Lip_{\|\cdot\|_2}(\mathbb{R}^d)$ and β is the strong error rate of the discretization scheme for the SDE in (1).

Proof.

$$\begin{aligned}
&\left|M_n^{(\theta,l)}(\varphi)(y) - M_n^{(\theta,l-1)}(\varphi)(y)\right| \\
&= \left|M_n^{(\theta,l,l-1)}(\varphi\otimes 1)(y,y) - M_n^{(\theta,l,l-1)}(1\otimes\varphi)(y,y)\right| \quad \text{by Definition 4.1} \\
&= \left|\int_{\mathbb{R}^{2d}} \varphi(y_n^l) M_n^{(\theta,l,l-1)}((y,y), d\breve{y}_n^l) - \int_{\mathbb{R}^{2d}} \varphi(y_n^{l-1}) M_n^{(\theta,l,l-1)}((y,y), d\breve{y}_n^l)\right| \\
&= \left|\int_{\mathbb{R}^{2d}} (\varphi(y_n^l) - \varphi(y_n^{l-1})) M_n^{(\theta,l,l-1)}((y,y), d\breve{y}_n^l)\right| \\
&\leq \left(\int_{\mathbb{R}^{2d}} (\varphi(y_n^l) - \varphi(y_n^{l-1}))^2 M_n^{(\theta,l,l-1)}((y,y), d\breve{y}_n^l)\right)^{1/2} \quad \text{by Jensen's inequality} \\
&\leq \left(\int_{\mathbb{R}^{2d}} |y_n^l - y_n^{l-1}|^2 M_n^{(\theta,l,l-1)}((y,y), d\breve{y}_n^l)\right)^{1/2} \quad \text{since } \varphi \in \mathrm{Lip}_{\|\cdot\|_2}(\mathbb{R}^d).
\end{aligned}$$

Applying the result in (13) in the above inequality concludes the proof. □

Proposition 4.1. *Under Assumption 4.1 (i)-(ii), for any $(l, \theta) \in \Theta \times \mathbb{N}$, $n \geq 0$ and $\varphi \in Lip_{\|\cdot\|_2}(\mathbb{R}^{nd}) \cap \mathcal{B}_b(\mathbb{R}^{nd})$ there exists a $C < \infty$ such that*

$$\left|\gamma_n^{(\theta,l)}(\varphi) - \gamma_n^{(\theta,l-1)}(\varphi)\right| \leq C\Delta_l^{\beta/2} \tag{16}$$

Proof. Since Assumption 4.1 (i)-(ii) holds for all $\theta \in \Theta$, the bound in (16) is uniform in θ, therefore one can adopt the proof in [8]. □

Theorem 4.1. *Under Assumption 4.1, for any $(l, \theta) \in \Theta \times \mathbb{N}$ and $\varphi \in \mathcal{B}_b(\mathbb{R}^{nd}) \cap Lip(\mathbb{R}^{nd})$, there exists a $C < \infty$ such that*

$$\mathbb{E}\left[\left(\sum_{i=1}^{2N_l} V^i\,\varphi(\mathbf{Y}_n^{l,i}) - \left\{\gamma_n^{(\theta,l)}(\varphi) - \gamma_n^{(\theta,l-1)}(\varphi)\right\}\right)^2\right] \leq C\frac{\Delta_l^{2\wedge\beta}}{N_l}, \tag{17}$$

where $\{\mathbf{Y}_n^{l,i},\, V^{l,i}\}_{i=1}^{2N_l}$ is the output of Algorithm 7.

Proof. The proof is essentially the same as in [8], except that one needs to use Lemma 4.1 and inequality (13) along the lines. □

