Multilevel Monte Carlo for Jump Processes

Yuan Xia

St Hugh's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2013

To my parents.

Acknowledgements

I'm indebted to my supervisor Mike Giles for his invaluable guidance and consistent encouragement during my study. Without his deep insight and patience, this work couldn't have been done.

It is my honor to have Ben Hambly and Andreas Kyprianou as my examiners whose comments are greatly appreciated. I am grateful to Lajos Gergely Gyurkó, Christoph Reisinger for their helpful comments on my transfer and confirmation reports. Part of this work also benefits from valuable comments by Hanqing Jin, Jan Obłój and discussions with Sylvestre Burgos and Lukas Szpruch.

Fellow students and faculties at Mathematical Institute and Oxford-Man Institute, including Bahman Angoshtari, Samuel N. Cohen, Nathan Korda, Gechun Liang, Arnaud Lionnet, Hao Ni, Bertrand Nortier, Peter Spoida, Pedro Vitória, Kaiwei Wang, Weijun Xu, Zichen Zhang, thank you for making my time at Oxford memorable.

Thank my parents Jianguo Xia and Huiqin Wang and my beloved Yijia Liu for their continuing supports.

Finally, I would like to thank China Scholarship Council and Oxford-Man Institute for their financial support.

Abstract

This thesis consists of two parts. The first part (Chapters 2-4) considers multilevel Monte Carlo for option pricing in finite activity jump-diffusion models. We use a jump-adapted Milstein discretisation for constant rate cases and with the thinning method for bounded state-dependent rate cases. Multilevel Monte Carlo estimators are constructed for Asian, lookback, barrier and digital options. The computational efficiency is numerically demonstrated and analytically justified.

The second part (Chapter 5) deals with option pricing problems in exponential Lévy models where the increments of the underlying process can be directly simulated. We discuss several examples: Variance Gamma, Normal Inverse Gaussian and α-stable processes and present numerical experiments of multilevel Monte Carlo for Asian, lookback, barrier options, where the running maximum of the Lévy process involved in lookback and barrier payoffs is approximated using discretely monitored maximum. To analytically verify the computational complexity of multilevel method, we also prove some upper bounds on L^p convergence rate of discretely monitored error for a broad class of Lévy processes.

Contents

List of Figures

Chapter 1

Introduction

1.1 Financial modelling using jump processes

In the Black-Scholes Model [6], the price of an option is given by the expected value of a payoff depending upon an asset price modelled by a stochastic differential equation driven by Brownian motion,

$$\mathrm{d}S(t) = a(S(t), t)\,\mathrm{d}t + b(S(t), t)\,\mathrm{d}W(t), \quad 0 \le t \le T, \tag{1.1}$$

with given initial data S_0. Although this model is widely used, stylized empirical facts of asset returns, which are statistical properties shared by a wide range of instruments supports the idea that asset returns are not log-normal; see [13]. This has motivated people to suggest models which better capture the characteristics of the asset price dynamics.

Merton [49] instead proposed a jump-diffusion process, in which the asset price follows a jump-diffusion SDE:

$$\mathrm{d}S(t) = a(S(t-), t)\,\mathrm{d}t + b(S(t-), t)\,\mathrm{d}W(t) + c(S(t-), t)\,\mathrm{d}J(t), \quad 0 \le t \le T, \tag{1.2}$$

where the jump term $J(t)$ is a compound Poisson process $\sum_{i=1}^{N(t)}(Y_i - 1)$, the jump magnitude Y_i has a prescribed distribution, and $N(t)$ is a Poisson process with intensity λ, independent of the Brownian motion. Due to the existence of jumps, the process is a càdlàg process, i.e. having right continuity with left limits. We note that $S(t-)$ denotes the left limit of the process while $S(t) = \lim_{s \to t+} S(s)$. In [49], Merton also assumed that $\log Y_i$ has a normal distribution.

There are several ways to generalize the Merton model. Motivated by the feedback effect caused by herding behavior, which means investors' panic will rise as price falls, the intensity could be modelled to depend on the state value of the asset. [40] proposes a jump-diffusion LIBOR market model where the LIBOR rate is generated by a state-dependent jump-diffusion process. Unlike the jump-diffusion SDE in the Merton model, the state-dependent jump-diffusion SDE doesn't admit explicit solutions; thus numerical methods are necessary for the computation.

Apart from jump-diffusion models, another possible way is to build the model on processes purely consisting of jumps. According to [14], models building on pure jump processes can reproduce the stylized facts of asset returns, like heavy tails and the asymmetric distribution of increments. They also reflect generic sudden discontinuous price movement compared to models with stochastic volatility or nonlinear diffusion coefficient diffusion processes. Other evidence to support this kind of model are from option markets. According to chapter 1 of [14], models with jumps give an intuitive explanation of implied volatility skew and smile in the index option market and foreign exchange market - the jump fear is mainly of downside in the equity market which produces a premium for low-strike options; the jump risk is symmetric in the foreign exchange market so the implied volatility has a smile shape.

Since pure jump processes of finite activity without a diffusion component cannot generate realistic price dynamics, it is necessary to allow the frequency of jumps within a finite horizon to be infinite. In the literature ([54, 13]) a class of tractable models is constructed based on the assumption that asset returns follow a Lévy process. Lévy processes allow a large class of distributions of increments, including pure jump processes. They have good flexibility to choose the number of parameters so that models can provide a higher calibration accuracy.

We briefly discuss risk-neutral pricing in models with jump processes. According to chapter 9 of [13], by the fundamental theorem of asset pricing, assuming the market is arbitrage-free there exists a risk-neutral pricing measure $\mathbb{Q}$ such that discounted assets are martingales under $\mathbb{Q}$. In the Black-Scholes Model, $\mathbb{Q}$ is unique. However, the uniqueness of $\mathbb{Q}$ is generally unobtainable in models involving jumps. We have to choose a $\mathbb{Q}$ for pricing purpose. For example, in the Merton model, by assuming the jump risk has no premium, under $\mathbb{Q}$ the underlying SDE has a different drift compared to under physical measure $\mathbb{P}$. Hence the parameters with regard to the jump distribution could be estimated from historical price series. Denoting r the risk-free rate, then under $\mathbb{Q}$, the jump-diffusion SDE (1.2) would be

$$\mathrm{d}S(t) = (r\,S(t-) - c(S(t-),t)\lambda m)\,\mathrm{d}t + b(S(t-),t)\,\mathrm{d}W(t) + c(S(t-),t)\,\mathrm{d}J(t), \quad 0 \le t \le T, \tag{1.3}$$

where $m = \mathbb{E}[Y_i] - 1$ is the compensator to ensure the discounted asset price is a martingale.

The $\mathbb{Q}$ we choose determines what kind of parameters of the model are calibrated from simple option products. In the first part of the thesis we choose a $\mathbb{Q}$ based on the same assumption, i.e. no jump risk premium. We use reasonable hypothetical parameters for numerical experiments. In the second part of the thesis, since exponential Lévy models are built for calibration from pricing perspective, the calibration is always under a specified $\mathbb{Q}$. We use the parameters of the model calibrated under mean-correcting pricing measure $\mathbb{Q}$ in [54] for numerical computation. Since we focus on the computation of models, we don't discuss other possible choice of risk-neutral pricing measures $\mathbb{Q}$.

The computation of option pricing in models with jumps involves three types of methods: Fourier transform methods, numerical methods for PIDE (Partial Integral Differential Equation) and Monte Carlo. Since the characteristic functions of Lévy processes in the modelling literature can be expressed in terms of elementary functions, the Fourier transform of option prices may be obtained; see chapter 11 of [14]. By numerically evaluating the Fourier transform, which is quite efficient by using the FFT algorithm, calibration is very efficient using Fourier transform methods. Its limitation lies in the complexity to price path-dependent payoffs involving a running maximum. Fourier transform of quantities related to running maximum often requires the Wiener-Hopf factorization of the Lévy process. However, for the Lévy processes we are concerned with, Wiener-Hopf factors are not known in closed form.

In the case of diffusion models, risk-neutral pricing theory shows that the option price satisfies a pricing PDE. In a model with jumps, the pricing equation is a PIDE which causes theoretical and numerical challenges to solve compared to a diffusion PDE. People have managed to modify numerical methods for PDE to deal with these issues. A good summary in chapter 12 of [14] discusses advantages and drawbacks of different numerical methods for PIDE.

The common limitation of these method is that they can not cope with high dimensional problems. This is the main point in favour of Monte Carlo. Monte Carlo can also handle options with path-dependent payoffs. Therefore, the focus of this work is to develop multilevel Monte Carlo for jump processes, rather than comparing it with other numerical methods.

1.2 Monte Carlo simulation

The Monte Carlo method estimates the expectation $Y = \mathbb{E}[f(X)]$ using

$$\hat{Y} := \frac{1}{M}\sum_{i=1}^{M} f(X^{(i)}).$$

where $X^{(i)}$ are i.i.d. samples of X. Suppose $\mathbb{E}[|f(X)|] < +\infty$, the Law of Large Number ensures the estimator converges to the true value as the number of draws increases. Suppose $0 < Var[f(X)] < +\infty$, the Central Limit Theorem gives the confidence interval of the estimator. As M goes to infinity,

$$\frac{\hat{Y} - Y}{\sqrt{\frac{1}{M}Var[f(X)]}} \sim N(0,1),$$

which provides information about the magnitude of the sampling error after a finite number of samples are drawn. In the following we also use the notation $\mathbb{V}[X] \equiv Var[X]$.

In financial applications of option pricing, we are interested in the expected value of a functional of the underlying process S, i.e. $\mathbb{E}[f(S)]$, where f is the payoff of the option, and S follows some SDE. The payoff of the option could depend on the terminal value of S, or be path-dependent where the valuation depends on the entire path $S(t), 0 \leq t \leq T$: Asian options (based on the average value of the underlying), lookback options (based on the running minimum or maximum of the underlying), barrier options (in which the payoff is zero if the underlying crosses, or fails to cross, a certain level) and digital options (for which the payoff is a step function or other discontinuous function of $S(T)$).

When exact sampling of the underlying SDE solution is not available, or we are interested in path-dependent options, the expected value can be estimated by a simple Monte Carlo method with a suitable approximation to the SDE solution. To demonstrate the application of Monte Carlo to our problem, we take European options with a uniform Lipschitz payoff as an example.

We assume the solution of the SDE (1.1) doesn't admit exact sampling. A simple Euler discretisation of the diffusion SDE with timestep h is

$$\widehat{S}_{n+1} = \widehat{S}_n + a\left(\widehat{S}_n, nh\right) h + b\left(\widehat{S}_n, nh\right) \Delta W_n,$$

where $\widehat{S}(0) = S(0)$, $\widehat{S}_n = \widehat{S}(nh)$, $\Delta W_n = W\left((n+1)h\right) - W\left(nh\right).$

We are interested in European options (based on the value of the underlying at the terminal state $S(T)$) with a uniform Lipschitz f, i.e. there exists a constant c such that

$$|f(x) - f(y)| \le c \|x - y\|, \ \forall x, \ y.$$

A simple estimator for $\mathbb{E}[f(S(T))]$ is

$$\hat{Y} = \frac{1}{M} \sum_{i=1}^{M} f(\widehat{S}^{(i)}_{T/h}).$$

The mean-square-error (MSE) of the estimator is defined as

$$\begin{aligned} \text{MSE} &= \mathbb{E}\left[\left(\mathbb{E}[f(S(T))] - \frac{1}{M}\sum_{i=1}^{M} f(\widehat{S}^{(i)}_{T/h})\right)^2\right] \\ &= \left(\mathbb{E}\left[f(S(T))\right] - \mathbb{E}\left[f(\widehat{S}_{T/h})\right]\right)^2 + \mathbb{E}\left[\left(\mathbb{E}\left[f(\widehat{S}_{T/h})\right] - \frac{1}{M}\sum_{i=1}^{M} f(\widehat{S}^{(i)}_{T/h})\right)^2\right] \\ &= \left(\mathbb{E}[f(S(T))] - \mathbb{E}\left[f(\widehat{S}_{T/h})\right]\right)^2 + \frac{1}{M}\mathbb{V}\left[f(\widehat{S}_{T/h})\right]. \end{aligned}$$

The RHS shows that the MSE comes from two sources: discretisation error and sampling error.

Provided some technical conditions on a, b, g, which are a bit more than twice continuous differentiability of a, b and fourth continuous differentiability of g (for details cf. Theorem 14.1.5 in [45]), the Euler discretisation achieves first order weak convergence, i.e. there exists a constant K such that

$$\left|\mathbb{E}[g(S(T))] - \mathbb{E}[g(\widehat{S}_{T/h})]\right| \le Kh. \tag{1.4}$$

For a uniform Lipschitz f, to get desired result we can let $g(x) \equiv x$. Therefore

$$\text{MSE} \le c^2 K^2 h^2 + \frac{1}{M} Var[f(S)].$$

The first term is attributed to bias caused by the discretisation, and the second part is due to the sampling error of Monte Carlo. Root mean square (RMS) error is defined as the square root of MSE. To achieve an $O(\epsilon)$ RMS requires $O(\epsilon^{-2})$ paths, each with $O(\epsilon^{-1})$ timesteps, leading to a computational complexity of $O(\epsilon^{-3})$.

1.3 Multilevel Monte Carlo

Giles [32] introduces a multilevel Monte Carlo path simulation method, inspired by the multigrid ideas for the iterative solution of systems of equations. The computational complexity for a general class of methods and applications is analysed, assuming that the discretisation satisfies certain conditions. As a corollary, it shows that the computational complexity can be reduced to $O(\epsilon^{-2}(\log\epsilon)^2)$ for European options using the Euler discretisation for Brownian SDEs and a Lipschitz payoff. For other path-dependent options, numerical results show significant computational savings.

For diffusion SDEs, suppose we perform Monte Carlo path simulations on different levels of resolution l, with 2^l uniform timesteps on level l. For a given Brownian path $W(t)$, let P denote the payoff, and let $\widehat{P}_l$ denote its approximation by a numerical scheme with timestep h_l. As a result of the linearity of the expectation operator, we have the following identity:

$$\mathbb{E}[\widehat{P}_L] = \mathbb{E}[\widehat{P}_0] + \sum_{l=1}^{L} \mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}]. \tag{1.5}$$

Let $\widehat{Y}_0$ denote the standard Monte Carlo estimate for $\mathbb{E}[\widehat{P}_0]$ using N_0 paths, and for $l > 0$, we use N_l independent paths to estimate $\mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}]$ using

$$\widehat{Y}_l = N_l^{-1} \sum_{i=1}^{N_l} \left(\widehat{P}_l^{(i)} - \widehat{P}_{l-1}^{(i)} \right). \tag{1.6}$$

The multilevel method exploits the fact that the multilevel correction variance $V_l := \mathbb{V}[\widehat{P}_l - \widehat{P}_{l-1}]$ decreases with l, and adaptively chooses N_l to minimise the computational cost to achieve a desired root-mean-square error. This is summarized in the following theorem:

Theorem 1.3.1. *Let P denote a functional of the solution of stochastic differential equation (1.1) for a given Brownian path $W(t)$, and let $\widehat{P}_l$ denote the corresponding approximation using a numerical discretisation with timestep $h_l = 2^{-l}\,T$.*

If there exist independent estimators $\widehat{Y}_l$ based on N_l Monte Carlo samples, and positive constants $\alpha \geq \frac{1}{2}\min(\beta, 1)$, β, c_1, c_2, c_3 such that

i) $\left| \mathbb{E}[\widehat{P}_l - P] \right| \leq c_1\, h_l^{\alpha}$

ii) $\mathbb{E}[\widehat{Y}_l] = \begin{cases} \mathbb{E}[\widehat{P}_0], & l = 0 \\ \mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}], & l > 0 \end{cases}$

iii) $\mathbb{V}[\widehat{Y}_l] \le c_2\, N_l^{-1} h_l^{\beta}$

iv) C_l*, the expected computational complexity of* $\widehat{Y}_l$*, is bounded by*

$$\mathbb{E}[C_l] \le c_3\, N_l\, h_l^{-1},$$

then there exists a positive constant c_4 *such that for any* $\epsilon < e^{-1}$ *there are values* L *and* N_l *for which the multilevel estimator*

$$\widehat{Y} = \sum_{l=0}^{L} \widehat{Y}_l,$$

has a mean-square-error with bound

$$MSE \equiv \mathbb{E}\left[\left(\widehat{Y} - \mathbb{E}[P]\right)^2\right] < \epsilon^2$$

with a expected computational complexity $\mathbb{E}[C]$ *with bound*

$$\mathbb{E}[C] \le \begin{cases} c_4\, \epsilon^{-2}, & \beta > 1, \\ c_4\, \epsilon^{-2}(\log \epsilon)^2, & \beta = 1, \\ c_4\, \epsilon^{-2-(1-\beta)/\alpha}, & 0 < \beta < 1. \end{cases}$$

Proof. See [32] for the original version and the generalisation to the case of random computational cost in [12]. □

The saving of the computational cost comes from the fact that as the level goes up, the N_l needed to achieve the accuracy for a given MSE is decreasing due to a decaying V_l. Thus the convergence rate of V_l plays a central role in achieving a good computational complexity.