Note that

$$\begin{aligned}
&\mathbb{E}\left[\left(\sum_{i=1}^{2N_l} V^i\,\varphi(\mathbf{Y}_n^{l,i})\right)^2\right] \\
&= \mathbb{E}\left[\left(\sum_{i=1}^{2N_l} V^i\,\varphi(\mathbf{Y}_n^{l,i}) - \left\{\gamma_n^{(\theta,l)}(\varphi) - \gamma_n^{(\theta,l-1)}(\varphi)\right\} + \left\{\gamma_n^{(\theta,l)}(\varphi) - \gamma_n^{(\theta,l-1)}(\varphi)\right\}\right)^2\right] \\
&\leq \mathbb{E}\left[\left(\sum_{i=1}^{2N_l} V^i\,\varphi(\mathbf{Y}_n^{l,i}) - \left\{\gamma_n^{(\theta,l)}(\varphi) - \gamma_n^{(\theta,l-1)}(\varphi)\right\}\right)^2\right] + \mathbb{E}\left[\left|\gamma_n^{(\theta,l)}(\varphi) - \gamma_n^{(\theta,l-1)}(\varphi)\right|^2\right]
\end{aligned}$$

by Theorem 4.1 and Proposition 4.1, we have

$$\mathbb{E}\left[\left(\sum_{i=1}^{2N_l} V^i\,\varphi(\mathbf{Y}_n^{l,i})\right)^2\right] \leq C\left(\frac{\Delta_l^{2\wedge\beta}}{N_l} + \Delta_l^{\beta}\right). \tag{18}$$

5 Overall Methodology

In this section, we provide the main algorithm of this paper, Algorithm 8. As previously discussed, the goal is to perform an unbiased inference for the Bayesian model posterior associated to the HMM model in (1)-(2), which is given by

$$\Pi(d\theta, dy_{1:n}) \propto \overline{\Pi}(d\theta)\prod_{p=1}^{n} G_p^{(\theta)}(y_p)M_p^{(\theta,\infty)}(y_{p-1}, dy_p), \tag{19}$$

where θ is the static parameter to infer with a prior measure $\overline{\Pi}(d\theta)$ and $y_{1:n}$ are the hidden states. We assume the transition densities $M_n^{(\theta,\infty)}$ cannot be simulated directly, but there are discretized densities $\{M_n^{(\theta,l)}\}_{l\in\mathbb{N}}$ that approximate $M_n^{(\theta,\infty)}$ which one can simulate from. Therefore, we work with the following posterior instead

$$\Pi^{(l)}(d\theta, dy_{1:n}) \propto \overline{\Pi}(d\theta) \prod_{p=1}^{n} G_p^{(\theta)}(y_p) M_p^{(\theta,l)}(y_{p-1}, dy_p). \tag{20}$$

Given some function φ that is bounded and measurable on $\Theta \times \mathbb{R}^{nd}$, we are interested in unbiasedly estimating the expectation $\Pi(\varphi)$. Algorithm 8 provides the methodology to do that. The algorithm consists of two main parts; the first is a direct implementation of PMMH [1] at the coarsest level $l_{\min}$, e.g. $l_{\min} = 1$, whilst the second part is just running the unbiased level difference scheme in Algorithm 7, conditionally upon the results from the first part at random levels l sampled from a given PMF $P_l = \{p_l\}_{l\in\{l_{\min}+1,\cdots\}}$. The first part of the algorithm does not yield unbiased inference, however the second part does [8].

Algorithm 8 : Unbiased estimator of $\Pi(\varphi)$

1. **Input:** A function φ defined on $\Theta \times \mathbb{R}^{nd}$, the number of particles $\{N_l\}_{l\geq l_{\min}}$, initial value Y_0, time n, $\mu^l(dx)$, λ_l, F_0^l, δ_{l-1}, F_0^{l-1}, $\epsilon > 0$, potential functions $\{G_k^{(\theta)}, \breve{G}_k^{(\theta)}\}_{k=1}^n$, a prior $\overline{\Pi}(d\theta)$, a proposal density $q(\theta,\theta')$ for the MH update, a number of iterations S, a PMF $P_l = \{p_l\}_{l\in\{l_{\min}+1,\cdots\}}$, and initial values $(\theta_0, \{\mathbf{Y}_{n,0}^i, V_0^i\}_{i=1}^N)$ such that $\sum_{i=1}^N V_0^i > 0$.

2. **Part 1**:

   Initialize: Set $\widehat{K} = 0$ and $D_{\widehat{K}} = 0$.

   For $k = 1, \cdots, S$, do:

   (i) Sample $\widetilde{\theta}_k$ from $q(\theta_{k-1}, \cdot)$.

   (ii) Run Algorithm 3 with input $l = l_{\min}$, N, Y_0, n, $\widetilde{\theta}_k$ and potentials $\{G_p^{(\widetilde{\theta}_k)}\}_{p=1}^n$ to return the output $\{\widetilde{\mathbf{Y}}_{n,k}^{l,i}, \widetilde{V}_k^i\}_{i=1}^N$.