1.4 Review of prior work

Based on [32]'s breakthrough, there has been a series of following work in the direction of financial applications. [31] presents significantly improved numerical results using the Milstein discretisation. For a Lipschitz European payoff, the Milstein method's improved strong convergence leads to $V_l = O(h_l^2)$, and therefore a computational complexity of $O(\epsilon^{-2})$ according to the computational complexity Theorem 1.3.1. For other payoffs, it shows that estimators using an approximation based on a Brownian interpolation technique achieves the same computational complexity numerically. [36] performs a numerical analysis of the multilevel Monte Carlo using the Euler discretisation, justifying the computational complexity results of other path-dependent payoffs in the numerical experiments of [32]. It is proved that $V_l = \mathcal{O}(h_l)$ for a Lipschitz payoff, Asian options and lookback options. It is also proved that $V_l = o(h_l^{1/2-\delta})$, for any $\delta > 0$, for barrier options and digital options. The final result has been tightened by Avikainen [4] who proved that $V_l = \mathcal{O}(h_l^{1/2} \log h_l)$. [38] applies the multilevel idea to quasi-Monte Carlo. By using randomised quasi-Monte Carlo techniques based on a rank-1 lattice rule, the computational complexity is further reduced. [27] examines the numerical performance of the multilevel treatment for basket options using the Milstein discretisation. [34] performs a numerical analysis of the multilevel Milstein method presented in paper [32]. [29] deals with an unusual parabolic SPDE which arises in a financial credit modelling application. [30] addresses the use of the Milstein approximation in multidimensional cases, which usually requires the simulation of Lévy areas. An antithetic technique without simulating Lévy areas is developed which gives a high rate of multilevel convergence, and numerical analysis is provided. [37] surveys the application of multilevel methods in computational finance. [28] reviews the progress on multilevel Monte Carlo, offering new ideas and suggesting future research.

In other direction of applications, [12, 55] apply the multilevel approach to elliptic SPDEs which arise in the modelling of nuclear waste repositories, with the permeability of the rock being modelled as a log-Normal stochastic field.

1.5 Thesis outline

This thesis concerns multilevel Monte Carlo for option pricing when the underlying follows a jump process. There have been a number of papers in related areas. The multilevel Monte Carlo method has been used by Dereich and Heidenreich [17, 16] for both finite and infinite activity Lévy-driven SDEs with payoffs Lipschitz continuous w.r.t. the supremum norm. The first part of the thesis differs in considering simpler finite activity jump-diffusion models, but has examples of more challenging non-Lipschitz payoffs (e.g. digital and barrier options), and also uses a more accurate

Milstein discretisation to achieve an improved order of convergence for the multilevel correction variance.

The second part of the thesis focuses on pricing in exponential Lévy models. This setting differs from models based on Lévy-driven SDEs due to a possible moment explosion issue which is discussed in section 5.3.4.2. Also, the Lévy processes we discuss allow exact simulation of increments where numerical discretisation is not needed for path generation. However, the running maximum is involved in lookback and barrier payoffs, and there is no straightforward method to simulate it. For the simulation of the running maximum of Lévy processes, [23] studies small-time asymptotic behavior of the exit probability and uses it for Monte Carlo computation of functionals of killed Lévy processes. [46] develops a novel Wiener-Hopf Monte-Carlo method to generate the joint distribution of $(S(T), \max_{0\leq t\leq T} S(t))$. This simulation technique is further extended to a multilevel setting in [22]. They obtain a computational complexity of $O\left(\epsilon^{-4}\right)$ for Lévy processes with infinite variation and $O\left(\epsilon^{-3}\right)$ for processes with bounded variation. In contrast to those advanced techniques, we take the discretely monitored maximum of the Lévy process as the approximation of the running supremum. With multilevel, this simple approximation achieves near optimal computational complexity in most cases.

The thesis is organised as follows. Chapter 2 discusses numerical simulation for constant rate and bounded state-dependent rate jump-diffusion processes, construction of a multilevel Monte Carlo estimator and presents numerical results for Asian, lookback, barrier and digital options. Chapter 3 rigorously justifies the computational complexity shown in Chapter 2 by numerical analysis of the variance of the multilevel estimators for European call, Asian, lookback, barrier and digital options in the constant rate and bounded state-dependent intensity jump-diffusion models. In Chapter 4 we deal with estimating sensitivities using multilevel Monte Carlo in jump-diffusion models. Especially for the sensitivity to the jump rate, two estimators are derived.

In Chapter 5 we discuss several examples of processes used in exponential Lévy models: Variance Gamma, Normal Inverse Gaussian and α-stable processes. Numerical results of multilevel Monte Carlo for Asian, lookback, barrier and digital options in exponential Lévy models are demonstrated. For the need of analysis, we estimate the p-th moment convergence rate of the discretely monitored running maximum for a large class of Lévy processes whose Lévy measures admit power law near the origin and have exponential tails. Based on the estimate, the numerical analysis of the variance of multilevel estimators is presented.

Chapter 2

Multilevel Monte Carlo Option Pricing In Jump-diffusion Models

2.1 Introduction

As mentioned in the introduction, we consider Merton[49] jump-diffusion model, in which the asset price follows a jump-diffusion SDE:

$$\mathrm{d}S(t) = a(S(t-),t)\,\mathrm{d}t + b(S(t-),t)\,\mathrm{d}W(t) + c(S(t-),t)\,\mathrm{d}J(t), \quad 0 \le t \le T,$$

where the jump term $J(t)$ is a compound Poisson process $\sum_{i=1}^{N(t)}(Y_i - 1)$, the jump magnitude Y_i has a prescribed distribution, and $N(t)$ is a Poisson process with intensity λ, independent of the Brownian motion. $\log Y_i$ has a normal distribution. with mean a and variance b^2, namely $\log Y_i \sim N(a, b^2)$.

There are several ways in which to generalize the Merton model. In section 2.5 we consider one case investigated by Glasserman & Merener [41], in which the jump rate depends on the asset price, namely $\lambda = \lambda(S(t-),t)$. Another possible way is to consider the case where the frequency of jump within finite horizon is infinite. In this direction, we discuss exponential Lévy models in chapter 5.

For European options, we are interested in the expected value of a function of the terminal state, $f(S(T))$, but in the case of exotic options the valuation depends on the entire path $S(t), 0 \le t \le T$. The expected value can be estimated by a simple Monte Carlo method with a suitable approximation to the SDE solution. However, if the discretisation has first order weak convergence then to achieve an $\mathcal{O}(\epsilon)$ root mean square (RMS) error requires $\mathcal{O}(\epsilon^{-2})$ paths, each with $\mathcal{O}(\epsilon^{-1})$ timesteps, leading to a computational complexity of $\mathcal{O}(\epsilon^{-3})$.

Giles [31, 32] introduced a multilevel Monte Carlo path simulation method, demonstrating that the computational cost can be reduced to $\mathcal{O}(\epsilon^{-2})$ for SDEs driven by Brownian motion. This has been extended by Dereich and Heidenreich [17, 16] to approximation methods for both finite and infinite activity Lévy-driven SDEs with globally Lipschitz payoffs. The first part of the thesis differs in considering simpler finite activity jump-diffusion models, but also one example of a more challenging non-Lipschitz payoff, and also uses a more accurate Milstein discretisation to achieve an improved order of convergence for the multilevel correction variance which will be defined later. The second part of the thesis emphasizes the exponential Lévy model which allows explicit simulation of increments where numerical discretisation is not needed for paths. The setting also differs from models based on Lévy-driven SDEs in the aspect of possible moment explosion.

In this chapter we apply the multilevel approach to the Monte Carlo simulation of path-dependent option pricing with jump-diffusion processes. We first consider the case where the jump rate is constant then take into account the state-dependent rate case. In both cases, in order to calculate coarse-path samples from fine-path samples using Brownian interpolation, we adopt a jump-adapted Milstein discretisation scheme proposed by [50], which explicitly simulates the times when jumps occur. Furthermore, we construct multilevel estimators for corresponding path-dependent payoffs coping with challenges caused by jumps. Through constructing payoff estimators by Brownian bridge technique, high order multilevel correction term variance convergence rate is achieved. In the state-dependent rate case, we use the thinning method to tackle the lack of synchronization of jump times in the fine and coarse grids. Numerical results show similar improvement in computational efficiency compared with previous achievements for diffusion processes [31]. Generally, using the jump-adapted Milstein scheme with the multilevel approach, we can reduce the computation cost to $\mathcal{O}(\epsilon^{-2})$ in terms of RMS error ϵ.

In the following sections of the chapter, we first review the Multilevel Monte Carlo method for diffusion processes. The next section describes the jump-adapted discretisation of jump-diffusion processes and its advantages for facilitating the multilevel approach. Then we discuss the path simulation and estimator construction for the jump-adapted discretisation with the multilevel approach and present numerical results of Asian, lookback, barrier and digital options. The next part establishes two methods to deal with state-dependent intensity. The final section draws conclusions and indicates directions of future research.

2.2 A Jump-adapted Milstein discretisation

To simulate jump-diffusion processes, it is possible to use fixed time grid schemes as for geometric Brownian motion. The Euler-Maruyama scheme for jump-diffusion processes has $\mathcal{O}(\sqrt{h})$ strong convergence ([51]). However, it would be more difficult to achieve higher order strong convergence. To achieve a higher order strong convergence for jump-diffusion processes, the Itô-Taylor expansion will involve some double integrals of white noise and the Poisson random measure [51], which increases the computational complexity of the simulation.

Another problem which might be encountered for fixed-time grid schemes is the construction of estimators for the path-dependent payoff. Within each timestep, running minimum or other functionals of paths is difficult to simulate due to the complex form of the joint density of diffusion and jump.

In order to avoid simulating double stochastic integrals as well as to identify the time at which the jump occurs, we use the so-called jump-adapted approximation proposed by Platen in [50]. This jump-adapted scheme would improve the computational tractability compared to other fixed time grid discretisation schemes with the same weak/strong convergence order.

For each path simulation, the set of jump times $\mathbb{J} = \{\tau_1, \tau_2, \ldots, \tau_m\}$ within the time interval $[0, T]$ is added to a set of uniformly spaced times $t'_i = i\,T/N,\ i = 0, \ldots, N$, to form a combined set of discretisation times $\mathbb{T} = \{0 = t_0 < t_1 < t_2 < \ldots < t_M = T\}$. As a result, the length of each timestep $h_n = t_{n+1} - t_n$ will be no greater than $h = T/N$.

Within each timestep the first order Milstein discretisation is used to approximate the SDE, and then the jump is simulated when the simulation time is equal to one of the jump times. This gives the following numerical method:

$$
\begin{aligned}
\widehat{S}^-_{n+1} &= \widehat{S}_n + a_n\, h_n + b_n\, \Delta W_n + \tfrac{1}{2}\, b'_n\, b_n\, (\Delta W_n^2 - h_n), \\
\widehat{S}_{n+1} &= \begin{cases} \widehat{S}^-_{n+1} \ + c(\widehat{S}^-_{n+1}, t_{n+1})(Y_i - 1), & \text{when } t_{n+1} = \tau_i; \\ \widehat{S}^-_{n+1}, & \text{otherwise,} \end{cases}
\end{aligned}
\tag{2.1}
$$

where the subscript n is used to denote the timestep index, $\widehat{S}^-_n = \widehat{S}(t_n-)$ is the left limit of the approximated path, ΔW_n is the Brownian increment during the timestep, a_n, b_n, b'_n are the values of a, b, b' based on $(\widehat{S}_n, t_n)$, and Y_i is the jump magnitude at τ_i.

The simulation algorithm via the jump-adapted scheme can be described as the following steps:

1. Set $i = 1,\ j = 1$, $t' = t'_0$;
2. Generate jump time τ_i in terms of its distribution;
3. While $(\tau_i < t'_j)$ do

   (1). Simulate the process within $[t', \tau_i)$, in which the process is driven purely by Brownian motion; then simulate the jump at τ_i;

   (2). Set $t' = \tau_i$;

   (3). $i = i + 1$, and generate next jump time τ_i in terms of its distribution;
4. Simulate the process within $[t', t'_j]$;
5. Set $j = j + 1$, $t' = t'_j$ and goto 3.

Jump-adapted schemes also save computational cost in our application. On average its computation cost for generating exponential random numbers for a sample path is λT. Compared with the cost of a fixed time schemes using N timesteps, the jump-adapted scheme saves substantial computation in applications where $\lambda T \ll N$.

2.3 Multilevel Monte Carlo method

Recall the multilevel Monte Carlo method in the introduction and the computational complexity Theorem 1.3.1. In the case of the jump-adapted discretisation, h_l should be taken to be the uniform timestep at level l, to which the jump times are added to form the set of discretisation times. Due to the presence of jump, the computational cost is

$$Cost = \sum_{l=0}^{L} \sum_{i=1}^{N_l} (N_T^{(i)} + 2^l), \tag{2.2}$$

where $N_T^{(i)}$ is the number of jumps in each scenario. We have to define the computational complexity as the expected computational cost since different paths may have different numbers of jumps. The expected number of jumps is finite and therefore the cost bound in assumption iv) will remain valid for an appropriate choice of the constant c_3. Hence, the multilevel algorithm framework works well.

2.4 Multilevel Monte Carlo for constant jump rate

2.4.1 Estimator construction

We elaborate on the definition of fine and coarse path approximations used for (1.6). Jump times are simulated by setting $\tau_j - \tau_{j-1} \sim \exp(\lambda)$, forming a jump time set $\mathbb{J} = \{\tau_1, \tau_2, \ldots, \tau_m\}$. The Brownian increments $\Delta W_n \equiv W(t_{n+1}) - W(t_n)$ are generated for each timestep of the fine grid $\mathbb{T}_l = \mathbb{J} \cup \left\{\frac{iT}{2^l}\right\} = \{0 = t_0 < t_1 < t_2 < \ldots < t_M = T\}$. Then the fine path is defined as the jump-adapted Milstein discretisation $\widehat{S}^f_n := \widehat{S}^f(t_n)\,,\; n = 0, \ldots, m + 2^l$ in (2.1). To define the coarse path approximation on $\mathbb{T}_{l-1} = \mathbb{J} \cup \left\{\frac{iT}{2^{l-1}}\right\}$, the Brownian increments generated for fine path are used in the following way. The Brownian increments for two consecutive timesteps on the fine grid $\mathbb{T}_l$ which contain so-called midpoints

$$\left\{\zeta_i = \frac{(2i-1)T}{2^l} : i = 1, \ldots, 2^{l-1}\right\}, \tag{2.3}$$

are summed to obtain the the Brownian increment for the timestep of $[\eta_i, \bar{\eta}_i]$ on the coarse grid $\mathbb{T}_{l-1}$ where

$$\begin{aligned} \eta_i &= \max\left\{t \in \mathbb{T}_{l-1} : t < \zeta_i\right\}, \\ \bar{\eta}_i &= \min\left\{t \in \mathbb{T}_{l-1} : t > \zeta_i\right\}. \end{aligned} \tag{2.4}$$

In the remaining timesteps of the coarse grid, the construction of the coarse path uses the same Brownian increment as the one in fine grid. So $\widehat{S}^c_n := \widehat{S}^c(t_n)\,,\; t_n \in \mathbb{T}_{l-1}$. This procedure is also illustrated in Figure 2.1.

The multilevel estimator (1.6) requires the calculation of the payoff difference $\widehat{P}^f_l - \widehat{P}^c_{l-1}$. Here $\widehat{P}^f_l$ is the fine-path estimate using timestep $h_l = 2^{-l}T$, and $\widehat{P}^c_{l-1}$ is the corresponding coarse-path estimate using timestep $h = 2^{-(l-1)}T$. As explained in [31], to ensure that the identity (1.5) is correctly respected, it is required that

$$\mathbb{E}[\widehat{P}^f_{l-1}] = \mathbb{E}[\widehat{P}^c_{l-1}]. \tag{2.5}$$

In the case of a European option, defining $\widehat{P}^f_{l-1}$ and $\widehat{P}^c_{l-1}$ as above leads to (2.5). However, for the path-dependent options, $\widehat{P}^c_{l-1}$ is sampled using random numbers generated for the sample of $\widehat{P}^f_l$. In these cases, it is necessary to verify (2.5). (2.5) holds since $\widehat{P}^c_{l-1}$ has the same distribution as $\widehat{P}^f_{l-1}$; they are based on the approximation that process has constant drift and volatility within timestep in $\mathbb{T}_{l-1}$.

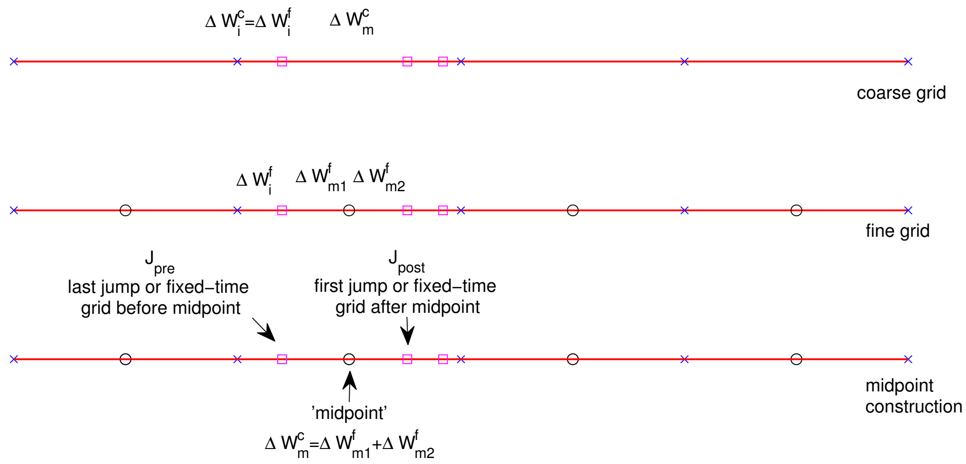

Figure 2.1: Midpoint construction, where crosses denote uniform points on the coarse grid; circles denote midpoints on the fine grid; squares denote jump times.

2.4.2 Numerical results

In the following we show numerical results for European call, Asian, lookback, barrier and digital options. All of the options are priced for the Merton model. We choose a risk-neutral measure where the jump risk can be diversified, under which the jump-diffusion SDE is

$$\frac{\mathrm{d}S(t)}{S(t-)} = (r - \lambda m)\,\mathrm{d}t + \sigma\,\mathrm{d}W(t) + \mathrm{d}J(t), \quad 0 \le t \le T,$$

where λ is the jump intensity, r is the risk-free interest rate, σ is the volatility, the jump magnitude satisfies $\log Y_i \sim N(a, b^2)$, and $m = \mathbb{E}[Y_i] - 1$ is the compensator to ensure the discounted asset price is a martingale. All of the simulations in this section use the parameter values $S_0 = 100$, $K = 100$, $T = 1$, $r = 0.05$, $\sigma = 0.2$, $a = 0.1$, $b = 0.2$, $\lambda = 1$. If we choose a relatively large value of $\lambda < 100$ the result would not change too much since in the expected computing cost expression (2.2), the average number of jump times is still inferior to the number uniform grid steps. The current code is based on Giles's code for [31], from which we generate standardised numerical results and figures.

2.4.2.1 European call option

Figure 2.2 shows the numerical results for the European call option with payoff $\exp(-rT)\,(S(T)-K)^+$, with $(x)^+ \equiv \max(x, 0)$ and strike $K = 100$.

The top left plot shows the behaviour of the variance of both $\widehat{P}_l$ and the multilevel correction $\widehat{P}_l - \widehat{P}_{l-1}$, estimated using 10^5 samples so that the Monte Carlo sampling error is negligible. The slope of the MLMC line indicates that $V_l \equiv \mathbb{V}[\widehat{P}_l - \widehat{P}_{l-1}] = \mathcal{O}(h_l^2)$, corresponding to $\beta = 2$ in condition *iii*) of Theorem 1.3.1. The top right plot shows that $\mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}]$ is approximately $\mathcal{O}(h_l)$, corresponding to $\alpha = 1$ in condition *i*). Noting that the payoff is Lipschitz, both of these are consistent with the first order strong convergence proved in [51].

The bottom two plots correspond to five different multilevel calculations with different user-specified accuracies to be achieved. These use the numerical algorithm given in [32] to determine the number of grid levels, and the optimal number of samples on each level, which are required to achieve the desired accuracy. We use the computational cost $\sum_{l=0}^{l} \sum_{i=1}^{N_l} (N_T^{(i)} + 2^l)$ to take into account the effect of jump. The left plot shows that in each case many more samples are used on level 0 than on any other level, with very few samples used on the finest level of resolution. The right plot shows that the the multilevel cost is approximately proportional to ϵ^{-2}, which agrees with the computational complexity bound in Theorem 1.3.1 for the $\beta > 1$ case.

2.4.2.2 Asian option

The payoff of the Asian option we consider is

$$P = \exp(-rT) \ \max\left(0, \overline{S} - K\right),$$

where

$$\overline{S} = T^{-1} \int_0^T S(t) \ \mathrm{d}t.$$

[31] shows that accuracy can be achieved by approximating the behaviour of a process within a timestep as an Itô process with constant drift and volatility, conditional on the endpoint values $\widehat{S}_n$ using Milstein schemes. The convergence of this approximation is proved in Chapter 3. In the jump-diffusion case, each timestep is random. Taking conditional expectations on the filtration generated by the jump times, each timestep is known. Let G denote the filtration $\sigma\{N_t,\ 0 \le t \le T\}$.

Suppose we use the jump-adapted Milstein discretisation (2.1), In other words, taking b_n to be the constant volatility within the timestep $[t_n, t_{n+1}]$, we define the Brownian interpolation at t as

$$\widehat{S}(t) = \widehat{S}_n + \upsilon_n(\widehat{S}_{n+1}^{-} - \widehat{S}_n) + b_n[W_t - W_n - \upsilon_n(W_{n+1} - W_n)],$$

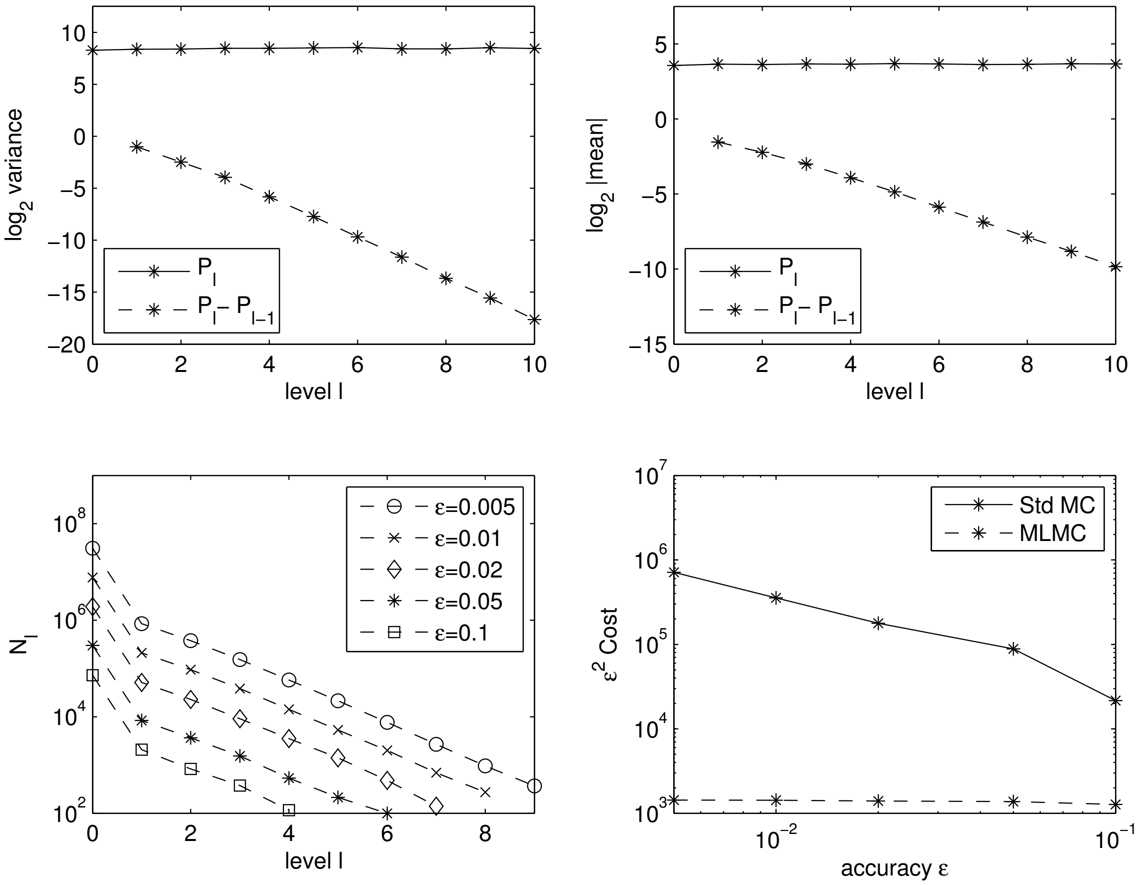

Figure 2.2: European call option with constant jump rate

where $\upsilon_n = (t - t_n)/h_n,\ h_n = t_{n+1} - t_n$.

By Lemma A.1.1, this implies that conditional on G,

$$\int_{t_n}^{t_{n+1}} \widehat{S}(t)\ \mathrm{d}t = \ \tfrac{1}{2} h_n(\widehat{S}_n + \widehat{S}^-_{n+1}) + b_n \Delta I_n,$$

where ΔI_n is

$$\Delta I_n := \int_{t_n}^{t_{n+1}} (W(t) - W(t_n))\ \mathrm{d}t \ - \ \tfrac{1}{2}\, h_n \Delta W_n,$$

where $h_n = t_{n+1} - t_n$, $\Delta W_n = W_{n+1} - W_n$. We have $\Delta I_n \mid G \sim N(0, h_n^3/12)$ and it is independent of ΔW_n conditional on G.

The approximated payoff for the fine path would be

$$\overline{S}^f = T^{-1} \sum_{n=0}^{n_T - 1} \Big(\ \tfrac{1}{2}\, h_n\, (\widehat{S}_n + \widehat{S}^-_{n+1}) + b_n \Delta I^f_n \Big).$$

$n_T = 2^l + N_T$ is the number of timesteps.

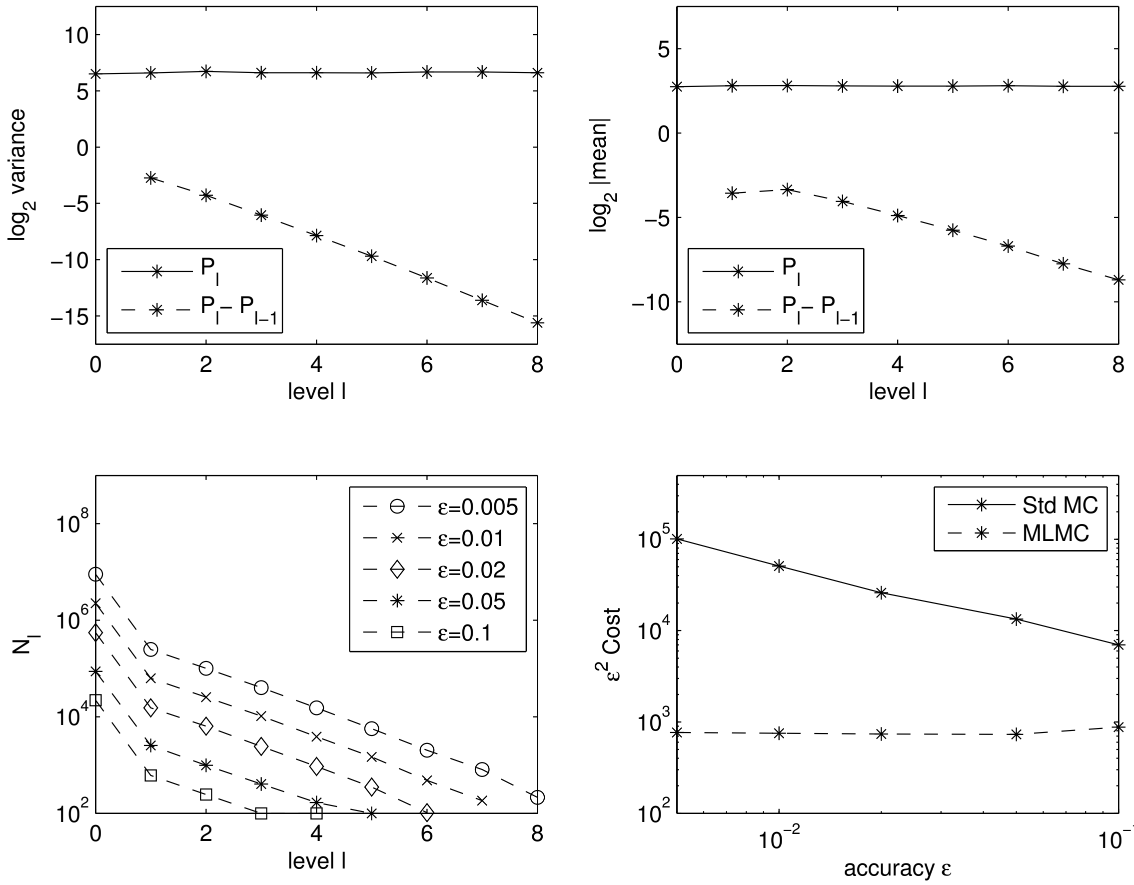

Figure 2.3: Asian option

For a particular timestep in the coarse path construction, we have two possible situations. If it does not contain one of the fine path discretisation times, and therefore corresponds exactly to one of the fine path timesteps, then it is treated in the same way as the fine path using exactly the same ΔI_n and ΔW_n generated for fine path. If the coarse timestep contains one of the fine path discretisation times, then it corresponds to $[t_n, t_{n+2}]$ using the index of the fine grid, where $n+1$ is the index of a midpoint. In this case, we have

$$\begin{aligned}
&\int_{t_n}^{t_{n+2}} (W(t) - W(t_n))\ \mathrm{d}t \ - \ \tfrac{1}{2}(t_{n+2}-t_n)(W(t_{n+2})-W(t_n)) \\
&= \ \int_{t_n}^{t_{n+1}} (W(t) - W(t_n))\ \mathrm{d}t \ - \ \tfrac{1}{2}\,(t_{n+1}-t_n)\,(W(t_{n+1})-W(t_n)) \\
&+ \ \int_{t_{n+1}}^{t_{n+2}} (W(t) - W(t_{n+1}))\ \mathrm{d}t \ - \ \tfrac{1}{2}\,(t_{n+2}-t_{n+1})\,(W(t_{n+2})-W(t_{n+1})) \\
&+ \ \tfrac{1}{2}\,(t_{n+2}-t_{n+1})\,(W(t_{n+1})-W(t_n)) - \ \tfrac{1}{2}\,(t_{n+1}-t_n)\,(W(t_{n+2})-W(t_{n+1}))\,,
\end{aligned}$$

and thus

$$\Delta I_n^c = \Delta I_n + \Delta I_{n+1} + \frac{1}{2}(t_{n+2} - t_n)(\mu_n \Delta W_n - (1-\mu_n)\Delta W_{n+1}),$$

where $\mu_n = (t_{n+2} - t_{n+1})/(t_{n+2} - t_n)$. ΔI^c corresponds to $[t_n, t_{n+2}]$ for the coarse path; on the RHS ΔI_n and ΔW_n are generated for the fine path.