   (iii) Set

   $$\alpha = \min\left\{1, \frac{q(\widetilde{\theta}_k, \theta_{k-1})\ \overline{\Pi}(\widetilde{\theta}_k)\ (\sum_{i=1}^N \widetilde{V}_k^i + \epsilon)}{q(\theta_{k-1}, \widetilde{\theta}_k)\ \overline{\Pi}(\theta_{k-1})\ (\sum_{i=1}^N V_{k-1}^i + \epsilon)}\right\}.$$

   Then accept with probability α and set $(\theta_k, \{\mathbf{Y}_{n,k}^i, V_k^i\}_{i=1}^N) = (\widetilde{\theta}_k, \{\widetilde{\mathbf{Y}}_{n,k}^i, \widetilde{V}_k^i\}_{i=1}^N)$, $\widehat{K} \leftarrow \widehat{K} + 1$, $D_{\widehat{K}} = 1$ and $(\widehat{\theta}_{\widehat{K}}, \{\widehat{\mathbf{Y}}_{n,\widehat{K}}^i, \widehat{V}_{\widehat{K}}^i\}_{i=1}^N) = (\theta_k, \{\mathbf{Y}_{n,k}^i, V_k^i\}_{i=1}^N)$. Otherwise set $(\theta_k, \{\mathbf{Y}_{n,k}^i, V_k^i\}_{i=1}^N) = (\theta_{k-1}, \{\mathbf{Y}_{n,k-1}^i, V_{k-1}^i\}_{i=1}^N)$ and $D_{\widehat{K}} \leftarrow D_{\widehat{K}} + 1$.

3. **Part 2**: Given $(\widehat{\theta}_k, \{\widehat{\mathbf{Y}}_{n,k}^i, \widehat{V}_k^i\}_{i=1}^N)_{k=1}^{\widehat{K}}$ and $\{D_k\}_{k=1}^{\widehat{K}}$ from Part 1. For $k = 1, \cdots, \widehat{K}$, do:

   (i) For $i = 1, \cdots, N$, set $\mathbf{Y}_{n,k}^{0,i} = \widehat{\mathbf{Y}}_{n,k}^i$ and $W_{k,0}^i = D_k\ \widehat{V}_k^i / (\sum_{i=1}^N \widehat{V}_k^i + \epsilon)$.

   (ii) Sample l_k from P_l.

   (iii) Run Algorithm 7 with input l_k, N, Y_0, n, $\widehat{\theta}_k$, λ_{l_k}, $F_0^{l_k}$, δ_{l_k-1}, $F_0^{l_k-1}$, $\mu^{l_k}(dx)$ and potential functions $\{G_p^{(\widehat{\theta}_k)}, \breve{G}_p^{(\widehat{\theta}_k)}\}_{p=1}^n$, to return $\{\mathbf{Y}_{n,k}^{l_k,i}, \overline{V}_k^{l_k,i}\}_{i=1}^{2N_l}$. For $i = 1, \cdots, N$, set

   $$W_{k,l_k}^i = \frac{\overline{V}_k^{l_k,i}}{p_{l_k}(\sum_{i=1}^N \widehat{V}_k^i + \epsilon)}.$$

4. **Output:** Return

$$\Pi_{ub}(\varphi) := \frac{\sum_{k=1}^{\widehat{K}} \left\{ \sum_{i=1}^{N} W_{k,0}^{i} \, \varphi(\widehat{\theta}_k, \mathbf{Y}_{n,k}^{0,i}) + \sum_{i=1}^{2N} W_{k,l_k}^{i} \, \varphi(\widehat{\theta}_k, \mathbf{Y}_{n,k}^{l_k,i}) \right\}}{\sum_{k=1}^{\widehat{K}} \left\{ \sum_{i=1}^{N} W_{k,0}^{i} + \sum_{i=1}^{2N} W_{k,l_k}^{i} \right\}}. \tag{21}$$

Remark 5.1. *(i) In Algorithm 8, it is important to note that PMMH is only implemented for the coarsest level, which makes it relatively cheap. In addition, the IS corrections in Part 2 may be calculated in parallel since Algorithm 7 is run independently for each $\widehat{\theta}_k$.*

(ii) As suggested in [8], based on (18) one may take $N_{l_{min}} = \overline{N}$, $N_l = \widetilde{N}$, with $\widetilde{N} < \overline{N}$, and $p_l \propto 2^{-3l/2}$ for $l > l_{min}$.