Figure 2.3 shows the numerical results for parameters $S(0)=100$, $K=100$, $T=1$, $r=0.05$, $\sigma=0.2$, $a=0.1$, $b=0.2$, $\lambda=1$. All the results are similar to the pure diffusion case obtained in [31]. The top two plots indicate second order variance convergence rate and first order weak convergence, both of which are consistent with the analysis in Chapter 3. The computational cost of the multilevel method is therefore proportional to ϵ^{-2}, as shown in the bottom right plot.

2.4.2.3 Lookback option

The payoff of the lookback option we consider is

$$P = \exp(-rT)\left(S(T) - \min_{0\le t\le T} S(t)\right).$$

Previous work [31] achieved a second order convergence rate for the multilevel correction variance using the Milstein discretisation and an estimator constructed by approximating the behaviour within a timestep as an Itô process with constant drift and volatility, conditional on the endpoint values $\widehat{S}_n$ and $\widehat{S}_{n+1}$. Brownian Bridge results (see section 6.4 in [39]) give the minimum value within the timestep $[t_n, t_{n+1}]$, conditional on the end values, as

$$\widehat{S}_{n,min} = \frac{1}{2}\left(\widehat{S}_n + \widehat{S}_{n+1} - \sqrt{\left(\widehat{S}_{n+1} - \widehat{S}_n\right)^2 - 2\, b_n^2\, h_n \log U_n}\right), \tag{2.6}$$

where b_n is the constant volatility and U_n is a uniform random variable on $[0,1]$. The same treatment can be used for the jump-adapted discretisation in this chapter, except that $\widehat{S}_{n+1}^-$ must be used in place of $\widehat{S}_{n+1}$ in (2.6).

Equation (2.6) is used for the fine path approximation, but a different treatment is used for the coarse path, as in [31]. This involves a change to the original telescoping sum in (1.5) which now becomes

$$\mathbb{E}[\widehat{P}_L^f] = \mathbb{E}[\widehat{P}_0^f] + \sum_{l=1}^{L} \mathbb{E}[\widehat{P}_l^f - \widehat{P}_{l-1}^c], \tag{2.7}$$

where $\widehat{P}_l^f$ is the approximation on level l when it is the finer of the two levels being considered, and $\widehat{P}_l^c$ is the approximation when it is the coarser of the two. This modified telescoping sum remains valid provided $\mathbb{E}[\widehat{P}_l^f] = \mathbb{E}[\widehat{P}_l^c]$.

Considering a particular timestep in the coarse path construction, we have two possible situations. If it does not contain one of the fine path discretisation times, and therefore corresponds exactly to one of the fine path timesteps, then it is treated in the same way as the fine path, using the same uniform random number U_n. This leads naturally to a very small difference in the respective minima for the two paths.

The more complicated case is the one in which the coarse timestep contains one of the fine path discretisation times t', and so corresponds to the union of two fine path timesteps. In this case, the value at time t' is given by the Brownian interpolant

$$\widehat{S}(t') = \widehat{S}_n + \upsilon_n\,(\widehat{S}^-_{n+1} - \widehat{S}_n) + b_n\,(W(t') - W_n - \upsilon_n\,(W_{n+1} - W_n))\,, \tag{2.8}$$

where $\upsilon_n = (t' - t_n)/(t_{n+1} - t_n)$ and the value of $W(t')$ comes from the fine path simulation. Given this value for $\widehat{S}(t')$, the minimum values for $S(t)$ within the two timesteps $[t_n, t']$ and $[t', t_{n+1}]$ can be simulated in the same way as before, using the same uniform random numbers as the two fine timesteps.

The equality $\mathbb{E}[\widehat{P}_l^f] = \mathbb{E}[\widehat{P}_l^c]$ is respected in this treatment because $W(t')$ comes from the correct distribution, conditional on W_{n+1}, W_n, and therefore, conditional on the values of the Brownian path at the set of coarse discretisation points, the computed value for the coarse path minimum has exactly the same distribution as it would have if the fine path algorithm were applied.

Further discussion and analysis of this is given in Chapter 3, including a proof that the strong error between the analytic solution and the interpolation approximation is at most $\mathcal{O}(h \log h)$.

Figure 2.4 presents the numerical results. The results are very similar to those obtained by Giles for geometric Brownian motion [31]. The top two plots indicate second order variance convergence rate and first order weak convergence, both of which are consistent with the $\mathcal{O}(h \log h)$ strong convergence. The computational cost of the multilevel method is therefore proportional to ϵ^{-2}, as shown in the bottom right plot.

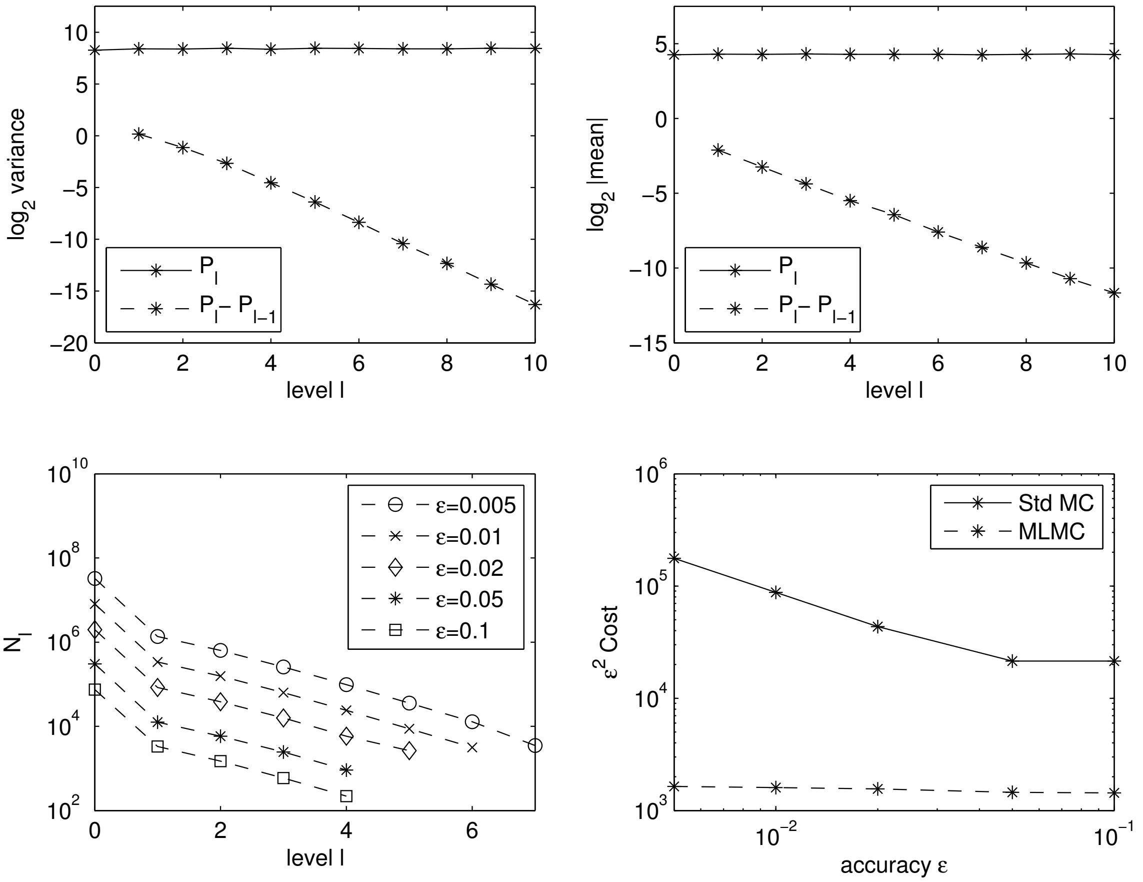

Figure 2.4: Lookback option with constant jump rate

2.4.2.4 Barrier option

We consider a down-and-out call barrier option for which the discounted payoff is

$$P = \exp(-rT)\,(S(T)-K)^+\,\mathbb{1}_{\{M_T>B\}},$$

where $M_T = \min_{0\le t\le T} S(t)$. The jump-adapted Milstein discretisation with the Brownian interpolation gives the approximation

$$\widehat{P} = \exp(-rT)\,(\widehat{S}(T)-K)^+\,\mathbb{1}_{\left\{\widehat{M}_T>B\right\}}$$

where $\widehat{M}_T = \min_{0\le t\le T} \widehat{S}(t)$. This could be simulated in exactly the same way as the lookback option, but in this case the payoff is a discontinuous function of the minimum M_T and an $\mathcal{O}(h)$ error in approximating M_T would lead to an $\mathcal{O}(h)$ variance for the multilevel correction.

Instead, following the approach of Cont & Tankov (see page 177 in [14]), it is better to use the expected value conditional on the values of the discrete Brownian increments and the jump times and magnitudes, all of which may be represented collectively as $\mathcal{F}$. This yields

$$
\begin{aligned}
&\mathbb{E}\left[\exp(-rT)\ (\widehat{S}(T)-K)^{+}\mathbb{1}_{\{\widehat{M}_T>B\}}\right] \\
&\quad = \mathbb{E}\left[\exp(-rT)\ (\widehat{S}(T)-K)^{+}\mathbb{E}\left[\mathbb{1}_{\{\widehat{M}_T>B\}} \mid \mathcal{F}\right]\right] \\
&\quad = \mathbb{E}\left[\exp(-rT)\ (\widehat{S}(T)-K)^{+} \prod_{n=0}^{n_T-1} \widehat{p}_n\right]
\end{aligned}
$$

where n_T is the number of timesteps, and $\widehat{p}_n$ denotes the conditional probability that the path does not cross the barrier B during the n^{th} timestep:

$$
\widehat{p}_n = 1-\exp\left(\frac{-2\,(\widehat{S}_n-B)^{+}(\widehat{S}^{-}_{n+1}-B)^{+}}{b_n^2\ (t_{n+1}-t_n)}\right). \tag{2.9}
$$

This barrier crossing probability is computed through conditional expectation and can be used to deduce (2.6).

For the coarse path calculation, we again deal separately with two cases. When the coarse timestep does not include a fine path time, we use (A.3). In the other case, when it includes a fine path time t' we evaluate the Brownian interpolant at t' and then use the conditional expectation to obtain

$$
\begin{aligned}
\widehat{p}_n &= \left\{1-\exp\left(\frac{-2\,(\widehat{S}_n-B)^{+}(\widehat{S}(t')-B)^{+}}{b_n^2\ (t'-t_n)}\right)\right\} \\
&\quad \times \left\{1-\exp\left(\frac{-2\,(\widehat{S}(t')-B)^{+}(\widehat{S}^{-}_{n+1}-B)^{+}}{b_n^2\ (t_{n+1}-t')}\right)\right\}.
\end{aligned} \tag{2.10}
$$

Figure 2.5 shows the numerical results for $K=100$, $B=85$. The top left plot shows that the multilevel variance is $\mathcal{O}(h_l^{\beta})$ for $\beta \approx 3/2$. This is similar to the behavior for a diffusion process [31]. The bottom right plot shows that the computational cost of the multilevel method is again almost perfectly proportional to ϵ^{-2}.

2.4.2.5 Digital option

The digital option considered here has the discounted payoff

$$
P = \exp(-rT)\ \mathbb{1}_{\{S(T)>K\}}.
$$

In [31], a multilevel variance convergence rate of $\mathcal{O}(h_l^{3/2})$ is achieved by smoothing the payoff using conditional expectation given the Brownian increments terminating

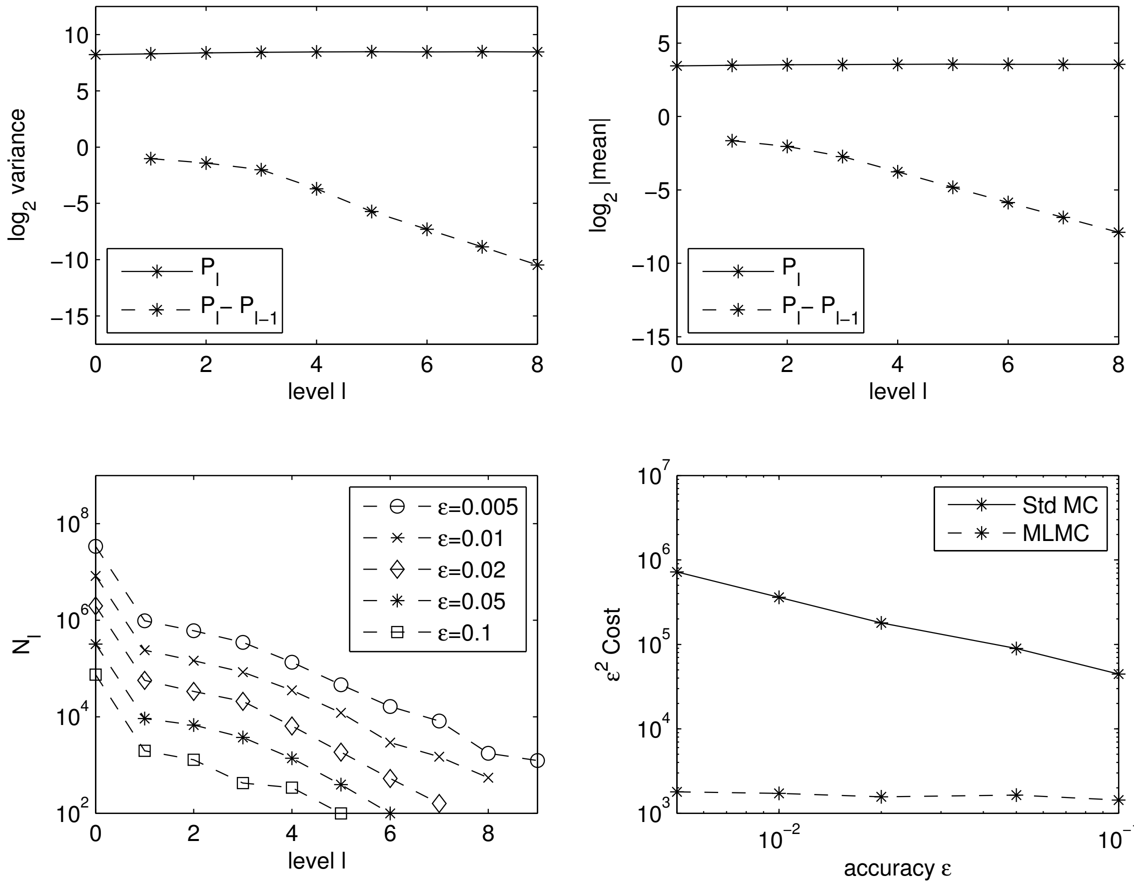

Figure 2.5: Barrier option with constant jump rate

one timestep before reaching the terminal time T. The estimator is the probability that $\widehat{S}_{n_l} > K$ under the assumption of simple Brownian motion with constant drift a_{n_l-1} and volatility b_{n_l-1} within the last timestep where :

$$
\begin{aligned}
& \mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}] \\
= \;& \mathbb{E}[\ \mathbb{E}[\ f(\widehat{S}^f_{n_l}) - f(\widehat{S}^c_{n_l}) \mid \Delta W_i,\ i = 1, \ldots, n_l - 1]\ ] \\
= \;& \mathbb{E}[\ \Phi\left(\frac{\widehat{S}^f_{n_l-1} + a^f_{n_l-1} h - K}{b^f_{n_l-1}\sqrt{h}}\right) - \Phi\left(\frac{\widehat{S}^c_{n_l-2} + 2a^c_{n_l-2} h + b^c_{n_l-2}\Delta W_{n_l-1} - K}{b^c_{n_l-2}\sqrt{h}}\right)\ ],
\end{aligned}
$$

where f is the discounted digital payoff, $n_l = 2^l$ denotes the number of fine-path timesteps and $h = T2^{-l}$, and Φ is the cumulative density function of the standard Normal distribution.

In the jump-adapted time grid, the relationship between the last jump time and the last timestep before expiry leads to different expressions for the above conditional

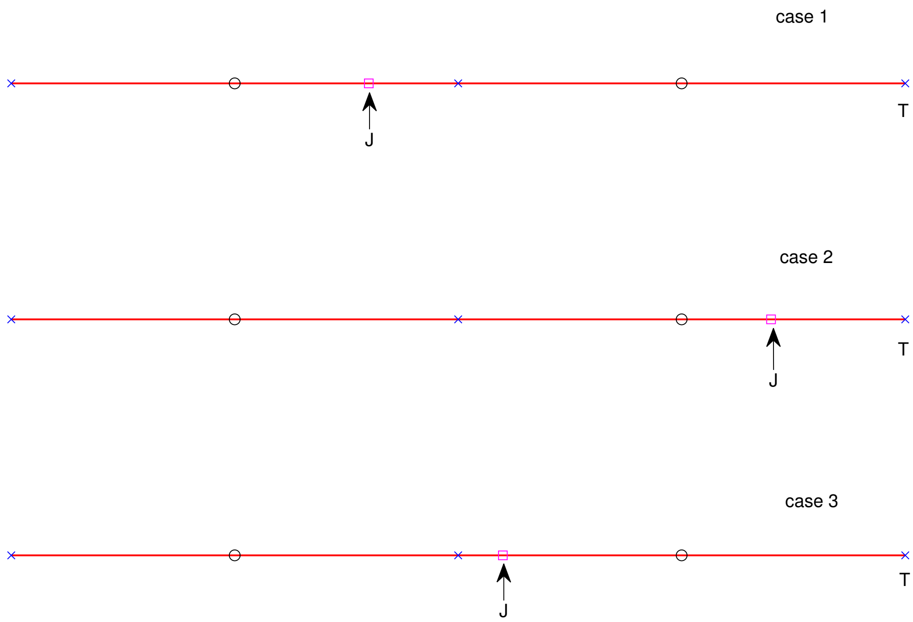

Figure 2.6: Construction of the conditional expectation estimator for the digital option

expectation estimator. Let n_T denote the number of timesteps in the discretisation grid. There are three cases:

1. The last jump time J happens before the penultimate fixed-time timestep, i.e. $J < (n_T - 2)\frac{T}{n_T}$;

2. The last jump time is within the last fixed-time timestep , i.e. $J > (n_T - 1)\frac{T}{n_T}$;

3. The last jump time is within penultimate fixed-time timestep, i.e. $(n_T - 1)\frac{T}{n_T} > J > (n_T - 2)\frac{T}{n_T}$.

Correspondingly, the expressions of fine-path and coarse-path estimators are shown in the following.

1. In the case 1, the last timestep of the fine grid and coarse grid is the same as the previous diffusion case. Since within each timestep, $\widehat{S}_t^f$ is a constant drift

and volatility diffusion process,

$$\widehat{P}_l^f = \Phi\left(\frac{\widehat{S}_{n_T-1}^f + a_{n_T-1}^f h - K}{\left|b_{n_T-1}^f\right|\sqrt{h}}\right),$$

where Φ is the cumulative Normal distribution function, $h = T/2^l$ is fine-path fixed timestep, drift $a_{n_T-1}^f \equiv a(\widehat{S}_{n_T-1}^f, T-h_{n_T})$ and volatility $b_{n_T-1}^f \equiv b(\widehat{S}_{n_T-1}^f, T-h_l)$.

To reduce variance we again use the Brownian increment $\Delta W_{n_T-2} \sim N(0,h)$ generated for the fine path,

$$\widehat{P}_{l-1}^c = \Phi\left(\frac{\widehat{S}_{n_T-2}^c + 2a_{n_T-2}^c h + b_{n_T-2}^c \Delta W_{n_T-2} - K}{\left|b_{n_T-2}^c\right|\sqrt{h}}\right).$$

2. In case 2, the last timestep of the fine grid would be $h_j = T - J$. Since the last timestep agrees on the fine and coarse discretisations, the expressions for the estimator for both fine and coarse path are

$$\begin{aligned}\widehat{P}_l^f &= \Phi\left(\frac{\widehat{S}_{n_T-1}^f + a_{n_T-1}^f h_j - K}{\left|b_{n_T-1}^f\right|\sqrt{h_j}}\right),\\ \widehat{P}_{l-1}^c &= \Phi\left(\frac{\widehat{S}_{n_T-1}^c + a_{n_T-1}^c h_j - K}{\left|b_{n_T-1}^c\right|\sqrt{h_j}}\right).\end{aligned} \tag{2.11}$$

3. In the last case, the last timestep of the fine grid is h, so

$$\widehat{P}_l^f = \Phi\left(\frac{\widehat{S}_{n_T-1}^f + a_{n_T-1}^f h - K}{\left|b_{n_T-1}^f\right|\sqrt{h}}\right).$$

Let J denote the last jump time, and the last timestep of the coarse grid is $h_j = T - J - h$. To reduce variance we again use Brownian increment $\Delta W_{n_T-2} \sim N(0, h_j)$ generated for the fine path,

$$\widehat{P}_{l-1}^c = \Phi\left(\frac{\widehat{S}_{n_T-2}^c + a_{n_T-2}^c h_j + b_{n_T-2}^c \Delta W_{n_T-2} - K}{\left|b_{n_T-2}^c\right|\sqrt{h}}\right).$$

In all three cases, the conditional expectation of the coarse-path estimator is equal to the fine-path one on the same level, thus equality (2.5) is justified. Figure 2.6 clearly demonstrates three cases.

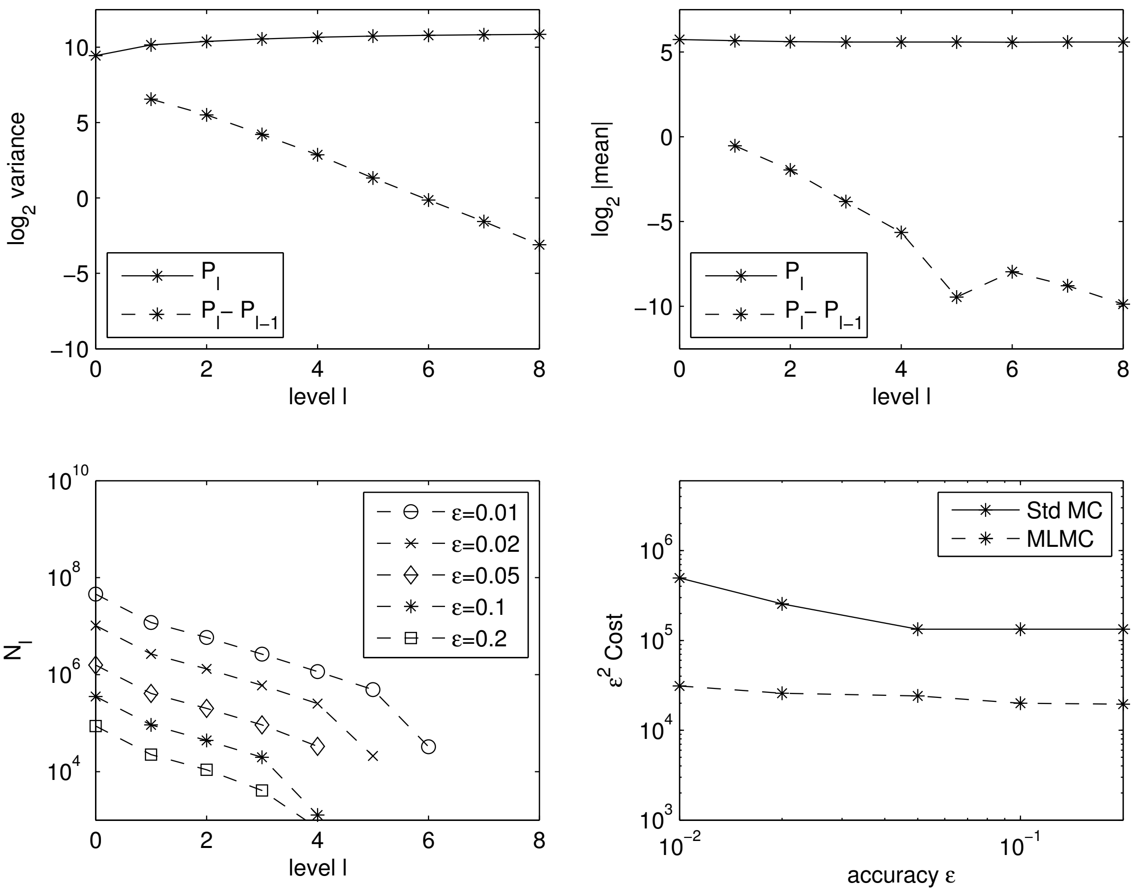

Figure 2.7: Digital option

Figure 2.7 shows the numerical results for parameters $S(0)=100$, $K=100$, $T=1$, $r=0.05$, $\sigma=0.2$. The top left plot shows that the variance is approximately $\mathcal{O}(h_l^{3/2})$, corresponding to $\beta=1.5$. The reason for this is similar to the argument in [31].

A different feature compared to the geometric Brownian motion case is that the variance of the level 0 estimator is a constant increasing with jump rate λ, instead of zero. The reason is simply that the trajectories at level 0 vary in each simulation due to the presence of jumps.

2.5 Path-dependent rates

In the case of a path-dependent jump rate $\lambda(S_t, t)$, the implementation of the multilevel method becomes more difficult because the coarse and fine path approximations may jump at different times. These differences could lead to a large difference between the coarse and fine path payoffs, and hence greatly increase the variance of the multilevel correction.

To tackle this obstacle, we modify the simulation approach through the acceptance-rejection technique of Glasserman & Merener [41] which uses "thinning" to treat the case when λ is bounded.

The idea of the thinning method is to construct a Poisson process with a constant rate λ_{sup} which is an upper bound of the state-dependent rate. This gives a set of candidate jump times, and these are then selected as true jump times with probability $\lambda(S_t, t)/\lambda_{\text{sup}}$.

If we approximate the process between two jump times using the Milstein scheme, we have the following jump-adapted Milstein scheme with thinning:

1. Generate the jump-adapted time grid for a Poisson process with constant rate λ_{sup};
2. Simulate each timestep using the Milstein discretisation;
3. When the endpoint t_{n+1} is a candidate jump time, generate a uniform random number $U \sim [0,1]$, and if $U < p_{t_{n+1}} = \dfrac{\lambda(\widehat{S}(t_{n+1}-), t_{n+1})}{\lambda_{\text{sup}}}$, then accept t_{n+1} as a real jump time and simulate the jump.

2.5.1 Multilevel treatment

In the multilevel implementation, if we use the above algorithm with different acceptance probabilities for fine and coarse level, there may be some samples in which a jump candidate is accepted for the fine path, but not for the coarse path, or vice versa. Because of first order strong convergence, the difference in acceptance probabilities will be $\mathcal{O}(h_l)$, and hence there is an $\mathcal{O}(h_l)$ probability of coarse and fine paths differing in accepting candidate jumps. Such differences will give an $\mathcal{O}(1)$ difference in the payoff value, and hence the multilevel variance will be $\mathcal{O}(h_l)$. A more detailed analysis is given in chapter 3.

To improve the variance convergence rate, we use a change of measure so that the acceptance probability is the same for both fine and coarse paths. This is achieved by taking the expectation with respect to a new measure $\mathbb{Q}$:

$$\mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}] = \mathbb{E}_Q[\widehat{P}_l \prod_\tau R_\tau^f - \widehat{P}_{l-1} \prod_\tau R_\tau^c]$$

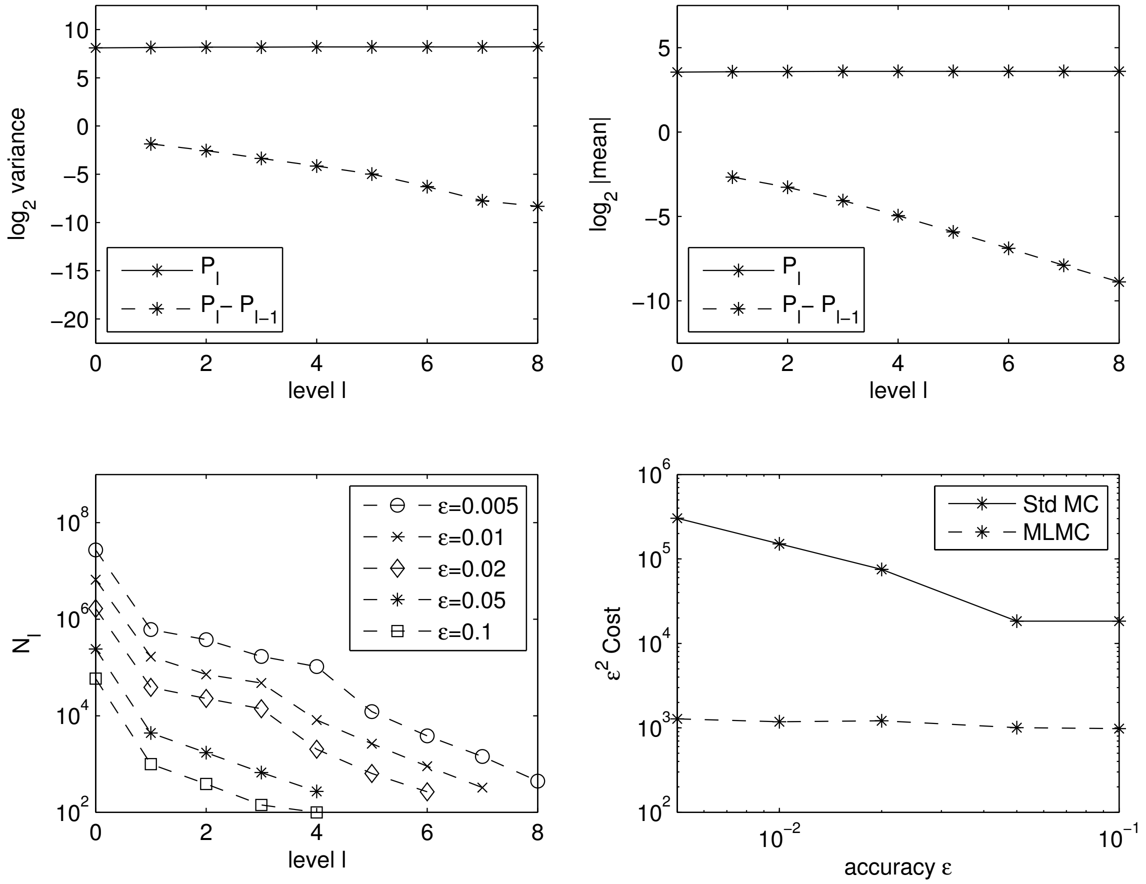

Figure 2.8: European call option with path-dependent jump rate using thinning without a change of measure

where τ are the jump times. The acceptance probability for a candidate jump under the measure $\mathbb{Q}$ is $\frac{1}{2}$ for both coarse and fine paths, instead of $p_\tau = \lambda(S(\tau-), \tau) \,/\, \lambda_{\text{sup}}$. The corresponding Radon-Nikodym derivatives are

$$R^f_\tau = \begin{cases} 2p^f_\tau, & \text{if } U_\tau < \frac{1}{2} \quad ; \\ 2(1-p^f_\tau), & \text{if } U_\tau \geq \frac{1}{2} \quad , \end{cases} \qquad R^c_\tau = \begin{cases} 2p^c_\tau, & \text{if } U_\tau < \frac{1}{2} \quad ; \\ 2(1-p^c_\tau), & \text{if } U_\tau \geq \frac{1}{2} \quad , \end{cases}$$

where U_τ are $(0,1)$ independent uniform random variables. Under Q, $\widehat{P}_l$ and $\widehat{P}_{l-1}$ are simulated on the jump-adapted time grid generated with a constant rate λ_{sup} where we simulate the jump magnitude if $U_\tau < \frac{1}{2}$; R^f_τ and R^c_τ are generated using the same U_τ.