(iii) Note that the estimator defined in (21) is equivalent to

$$\widehat{\Pi_{ub}(\varphi)} := \frac{\sum_{k=1}^{S} \left\{ \sum_{i=1}^{N} W_{k,0}^{i} \, \varphi(\theta_k, \mathbf{Y}_{n,k}^{0,i}) + \sum_{i=1}^{2N} W_{k,l_k}^{i} \, \varphi(\theta_k, \mathbf{Y}_{n,k}^{l_k,i}) \right\}}{\sum_{k=1}^{S} \left\{ \sum_{i=1}^{N} W_{k,0}^{i} + \sum_{i=1}^{2N} W_{k,l_k}^{i} \right\}},$$

where $(\{W_{k,0}^i, \mathbf{Y}_{n,k}^{0,i}\}_{i=1}^N)_{k=1}^S$ and $(\{W_{k,l_k}^i, \mathbf{Y}_{n,k}^{l_k,i}\}_{i=1}^{2N})_{k=1}^S$ are the output of running Part 2 in Algorithm 8 conditional upon $(\theta_k, \{\mathbf{Y}_{n,k}^i, V_k^i\}_{i=1}^N)_{k=1}^S$ from Part 1.

6 Numerical Simulations

Consider the model in (1)-(2), where $\{X_t\}_{t\in[0,T_f]}$ is a 1-dimensional Lévy process characterized by the triplet (ν, Σ, b). We consider a symmetric truncated Lévy process, that is, $\Sigma = b = 0$ which implies that the process considered here has no drift and Brownian motion components. We define the Lévy measure ν on $\mathbb{R} \setminus \{0\}$ as

$$\nu(dx) = \left(\mathbb{I}_{[-u,0)}(x) \frac{c}{(-x)^{1+\alpha}} + \mathbb{I}_{(0,u]}(x) \frac{c}{x^{1+\alpha}} \right) dx,$$

where $c > 0$, $\alpha \in (0, 2)$ and $u \geq 1$ is the truncation threshold. This is the same Lévy measure used in [10, 23]. Then the measure μ^l, $l \in \mathbb{N}$, is given by

$$\mu^l(dx) = \frac{c}{\lambda_l} \left[\mathbb{I}_{[-u,-\delta_l]}(x) \frac{1}{(-x)^{1+\alpha}} + \mathbb{I}_{[\delta_l,u]}(x) \frac{1}{x^{1+\alpha}} \right] dx,$$

where λ_l can be computed analytically as

$$\lambda_l = \nu(B_{\delta_l}^c) = 2c \int_{\delta_l}^{u} x^{-1-\alpha} \, dx = \frac{2c}{\alpha} \left(\frac{1}{\delta_l^\alpha} - \frac{1}{u^\alpha} \right).$$

Recall that δ_l is chosen such that $\lambda_l = \Delta_l^{-1}$, therefore, we have

$$\delta_l = \left(\frac{\alpha}{2c\Delta_l} + \frac{1}{u^\alpha} \right)^{-1/\alpha}.$$

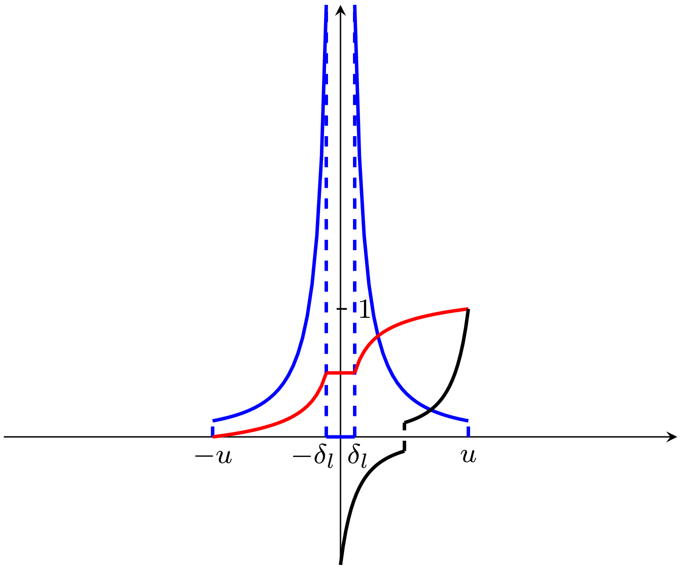

Figure 1: The blue, red and black curves correspond to the density of μ^l, the CDF and the inverse of the CDF, respectively.