Since $R^f_\tau - R^c_\tau = \mathcal{O}(h_l)$ and $\widehat{P}_l - \widehat{P}_{l-1} = \mathcal{O}(h_l)$, this results in the multilevel correction variance $\mathbb{V}_Q[\widehat{P}_l \prod_\tau R^f_\tau - \widehat{P}_{l-1} \prod_\tau R^c_\tau]$ being $\mathcal{O}(h_l^2)$.

If the analytic formulation is expressed using the same thinning and change of

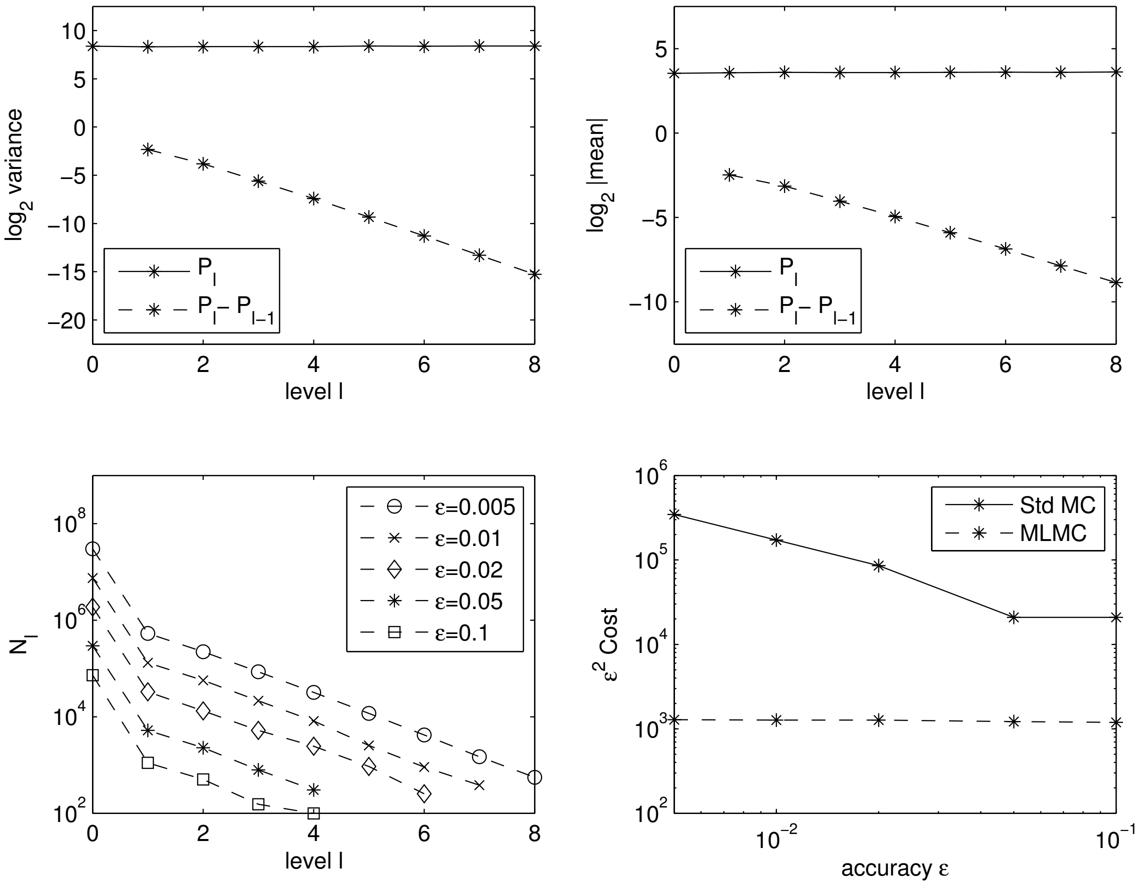

Figure 2.9: European call option with path-dependent jump rate using thinning with a change of measure

measure, the weak error can be decomposed into two terms as follows:

$$\mathbb{E}_Q\left[\widehat{P}_l \prod_\tau R^f_\tau - P \prod_\tau R_\tau\right] \;=\; \mathbb{E}_Q\left[(\widehat{P}_l - P)\;\prod_\tau R^f_\tau\right] \;+\; \mathbb{E}_Q\left[P\;(\prod_\tau R^f_\tau - \prod_\tau R_\tau)\right].$$

Using Hölder's inequality, the bound $\max(R_\tau, R^f_\tau) \leq 2$ and standard results for a Poisson process, the first term can be bounded using weak convergence results for the constant rate process, and the second term can be bounded using the corresponding strong convergence results which are explained in Chapter 3. This guarantees that the multilevel procedure does converge to the correct value.

2.5.2 Numerical results

We show numerical results for a European call option using

$$\lambda = \frac{1}{1 + (S(t-)/S_0)^2}, \quad \lambda_{\text{sup}} = 1,$$

and with all other parameters as used previously for the constant rate cases.

Comparing Figures 2.8 and 2.9 we see that the variance convergence rate is significantly improved by the change of measure, but there is little change in the computational cost. This is due to the main computational effort being on the coarsest level, which suggests using quasi-Monte Carlo on that level [38].

The bottom left plot in Figure 2.8 shows a slightly erratic behaviour. This is because the $\mathcal{O}(h_l)$ variance is due to a small fraction of the paths having an $\mathcal{O}(1)$ value for $\widehat{P}_l - \widehat{P}_{l-1}$. In the numerical procedure, the variance is estimated using an initial sample of 100 paths. When the variance is dominated by a few outliers, this sample size is not sufficient to provide an accurate estimate, leading to this variability.

2.6 Conclusions

In this chapter we have extended the multilevel Monte Carlo method to scalar jump-diffusion SDEs using a jump-adapted discretisation. Second order variance convergence is maintained in the constant rate case for European options with Lipschitz payoffs, and also for lookback options by constructing estimators using a previous Brownian interpolation technique. Variance convergence of order 1.5 is obtained for barrier and digital options, which again matches the convergence order which has been achieved previously for scalar SDEs without jumps. In the state-dependent rate case, we use thinning with a change of measure to avoid asynchronous jumps in the fine and coarse levels.

Chapter 3

Numerical Analysis Of Multilevel Monte Carlo For Scalar Jump-diffusion SDEs

3.1 Introduction

In the previous chapter, we applied multilevel Monte Carlo(MLMC) approach to finite-rate jump-diffusion SDEs, using a jump-adapted Milstein discretisation scheme for the constant rate case and with a thinning procedure for the state-dependent intensity case. The numerical results showed that in terms of RMS error ϵ the computational cost of $\mathcal{O}(\epsilon^{-2})$ is obtained for the estimation of several payoffs. Recall that due to Theorem 1.3.1, it is the **convergence rate of** V_l, i.e. the variance of the multilevel (ML) correction term, as $l \to \infty$ that determines the computational complexity of MLMC. Extending the analysis in [34], in this chapter we will establish the analysis of the variance convergence order needed in each case of Lipschitz, Asian, lookback, barrier, and digital payoffs. Those numerical and analysis results are summarised in Table 3.1.

The structure of this chapter is as follows. Section 3.2 presents notation of Poisson random measure and lists the assumptions on the jump-diffusion SDEs. Section 3.3 deals with the constant jump rate setting. We first review the jump-adapted Milstein scheme and prove its strong convergence. Then we define Brownian Interpolation, on which the construction of estimators for the running maximum and the crossing probability for path-dependent options are based. In order to perform numerical analysis later, we bound the distance between the Brownian Interpolant and the Kloeden-Platen interpolant in several norms. Afterwards we use bounds for

	Euler for diffusions		Milstein for diffusions		Jump-adapted Milstein	
option	numerical	analysis	numerical	analysis	numerical	analysis
Lipschitz	$\mathcal{O}(h)$	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h^2)$
Asian	$\mathcal{O}(h)$	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h^2)$
lookback	$\mathcal{O}(h)$	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$	$o(h^{2-\delta})$	$\mathcal{O}(h^2)$	$o(h^{2-\delta})$
barrier	$\mathcal{O}(h^{1/2})$	$o(h^{1/2-\delta})$	$\mathcal{O}(h^{3/2})$	$o(h^{3/2-\delta})$	$\mathcal{O}(h^{3/2})$	$o(h^{1-\delta})$
digital	$\mathcal{O}(h^{1/2})$	$\mathcal{O}(h^{1/2}\log h)$	$\mathcal{O}(h^{3/2})$	$o(h^{3/2-\delta})$	$\mathcal{O}(h^{3/2})$	$o(h^{3/2-\delta})$

Table 3.1: Orders of convergence for V_l as observed numerically and proved analytically for both the Euler and Milstein discretisations for pure diffusions and jump-diffusions; δ can be any strictly positive constant. $h \equiv h_l$.

moments of the distance between numerical and analytic solutions to establish an extreme path argument. Consequently, the estimates for convergence rates of V_l for Lipschitz, Asian, lookback, barrier and digital option are obtained. Section 3.4 deals with the state-dependent intensity setting, where the estimator is constructed by the thinning scheme with a change of measure to reduce variance. The variance of the ML estimator is analysed through a decomposition of the estimator into two parts: a ML correction term for a constant rate, and the discrepancy between Radon-Nikodym derivatives on fine and coarse levels.

3.2 Notations and assumptions

3.2.1 Jump-diffusion SDE notation

We follow the notations in [51]. Given a filtered probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$ satisfying the usual conditions, and a Borel set called the mark space $E \subseteq R^r \setminus \{0\}$, we have a Poisson random measure $p_\varphi(\mathrm{d}z, \mathrm{d}t)$ on $E \times [0, \infty]$. Its compensator is $\varphi(\mathrm{d}z)\mathrm{d}t = \lambda g(z)\mathrm{d}z\mathrm{d}t$, where λ is the intensity, and $g(z)$ is the p.d.f. of the mark. Intuitively, $p_\varphi(A, [0, t])$ counts the number of jumps with magnitude (mark) belonging to $A \subseteq E$ occurring in the time interval $[0, t]$.

The definition of random measure can be generalized to allow flexibility in time and state dependence on the intensity λ and the mark z. For rigorous details we refer to [42]. Following the notation in [51], the dynamics of a state-dependent jump-diffusion SDEs can be written as

$$\mathrm{d}S(t) = a(S(t-),t)\mathrm{d}t + b(S(t-),t)\mathrm{d}W(t) + \int_{z\in E} c(S(t-),t,z)\mu(\mathrm{d}z,\mathrm{d}t), \quad 0 \le t \le T, \tag{3.1}$$

where the compensator of the random measure μ is $\varphi(\mathrm{d}z)\mathrm{d}t = \lambda(S(t-),t)g(z)\mathrm{d}z\mathrm{d}t$. Note that the scope of this dynamics is wider than (1.2) in Chapter 2 since it allows the form of jump coefficient to be a general $c(S(t-),t,z)$ instead of a separable form of $c(S(t-),t)(z-1)$ in (1.2).

In the following we will focus on analysing (3.1) where $\mu(\omega;\cdot,\cdot)$ is a Poisson random measure $p_\varphi(\omega;\cdot,\cdot)$. This is because in Section 3.4 we will demonstrate how to use thinning to relate general SDE with state-dependent intensity to SDE with a Poisson random measure and a new jump coefficient c.

3.2.2 Some assumptions on the jump-diffusion SDEs

We shall restrict our model within scalar jump-diffusion SDE (3.1), namely the vector $S(t)$ is scalar and the mark space $E \subset \mathbb{R}$. We also impose that the drift function $a \in C^{1,1}(\mathbb{R}\times\mathbb{R}^+)$, volatility function $b \in C^{2,1}(\mathbb{R}\times\mathbb{R}^+)$ and jump coefficient $c \in C^{2,2,1}(\mathbb{R}\times\mathbb{R}^+\times E)$ satisfy the following standard conditions in which we use the operator $L_0 \equiv \partial/\partial t + a\,\partial/\partial S$ and $L_1 \equiv b\,\partial/\partial S$.

- A1 (bounded intensity): there exists $\lambda_{\sup}$ such that $\lambda(x,t) \le \lambda_{\sup}$. λ is absolutely continuous in t.
- A2 (uniform Lipschitz condition): there exists K_1 such that

$$\begin{aligned}|a(x,t)-a(y,t)| + |b(x,t)-b(y,t)| + |L_1b(x,t)-L_1b(y,t)| \\ + |c(x,t,z)-c(y,t,z)| \le K_1\,|x-y|.\end{aligned}$$

- A3 (linear growth bound): there exists K_2 such that

$$\begin{aligned}&|a(x,t)| + |L_0a(x,t)| + |L_1a(x,t)| + |b(x,t)| + |L_0b(x,t)| \\ &+ |L_1b(x,t)| + |L_0L_1b(x,t)| + |L_1L_1b(x,t)| + \left(\int_{z\in E} |c(x,t,z)|^2\,\lambda_{\sup}g(z)\mathrm{d}z\mathrm{d}t\right)^{1/2} \\ \le\;& K_2\,(1+|x|).\end{aligned}$$

- A4 (additional Lipschitz condition on diffusion coefficient): there exists K_3 such that

$$|b(x,t)-b(x,s)| \le K_3\,(1+|x|)\,\sqrt{|t-s|}.$$

- A5 (uniform Lipschitz condition on intensity): there exists K_4 such that

$$|\lambda(x,t) - \lambda(y,t)| \leq K_4\,|x-y|.$$

Bounded intensity is to guarantee that the thinning method is correct. Extension to locally bounded intensity is possible; see [26]. The absolutely continuity requirement in A1 is for the weak convergence proof of the jump-adapted Milstein scheme with the thinning in [41]. Assumption A5 is used to bound the discrepancy between Radon-Nikodym derivatives on fine and coarse levels in Section 3.4.

The following result is a general version of Theorem 1.9.3 in [51]. It can be proved using Theorem 66 mimicking the proof of Theorem 67 in [52].

Theorem 3.2.1. *Provided Assumption A2-A4 are satisfied, for any integer $m > 1$, the solution to (3.1) where $\mu(\omega;\cdot,\cdot)$ is a Poisson random measure $p_\varphi(\omega;\cdot,\cdot)$ admits*

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |S(t)|^m\right] < C\left(1+|S_0|^m\right).$$

where C is a positive constant.