Notice that $\delta_l < u$, so the definition of μ^l makes sense. We take $u = 1$, $c = 0.8$ and $\alpha = 0.5$. Figure 1 shows the graphs of the density associated with μ^l for $l = 2$, the cumulative distribution function (CDF) and its inverse. Note that due to the symmetric structure of ν, we have $F_0^l = 0$. Since the CDF inverse exists and can be computed analytically, we use the inverse transform sampling method to sample the heights of the Lévy process jumps from μ^l in Step 4. of Algorithm 1. One can compute the strong error rate of the Euler scheme associated with the measure ν. The term $\sigma^2_{\delta_l}$ in (7) is computed analytically as

$$\sigma^2_{\delta_l} = \int_{B_{\delta_l}} |x|^2 \nu(dx) = 2c \int_0^{\delta_l} x^{1-\alpha} dx = \frac{2c}{2-\alpha} \delta_l^{2-\alpha} = \mathcal{O}(\delta_l^{2-\alpha}),$$

hence by (7),

$$\mathbb{E}\Big[\sup_{t\in[0,1]} \big|Y_t - Y_t^l\big|^2 \Big] \leq C\delta_l^{2-\alpha} = C\left(\frac{\alpha}{2c\Delta_l} + 1\right)^{-(2-\alpha)/\alpha} = \mathcal{O}(\Delta_l^{(2-\alpha)/\alpha}).$$

Since $\alpha = 0.5$, we have the strong error rate $\beta = 3$.

The Lévy-driven SDE considered here has the form

$$dY_t = \theta\; Y_{t^-}\; dX_t, \quad Y_0 = y_0,$$

with $y_0 = 1$. The observations set is the daily S & P 500 log-returns from Jan 3, 2012 to May 25, 2013 taken from Yahoo Finance website. In Figure 2, we plot the MSE for Algorithm 8 and PMMH against the computational cost measured in seconds. The MSE is computed by running 52 independent simulations of each method

$$\text{MSE}_{ub} = \frac{1}{52}\sum_{i=1}^{52}[\Pi_{ub}(\varphi)^i - \Pi(\varphi)]^2, \qquad \text{MSE}_{pmmh} = \frac{1}{52}\sum_{i=1}^{52}[\Pi_{pmmh}(\varphi)^i - \Pi(\varphi)]^2,$$

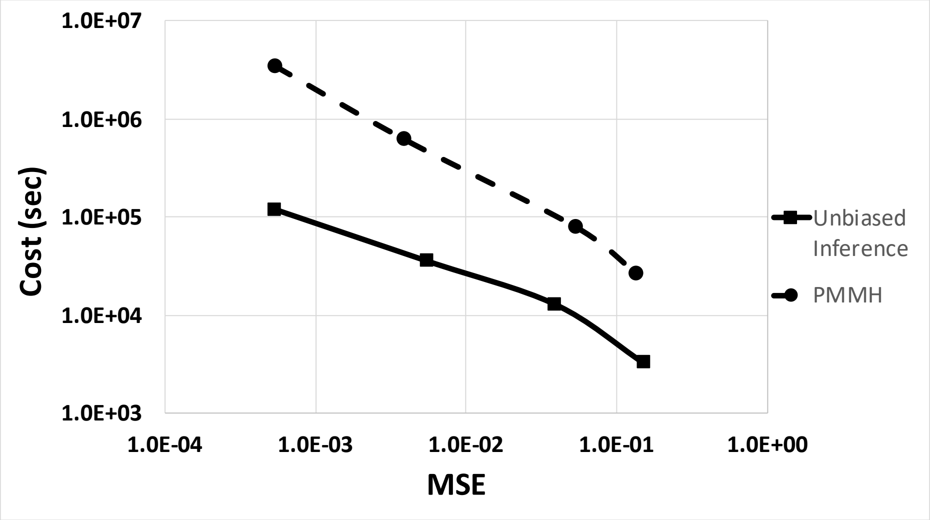

Figure 2: The computational cost (in seconds) versus MSE for both the unbiased inference method in Algorithm 8 and PMMH algorithm in [1].

where $\Pi(\varphi)$ is the ground truth. We explain below how the ground truth is estimated. In practice, one has to truncate the values of l in Part 2 of Algorithm 8, therefore, for each simulation $\Pi_{ub}(\varphi)^i$ in the MSE identity above we set $P_l = (p_l)_{l=l_{\min+1}}^{l_{\max}}$ with $l_{\min} = 1$, $l_{\max} = 12$ and $p_l = 2^{-3l/2}$. The reference $\Pi(\varphi)$ is estimated as the mean of 52 runs of Algorithm 8 with $l_{\max} = 14$, $N = 100$ and $S = 10^5$. We choose a positive algorithm constant $\epsilon = 10^{-8}$. The number of particles is at most 60 in both algorithms and PMMH has a discretization level of at most 8 (higher levels will result in a much larger computational cost.) We set the function $\varphi(\theta, x) = \theta$ in our simulations. In Table 1, we compare the expected value of θ obtained by both methods, the MSE and ratio of computational costs. The reference value is 0.76249.

θ_{ub}	MSE_{ub}	θ_{pmmh}	MSE_{pmmh}	$\text{Cost}_{pmmh}/\text{Cost}_{ub}$
0.78446	1.5054 E-01	0.78748	1.3591 E-01	7.98
0.76129	3.9087 E-02	0.78309	5.4409 E-02	6.02
0.76089	5.4825 E-03	0.78069	3.9315 E-03	17.1
0.76183	5.3014 E-04	0.77012	5.4522 E-04	28.5

Table 1: First and third columns are the expected values of θ obtained by Algorithm 8 and PMMH, respectively. The second and fourth columns are the corresponding MSE and the fifth column is the ratio of the computational costs.

Acknowldegements

HR & AJ were supported by KAUST baseline funding.

References

[1] Andrieu, C., Doucet, A., & Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **72**(3), 269–342 (with discussion).

[2] Applebaum, D. (2004). *Lévy Processes and Stochastic Calculus.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge.

[3] Barndorff-Nielsen, O. E., Mikosch, T. & Resnick, S. (Eds.) (2001). *Lévy Processes: Theory and Applications.* Birkhäuser.

[4] Barndorff-Nielsen, O. E. & Shephard, N. (2001). Non-Gaussian OU-based models and some of their uses in financial economics (with discussion). *J. Roy. Statist. Soc. B*, **63**(2), 167–241.

[5] Bertoin, J. (1996). *Lévy Processes.* Cambridge Tracts in Mathematics, **121**. Cambridge University Press, Cambridge.

[6] Blumenthal, R. M. & Getoor, R. K. (1961). Sample functions of stochastic processes with stationary independent increments. *J. Math. Mech.* **10**, 493–516.

[7] Cappé, O., Ryden, T. & Moulines, E. (2005). *Inference in Hidden Markov Models*, Springer, New York.

[8] Chada, N., Franks, J., Jasra, A., Law, K. J. H. & Vihola, M. (2021). Unbiased Inference for Discretely Observed Hidden Markov Model Diffusions. *SIAM/ASA JUQ*, **9**(2), 763–787.

[9] Del Moral, P. (2004). *Feyman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications.* Springer, New York.

[10] Dereich, S. & Heidenreich, F. (2011). A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations. *Stoc. Proc. Appl.*, **121**(7), 1565–1587.

[11] Douc, R., Cappé, O. & Moulines, E. (2005). Comparison of resampling schemes for particle filtering. *Proc. 4th Int. Symp. on Image and Signal Processing and Analysis*, pp 64–69.

[12] Doucet, A., De Freitas, N., Gordon, N. J. (2001). *Sequential Monte Carlo Methods in Practice.* Springer, New York.

[13] Ferreiro-Castilla, A., Kyprianou, A. E., Scheichl, R. & Suryanarayana, G. (2014). Multilevel Monte Carlo simulation for Lévy processes based on the Wiener-Hopf factorization. *Stoch. Proc. Appl.*, **124**(2), 985–1010.