3.3 Analysis in the constant rate setting

In this section we discuss the case in which the compensator $\varphi(\mathrm{d}z)\mathrm{d}t = \lambda(S(t-),t)g(z)\mathrm{d}z\mathrm{d}t$ of (3.1) has a constant intensity $\lambda(S(t-),t) = \lambda$.

3.3.1 A jump-adapted Milstein discretisation

We briefly review the jump-adapted Milstein scheme presented in chapter 2. Suppose that we have simulated the jump time grid $\mathbb{J} = \{\tau_1, \tau_2, \ldots, \tau_{N_T}\}$, which includes times at which jumps occur in $[0,T]$. On the other hand, consider a fixed time grid constituted of N timesteps, $t'_i = \frac{iT}{N}$, $i = 0, 1, \ldots, N$, which is used in discretisation schemes for diffusive SDEs. Now the superposition of them will be a jump-adapted grid $\mathbb{T} = \{0 = t_0 < t_1 < t_2 < \ldots < t_{N_T+N} = T\}$, and the timestep of this grid is denoted by $h_n = t_{n+1} - t_n, n = 0, \ldots, N_T + N - 1$. We define a jump-adapted Milstein scheme as

$$\begin{aligned}
&\widehat{S}^-_{n+1} = \widehat{S}_n + a_n\,h_n + b_n\,\Delta W_n + \tfrac{1}{2}\,b'_n\,b_n\,(\Delta W_n^2 - h_n),\\
&\widehat{S}_{n+1} = \begin{cases} \widehat{S}^-_{n+1} \;+ c(\widehat{S}^-_{n+1}, t_{n+1})(Y_i - 1), & \text{when } t_{n+1} \in \mathbb{J}; \\ \widehat{S}^-_{n+1}, & \text{otherwise.} \end{cases} \quad \text{Merton model} \\
&\widehat{S}_{n+1} = \widehat{S}^-_{n+1} + \textstyle\int_{z\in E} c(\widehat{S}^-_{n+1}, t_{n+1}, z)p_\lambda(\mathrm{d}z, t_{n+1}). \text{ General model}
\end{aligned} \qquad (3.2)$$

where $Y_i = \int_{z\in E} p_\lambda(\mathrm{d}z, t_{n+1})$ is the amplitude of the $i-$th jump, $\widehat{S}_n^- = \widehat{S}(t_n-)$, $b' \equiv \partial b/\partial S$, the subscript n denotes the timestep index, and a_n, b_n and b'_n are evaluated at $\widehat{S}_n, t_n$. With initial data $\widehat{S}_0 = S(0)$, this is the jump-adapted Milstein discretisation of equation (3.1) using a maximum timestep of size $h = \frac{T}{N}$.

[51] defines the Kloeden-Platen (KP) interpolant of (3.2):

$$\widehat{S}_{KP}(t) = \widehat{S}_n + a_n\,(t-t_n) + b_n\,(W(t)-W_n) + \tfrac{1}{2}\,b'_n\,b_n\,\left((W(t)-W_n)^2 - (t-t_n)\right), \quad (3.3)$$

for $t_n \le t < t_{n+1}$ and proves the following result.

Theorem 3.3.1. *Provided the assumptions A2-A4 are satisfied, then for* $m = 2$ *there exists a constant* $C\,(m,T)$ *such that for the solution to (3.1) and (3.3)*

$$\begin{aligned} \mathbb{E}\left[\sup_{0\le t\le T} |S(t) - \widehat{S}_{KP}(t)|^m\right] &< C\,(m,\lambda,T,K_{1,2})\,(1+|S_0|^m)\;h^m, \\ \mathbb{E}\left[\sup_{0\le t\le T} |\widehat{S}_{KP}(t)|^m\right] &< C\,(1+|S_0|^m)\,. \end{aligned}$$

Basically their result can be generalised to the case for any integer $m \geq 2$ with the same methodology:

Proof. The second result is implied by the first one and Theorem 3.2.1. It can also be seen as an extension of Lemma 6.6.1 in [51]. We prove the first one in the following.

We introduce a notation for $t_n \le t < t_{n+1}$:

$$\begin{aligned} & S(t) \\ = \;& S(t_n) + a_n\,(t-t_n) + b_n\,(W(t)-W_n) + \tfrac{1}{2}\,b'_n\,b_n\,\left((W(t)-W_n)^2 - (t-t_n)\right) + R_3(t_n,t) \\ \equiv \;& S(t_n) + f(t_n,t,S(t_n)) + R_3(t_n,t), \end{aligned}$$

where R_3 is the remainder of a Wagner-Platen expansion on page 189 of [51]. Then

$$\widehat{S}_{KP}(t) = \widehat{S}(t_n) + f(t_n,t,\widehat{S}(t_n)).$$

Follow the proof of Theorem 8.7.1 in [51],

$$Z\,(T) := \mathbb{E}\left[\sup_{0\le t\le T} |S(t) - \widehat{S}_{KP}(t)|^m\right]$$

can be bounded by the sum of three parts. The first part is

$$\mathbb{E}\Big[\sup_{0\le t\le T} \Big|\sum_{n=0}^{n_t-1} f(t_n,t_{n+1},S(t_n)) - f(t_n,t_{n+1},\widehat{S}(t_n)) + f(t_{n_t},t,S(t_{n_t})) - f(t_{n_t},t,\widehat{S}(t_{n_t}))\Big|^p\Big],$$

where n_t is the last grid point before t.

The second part is

$$\mathbb{E}\left[\sup_{0\le t\le T}\left|\sum_{n=0}^{n_t-1} R_3(t_n,t_{n+1})+R_3(t_{n_t},t)\right|^p\right].$$

According to the results in [45], the bounds on the first and second parts are

$$C\left(m,K_1\right)\int_0^T Z\left(u\right)\mathrm{d}u+C\left(K_2\right)\left(1+\left|S_0\right|^m\right)h^m.$$

The third part is

$$P_T:=\mathbb{E}\left[\sup_{0\le t\le T}\left|\int_0^t\int_{z\in E}\left(c(S(t-),t,z)-c(\widehat{S}_{KP}(t),t,z)\right)p_\varphi(\mathrm{d}z,\mathrm{d}t)\right|^m\right].$$

Let λ be the intensity rate of p_φ. $\tilde{p}_\varphi=p_\varphi-\varphi(\mathrm{d}z)\mathrm{d}t$ denotes the compensated measure of the random measure $p_\varphi(\mathrm{d}z,\mathrm{d}t)$. It follows that the stochastic integral of any predictive process w.r.t. $\tilde{p}_\varphi$ is a martingale. Applying the Burkholder–Davis–Gundy inequality (A.2.9) in the appendix to the stochastic integral w.r.t. $\tilde{p}_\varphi$, we have

$$\begin{aligned}
P_T \;\le\;& 2^{m-1}\Bigg(\mathbb{E}\left[\sup_{0\le t\le T}\left|\int_0^t\int_{z\in E}\left(c(S(t-),t,z)-c(\widehat{S}_{KP}(t),t,z)\right)\tilde{p}_\varphi(\mathrm{d}z,\mathrm{d}t)\right|^m\right]\\
&+\mathbb{E}\left[\sup_{0\le t\le T}\left|\int_0^t\int_{z\in E}\left(c(S(t-),t,z)-c(\widehat{S}_{KP}(t),t,z)\right)\varphi(\mathrm{d}z)\mathrm{d}t\right|^m\right]\Bigg)\\
\le\;& 2^{m-1}C_m\mathbb{E}\left[\left(\int_0^T\int_{z\in E}\left(c(S(t-),t,z)-c(\widehat{S}_{KP}(t),t,z)\right)^2\varphi(\mathrm{d}z)\mathrm{d}t\right)^{m/2}\right]\\
&+2^{m-1}\mathbb{E}\left[\sup_{0\le t\le T}\left|\int_0^t\int_{z\in E}\left(c(S(t-),t,z)-c(\widehat{S}_{KP}(t),t,z)\right)\varphi(\mathrm{d}z)\mathrm{d}t\right|^m\right]
\end{aligned}$$

Using the Jensen's inequality by noting that $\varphi(\mathrm{d}z)\mathrm{d}t/\left(\lambda T\right)$ can be seen as a probability measure since $\int_0^T\int_{z\in E}\varphi(\mathrm{d}z)\mathrm{d}t/\left(\lambda T\right)=1$, it follows that for $m\ge 2$

$$\begin{aligned}
&\left(\int_0^T\int_{z\in E}\left(c(S(t-),t,z)-c(\widehat{S}_{KP}(t),t,z)\right)^2\varphi(\mathrm{d}z)\mathrm{d}t/\left(\lambda T\right)\right)^{1/2}\\
\le\;&\left(\int_0^T\int_{z\in E}\left|c(S(t-),t,z)-c(\widehat{S}_{KP}(t),t,z)\right|^m\varphi(\mathrm{d}z)\mathrm{d}t/\left(\lambda T\right)\right)^{1/m}.
\end{aligned}$$

We have a similar inequality for the other term. Hence

$$\begin{aligned} P_T &\leq 2^{m-1} C_m \left(\lambda T\right)^{m/2-1} \mathbb{E}\left[\int_0^T \int_{z\in E} \left|c(S(t-),t,z) - c(\widehat{S}_{KP}(t),t,z)\right|^m \varphi(\mathrm{d}z)\mathrm{d}t\right] \\ &\quad + 2^{m-1} \left(\lambda T\right)^{m-1} \mathbb{E}\left[\int_0^T \int_{z\in E} \left|c(S(t-),t,z) - c(\widehat{S}_{KP}(t),t,z)\right|^m \varphi(\mathrm{d}z)\mathrm{d}t\right] \\ &\leq C\left(m,\lambda,T,K_1\right) \mathbb{E}\left[\int_0^T \left|S(t) - \widehat{S}_{KP}(t)\right|^m \mathrm{d}t\right] \\ &\leq C\left(m,\lambda,T,K_1\right) \int_0^T Z\left(u\right) \mathrm{d}u. \end{aligned}$$

In sum,

$$Z\left(T\right) \leq C\left(m,K_1\right) \int_0^T Z\left(u\right) \mathrm{d}u + C\left(K_2\right)\left(1+\left|S_0\right|^m\right) h^m + C\left(m,\lambda,T,K_1\right) \int_0^T Z\left(u\right) \mathrm{d}u.$$

By Grönwall's inequality we conclude that

$$\mathbb{E}\left[\sup_{0\leq t\leq T} \left|S(t) - \widehat{S}_{KP}(t)\right|^m\right] \leq C\left(m,\lambda,T,K_{1,2}\right)\left(1+\left|S_0\right|^m\right)\, h^m.$$

□

3.3.2 Interpolation approximation

In [34] the Brownian Interpolation is constructed by approximating the path as a constant drift and volatility diffusion between grid points. Here it is adapted with a slight difference-the length and number of timesteps are no longer deterministic. We can define

$$\widehat{S}(t) = \widehat{S}_n + v_n\left(\widehat{S}_{n+1}^{-} - \widehat{S}_n\right) + b_n\left(W(t) - W_n - v_n\left(W_{n+1} - W_n\right)\right) \tag{3.4}$$

for $t_n \leq t < t_{n+1}$, where $v_n \equiv (t-t_n)/h_n$. By definition it is eligible to generate running maximum/ crossing probability for $\widehat{S}(t)$ within grid points by Brownian Bridge results. In order to justify that the approximation constructed based on $\widehat{S}$ has the weak/ strong convergence, and to bound the variance of the ML term using extreme path theory, we shall bound the deviation between the KP interpolant and the Brownian interpolant in the sense of both strong convergence and weak convergence. Moreover we bound the distance between the two integrated interpolants in L^2 norm for the Asian payoff.

Theorem 3.3.2. *If $\widehat{S}(t)$ is the interpolant defined by (3.4) and $\widehat{S}_{KP}(t)$ is the Kloeden-Platen interpolant defined by (3.3) then for any integer $m > 0$*

i)

$$\mathbb{E}\left[\sup_{[0,T]}\left|\hat{S}(t)-\hat{S}_{KP}(t)\right|^m\right] = \mathcal{O}\left((h\log h)^m\right),$$

ii)

$$\sup_{[0,T]}\mathbb{E}\left[\left|\hat{S}(t)-\hat{S}_{KP}(t)\right|^m\right] = \mathcal{O}(h^m),$$

iii)

$$\mathbb{E}\left[\left(\int_0^T (\hat{S}(t)-\hat{S}_{KP}(t))\,\mathrm{d}t\right)^2\right] = \mathcal{O}(h^3).$$

Proof. In the following G denotes the filtration $\sigma\{N_t,\ 0 \le t \le T\}$. G gives the information of the jump-adapted grid.

By Theorem 3.3.1 and Assumption A2 we have a bound for $\mathbb{E}[\sup_n |b_n' b_n|^m]$. Let the number of fixed time grids be $N = T/h$. Then for $t \in [t_n, t_{n+1}]$, the difference between the two interpolants is

$$\widehat{S}(t) - \widehat{S}_{KP}(t) = \tfrac{1}{2}\, b_n' b_n\, Y(t),$$

where according to (3.4)

$$Y(t) = v_n(W_{n+1}-W_n)^2 - (W(t)-W_n)^2.$$

i) Using the Cauchy-Schwarz inequality, we have

$$\mathbb{E}\left[\sup_{[0,T]}\left|\hat{S}(t)-\hat{S}_{KP}(t)\right|^m\right] \le 2^{-m}\sqrt{\mathbb{E}\left[\sup_n |b_n' b_n|^{2m}\right]\ \mathbb{E}\left[\sup_{[0,T]}|Y(t)|^{2m}\right]}.$$

Moreover, there holds

$$|Y(t)| \le (W_{n+1}-W_n)^2 + (W(t)-W_n)^2.$$

By Lévy's modulus of continuity (see [20]):

$$\lim_{h\to 0}\sup_{\substack{0<|t-s|<h\\ 0<t,s<1}}\frac{|W(t)-W(s)|}{\sqrt{2h\log\frac{1}{h}}} = 1, \quad \text{a.s.}$$

we have

$$\mathbb{E}\left[\sup_{0\le t\le T} |Y(t)|^{2m}\right] \le C_m \left(h \log\frac{1}{h}\right)^{2m}.$$

Therefore the assertion holds.

ii) Let $W(t)-W_n = \sqrt{v_n\, h_n}\; Z_1$ and $W_{n+1}-W(t) = \sqrt{(1-v_n)\, h_n}\; Z_2$, where Z_1, Z_2 denote independent standard Normal random variables, one can prove that

$$|Y| \le h\ \max(Z_1^2, Z_2^2),$$

and hence $\mathbb{E}[\,|Y|^m] \le 2^{2-m}\,(2m-1)!\,h^m/(m-1)!$. The assertion then follows from

$$\mathbb{E}\left[\left|\hat{S}(t) - \hat{S}_{KP}(t)\right|^m\right] = 2^{-m}\,\mathbb{E}[\,|b_n' b_n|^m]\ \ \mathbb{E}[\,|Y|^m].$$

iii) Defining $X_n := \displaystyle\int_{t_n}^{t_{n+1}} Y(t)\ \mathrm{d}t$, we obtain

$$\mathbb{E}\left[\left(\int_0^T (\hat{S}(t) - \hat{S}_{KP}(t))\ \mathrm{d}t\right)^2\right] = \tfrac{1}{4}\mathbb{E}\left[\mathbb{E}\left[\left(\sum_{n=0}^{N+N_T-1} b_n' b_n X_n\right)^2 \mid G\right]\right].$$

For $n > m$, $\mathbb{E}[b_m' b_m X_m b_n' b_n X_n] = 0$ since X_n is independent of $b_m' b_m X_m b_n' b_n$ and $\mathbb{E}[X_n \mid G] = 0$. In addition, the X_n are independent of $b_n' b_n$ conditional on G. Therefore

$$\begin{aligned}
&\mathbb{E}\left[\left(\int_0^T (\hat{S}(t) - \hat{S}_{KP}(t))\ \mathrm{d}t\right)^2\right] \\
=\ & \tfrac{1}{4}\mathbb{E}\left[\mathbb{E}\left[\sum_{n=0}^{N+N_T-1} (b_n' b_n)^2\ \ X_n^2 \mid G\right]\right] \\
=\ & \tfrac{1}{4}\mathbb{E}\left[\sum_{n=0}^{N+N_T-1} \mathbb{E}\left[(b_n' b_n)^2\ \mid G\right] \mathbb{E}\left[\ X_n^2 \mid G\right]\right] \\
\le\ & \tfrac{1}{4}Ch^4\mathbb{E}\left[(N+N_T)\mathbb{E}\left[\sup_n\,(b_n' b_n)^2 \mid G\right]\right] \\
\le\ & \tfrac{1}{4}Ch^4\sqrt{\mathbb{E}\left[(N+N_T)^2\right]\mathbb{E}\left[\mathbb{E}\left[\sup_n\,(b_n' b_n)^2 \mid G\right]^2\right]} \\
\le\ & \tfrac{1}{4}Ch^4\sqrt{\mathbb{E}\left[(\tfrac{T}{h}+N_T)^2\right]\mathbb{E}\left[\sup_n\,(b_n' b_n)^4\right]} \\
=\ & \mathcal{O}(h^3)
\end{aligned}$$

where we use

$$\begin{aligned}\mathbb{E}\left[X_n^2 \mid G\right] &\le h_n \mathbb{E}\left[\int_{t_n}^{t_{n+1}} Y(t)^2 \,\mathrm{d}t \mid G\right] \\ &\le Ch^4\end{aligned}$$

due to the bound on $\mathbb{E}[\,|Y(t)|^m]$ in ii). The last two steps follow from Cauchy-Schwartz inequality to separate $N+N_T$ term, Jensen's inequality, linear growth bound and Theorem 3.3.1 for a finite pth moment of $\mathbb{E}\left[\sup_n (b_n' b_n)^4\right]$.