[14] Gander, M. P. S. & Stephens, D. A. (2007). Simulation and inference for stochastic volatility models driven by Lévy processes. *Biometrika*, **94**(3), 627–646.

[15] Giles. M. B. (2008). Multilevel Monte Carlo Path Simulation. *Oper. Res.*, **56**(3), 607–617.

[16] Giles, M.B. & Xia, Y. (2012). Multilevel path simulation for jump-diffusion SDEs. *In Plaskota, L., Woźniakowski, H. (eds.) Monte Carlo and Quasi-Monte Carlo Methods 2010.* Springer, Berlin.

[17] Golightly, A., & Wilkinson, D. J. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Comp. Stat. Data Anal.*, **52**(3), 1674–1693.

[18] Griffin, J. & Steel, M. (2006). Inference with non-Gaussian Ornstein-Uhlenbeck processes for stochastic volatility. *J. Econom.*, **134**(2), 605–644.

[19] Jacod, J., Kurtz, T. G., Méléard, S. & Protter, P. (2005). The approximate Euler method for Lévy driven stochastic differential equations. *Ann. Inst. Henri PoincarÂ ´e Probab. Stat.*, **41**(3), 523–558.

[20] Jasra, A., Kamatani, K., Law, K. J. H. and Zhou, Y. (2017). Multilevel Particle Filters. *SIAM J. Numer. Anal.*, **55**(6), 3068–3096.

[21] Jasra, A., Kamatani, K., Law, K. J. H. and Zhou, Y. (2018). Bayesian static parameter estimation for partially observed diffusions via multilevel Monte Carlo. *SIAM J. Sci. Comp.*, **40**(2), A887–A902.

[22] Jasra, A., Kamatani, K. & Masuda H. (2019). Bayesian inference for stable Lévy driven stochastic differential equations with high-frequency data. *Scand. J. Stat.*, **46**(2), 545–574.

[23] Jasra, A., Law, K.J.H. & Osei, P.P. (2019). Multilevel particle filters for Lévy-driven stochastic differential equations. *Stat. Comput.*, **29**, 775–789.

[24] Jasra, A., Stephens, D. A., Doucet, A. & Tsagaris, T. (2011). Inference for Lévy driven stochastic volatility models via adaptive sequential Monte Carlo. *Scand. J. Statist.*, **38**(1), 1–22.

[25] Kyprianou, A. E. (2014). *Fluctuations of Lévy Processes with Applications*, Universitext, Springer, Berlin, Heidelberg, pp. 35–69.

[26] Kyprianou, A. E., Schoutens, W. & Wilmott, P. (Eds.) (2005). *Exotic Option Pricing and Advanced Lévy Models.* Wiley.

[27] McLeish, D. (2011). A general method for debiasing a Monte Carlo estimator. *Monte Carlo Methods Appl.*, **17**(4), 301–315.

[28] Protter, P. (2004). *Stochastic Integration and Differential Equations*. Stochastic Modeling and Applied Probability, **21**, 2nd edn. Springer, Berlin.

[29] Protter, P. & Talay, D. (1997). The Euler scheme for Lévy driven stochastic differential equations. *The Annals of Probability, Ann. Probab.*, **25**(1), 393–423.

[30] Rhee, C.-H., Glynn, P. W. (2015). Unbiased Estimation with Square Root Convergence for SDE Models. *Oper. Res.*, **63**(5), 1026–1043.

[31] Roberts, G. O. & Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, **88**(3), 603–621.

[32] Rubenthaler S. (2003). Numerical simulation of the solution of a stochastic differential equation driven by a Lévy process. *Stochastic Process. Appl.* **103**(2), 311–349.

[33] Ruzayqat, H. M. & Jasra, A. (2020). Unbiased estimation of the solution to Zakai's equation. *Monte Carlo Methods Appl.*, **26**(2), 113–129.

[34] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*. Cambridge studies in Advanced Mathematics, **68**. Cambridge University Press, Cambridge.

[35] Vihola, M., Helske, J. & Franks, J. (2020). Importance sampling type estimators based on approximate marginal MCMC. *Scand. J. Statist.*, **47**(4), 1339–1376.