□

The following decomposition is useful in the analysis of variance of ML estimator:

$$\begin{aligned}\widehat{S}^f(t) - \widehat{S}^c(t) &= (\widehat{S}^f(t) - \widehat{S}^f_{KP}(t)) - (\widehat{S}^c(t) - \widehat{S}^c_{KP}(t)) \\ &\quad + (\widehat{S}^f_{KP}(t) - S(t)) - (\widehat{S}^c_{KP}(t) - S(t))\end{aligned} \tag{3.5}$$

with Theorem 3.3.2 bounding the error in the first two terms, and Theorem 3.3.1 bounding the error in the last two terms.

3.3.3 Lipschitz payoffs

The payoff of European options, such as vanilla call and put options, is a Lipschitz function of the value of the underlying asset at maturity,

$$P = f(S(T)).$$

A more general class of Lipschitz payoffs in which the payoff is a Lipschitz function of the values of the underlying asset at a finite number of times T_m,

$$P = f(S(T_1), S(T_2), \ldots, S(T_M)),$$

with the Lipschitz bound

$$\left| f(S_1^{(2)}, S_2^{(2)}, \ldots, S_M^{(2)}) - f(S_1^{(1)}, S_2^{(1)}, \ldots, S_M^{(1)}) \right| \le L \sum_{m=1}^{M} \left| S_m^{(2)} - S_m^{(1)} \right|,$$

for some constant L. In the numerical discretisation the fine and coarse path payoffs are both defined by

$$\widehat{P} = f(\widehat{S}(T_1), \widehat{S}(T_2), \ldots, \widehat{S}(T_M)),$$

with $\widehat{S}(t)$ given by the Brownian interpolation. Note that this will require the additional simulation of $W(T_m)$ if T_m does not correspond to one of the existing timesteps.

We get the following result concerning the variance of the multilevel estimator:

Theorem 3.3.3. *This approximation for Lipschitz payoffs has $V_l = \mathcal{O}(h_l^2)$.*

Proof. From the Lipschitz bound and Jensen's inequality we obtain

$$\mathbb{V}[\widehat{P}_l^f - \widehat{P}_{l-1}^c] \;\leq\; \mathbb{E}[(\widehat{P}_l^f - \widehat{P}_{l-1}^c)^2] \;\leq\; L^2 M \sum_{m=1}^{M} \mathbb{E}[(\widehat{S}^f(T_m) - \widehat{S}^c(T_m))^2].$$

The decomposition (3.5) implies that

$$\mathbb{E}[(\widehat{S}^f(T_m) - \widehat{S}^c(T_m))^2] \;\leq\; 4\Big(\mathbb{E}[(\widehat{S}^f(T_m) - \widehat{S}^f_{KP}(T_m))^2] + \mathbb{E}[(\widehat{S}^c(T_m) - \widehat{S}^c_{KP}(T_m))^2] + \mathbb{E}[(\widehat{S}^f_{KP}(T_m) - S(T_m))^2] + \mathbb{E}[(\widehat{S}^c_{KP}(T_m) - S(T_m))^2]\Big)$$

and the proof is completed using the results from Theorems 3.3.1 and 3.3.2. □

3.3.4 Asian options

Continuously monitored Asian options have a payoff $P = f(\overline{S}, S(T))$, that is a uniform Lipschitz function the average over the time interval

$$\overline{S} \equiv T^{-1} \int_0^T S(t)\ \mathrm{d}t,$$

and $S(T)$. We now consider two alternative numerical approximations.

3.3.4.1 Treatment 1

In chapter 2, the first treatment is that the approximated path averages $\overline{\widehat{S}}$ are defined by integrating the interpolant (3.4). Because of Lemma A.1.1, this gives

$$\int_0^T \widehat{S}^f(t)\,\mathrm{d}t = \sum_{n=0}^{N+N_T-1} \tfrac{1}{2}h_n\left(\widehat{S}^f_n + \widehat{S}^{f-}_{n+1}\right) + b_n\,\Delta I^f_n \tag{3.6}$$

where ΔI^f_n are independent $N(0,\ \frac{1}{12}h_n^3)$ variables, conditional on G which denotes the filtration $\sigma\{N_t,\ 0 \leq t \leq T\}$. The coarse path average is defined accordingly, and a straightforward calculation gives

$$\begin{aligned}\Delta I^c_n \;&\equiv\; \int_{t_n}^{t_{n+2}} (W(t) - W_n)\ \mathrm{d}t - \frac{1}{2}(t_{n+2} - t_n)(W(t_{n+2}) - W(t_n)) \\ &=\; \Delta I^f_n + \Delta I^f_{n+1} - \tfrac{1}{2}\,(t_{n+2} - t_n)\Bigg(\nu_n(W(t_{n+1}) - W(t_n)) \\ &\qquad -(1-\nu_n)\,(W(t_{n+2}) - W(t_{n+1}))\Bigg),\end{aligned}$$

where $\nu_n = (t_{n+2} - t_{n+1})/(t_{n+2} - t_n)$. Thus for the midpoint interval I_n^c is computed from two ΔI^f and the Brownian increments used to generate the fine path.

This approximation has weak convergence since

$$\left| \mathbb{E}\left[\int_0^T \widehat{S}(t)\,\mathrm{d}t - \int_0^T S(t)\,\mathrm{d}t \right] \right|$$

is bounded by Theorem 3.3.2.

Theorem 3.3.4. *The approximation* (3.6) *for continuous Asian payoffs has variance* $V_l = \mathcal{O}(h_l^2)$.

Proof. Integrating (3.5) gives

$$\begin{aligned} \mathbb{E}[(\overline{\widehat{S}^f} - \overline{\widehat{S}^c})^2] \;\leq\; 4 \Big(& \mathbb{E}[(\overline{\widehat{S}^f} - \overline{\widehat{S}^f_{KP}})^2] \;+\; \mathbb{E}[(\overline{\widehat{S}^c} - \overline{\widehat{S}^c_{KP}})^2] \\ & + \mathbb{E}[(\overline{\widehat{S}^f_{KP}} - \overline{S})^2] \;+\; \mathbb{E}[(\overline{\widehat{S}^c_{KP}} - \overline{S})^2] \Big), \end{aligned}$$

Using the Lipschitz bound and Theorems 3.3.1 and 3.3.2 we can conclude the proof. □

3.3.4.2 Treatment 2

The second treatment is the same as the first except that it omits the terms I_n^f and I_n^c and so the averages correspond to trapezoidal integration of the two interpolants, or alternatively they can be viewed as averages of the piecewise linear interpolants $\widehat{S}_{PL}(t)$.

$$\int_0^T \widehat{S}^f(t)\,\mathrm{d}t = \sum_{n=0}^{N+N_T-1} \tfrac{1}{2} h_n \left(\widehat{S}_n^f + \widehat{S}_{n+1}^{f-} \right). \tag{3.7}$$

Theorem 3.3.5. *The approximation* (3.7) *for continuous Asian options has variance* $V_l = \mathcal{O}(h_l^2)$.

Proof. The difference between the averages of the Brownian and piecewise linear interpolants is

$$T^{-1} \int_0^T \left(\widehat{S}(t) - \widehat{S}_{PL}(t) \right) \,\mathrm{d}t = T^{-1} \sum_n b_n \Delta I_n.$$

Since the I_n are independent $N(0,\ \frac{1}{12}h_n^3)$ variables and independent of b_n conditional on G, it follows that

$$\begin{aligned}\mathbb{E}\left[(\overline{\widehat{S}} - \overline{\widehat{S}_{PL}})^2\right] &= \tfrac{1}{12}T^{-2}\mathbb{E}\left[\sum_{n=0}^{N+N_T-1}\mathbb{E}\left[b_n^2 \mid G\right]\mathbb{E}\left[\ \Delta I_n^2 \mid G\right]\right] \\ &\leq \frac{1}{12}T^{-2}h^3\sqrt{\mathbb{E}\left[(N+N_T)^2\right]\mathbb{E}\left[\sup_n b_n^4\right]},\end{aligned}$$

and this is $\mathcal{O}(h^2)$ due to the finite bound for $\mathbb{E}[\sup_n b_n^4]$.

Since

$$\overline{\widehat{S}^f_{PL}} - \overline{\widehat{S}^c_{PL}} = (\overline{\widehat{S}^f} - \overline{\widehat{S}^c}) - (\overline{\widehat{S}^f} - \overline{\widehat{S}^f_{PL}}) + (\overline{\widehat{S}^c} - \overline{\widehat{S}^c_{PL}})$$

it follows that

$$\mathbb{E}[(\overline{\widehat{S}^f_{PL}} - \overline{\widehat{S}^c_{PL}})^2] \leq 3\left(\mathbb{E}[(\overline{\widehat{S}^f} - \overline{\widehat{S}^c})^2]\ +\ \mathbb{E}[(\overline{\widehat{S}^f} - \overline{\widehat{S}^f_{PL}})^2]\ +\ \mathbb{E}[(\overline{\widehat{S}^c} - \overline{\widehat{S}^c_{PL}})^2]\right).$$

The bounds on $\mathbb{E}[(\overline{\widehat{S}} - \overline{\widehat{S}_{PL}})^2]$ together with the bound on $\mathbb{E}[(\overline{\widehat{S}^f} - \overline{\widehat{S}^c})^2]$ from the proof of Theorem 3.3.4 prove that $\mathbb{E}[(\overline{\widehat{S}^f_{PL}} - \overline{\widehat{S}^c_{PL}})^2] = \mathcal{O}(h_l^2)$, and the result then follows from the assumed Lipschitz property of the payoff. □

3.3.5 Lookback options

The payoff of a lookback option is uniform Lipschitz in terms of the value of the underlying at maturity $S(T)$ and either the running minimum or the maximum of the underlying over the time horizon. We consider the case of the running minimum; the analysis for cases involving the maximum is similar.

$$P = \exp(-rT)\mathbb{E}\left[\left(S(T) - \min_{0\leq t\leq T} S(t)\right)\right]$$

For the fine path calculation, on level l, $\mathbb{T}_l = \mathbb{J} \cup \left\{\frac{iT}{2^l}\right\} = \{0 = t_0 < t_1 < \ldots < t_{n_l} = T\}$, $n_l = 2^l + N_T$ is the jump-adapted time grid. Define $\widehat{S}(t)$ to be the Brownian interpolant of $\widehat{S}_n$ within the fine grid $\mathbb{T}_l = \mathbb{J} \cup \left\{\frac{iT}{2^l}\right\}$, as in (3.4). The corresponding MC estimator constructed by the jump-adapted Milstein scheme with Brownian interpolation between grid points would be

$$\widehat{P} = \exp(-rT)\left(\widehat{S}(T) - \min_{0\leq t\leq T}\widehat{S}(t)\right).$$

Now we derive a method to calculate $\min_{0\le t\le T}\widehat{S}(t)$. Let $\mathcal{F}_l = \sigma\{N_t,\ 0 \le t \le T; W_1 = W(t_1), W_2, ..., W_{2^l+N_T}; Y_i, i \in \mathbb{J}\}$ where N_t is the Poisson process associated with the jump component, and $\widehat{S}_{i,min} = \min_{t_i\le t<t_{i+1}}\widehat{S}(t)$. The running minimum of $\widehat{S}$ can be computed by:

$$\mathbb{E}\left[\min_{0\le t\le T}\widehat{S}(t)\right] = \mathbb{E}\left[\mathbb{E}\left[\min_{0\le t\le T}\widehat{S}(t) \mid \mathcal{F}_l\right]\right] = \mathbb{E}\left[\mathbb{E}\left[\min_{0\le n\le n_l-1}\widehat{S}_{n,min} \mid \mathcal{F}_l\right]\right].$$

By the Markov property, we have for arbitrary $x > 0$,

$$P(\widehat{S}_{n,min} < x \mid \mathcal{F}_l) = P\left(\widehat{S}_{n,min} < x \mid \widehat{S}(t_n),\ \widehat{S}(t_{n+1}-)\right).$$

Thus by Lemma A.1.2 given $\mathcal{F}_l$ we can generate a sample of $\widehat{S}_{n,min}$:

$$\widehat{S}_{n,min} = \ \tfrac{1}{2}\left(\widehat{S}_n + \widehat{S}^-_{n+1} - \sqrt{\left(\widehat{S}^-_{n+1} - \widehat{S}_n\right)^2 - 2\,(b_n)^2\, h_n \log U_n}\ \right), \tag{3.8}$$

where U_n are independent uniform random variables on $[0,1]$.

Therefore, taking the minimum (3.8) over all timesteps gives a sample of the global minimum which is used to compute the fine path value $\widehat{P}_l$.

For the coarse path value $\widehat{P}^c_{l-1}$, define $\widehat{S}^c(t)$ to be the Brownian interpolant of $\widehat{S}^c_n$ within the coarse grid $\mathbb{T}_{l-1} = \mathbb{J}\cup\left\{\frac{iT}{2^{l-1}}\right\}$. Considering a particular timestep in the coarse path construction, we have two possible situations. If it does not contain one of the fine path discretisation times, and therefore corresponds exactly to one of the fine path timesteps, then it is treated in the same way as the fine path, using the same uniform random number U_n. If the coarse timestep contains midpoint $\zeta_i = \frac{(2i-1)T}{2^l}$, let $n(\zeta_i)$ be the index of grid point in $\mathbb{T}_l$ of the midpoint ζ_i. Using grid index $\mathbb{T}_l$, this coarse timestep is the midinterval $[t_{n-1}, t_{n+1}]$. We have to evaluate $\min_{[t_{n-1},t_{n+1}]}\widehat{S}^c(t)$ conditional on $\mathcal{F}_l$. since Brownian path value at the midpoint $\zeta_i = \frac{(2i-1)T}{2^l}$ (2.3) is known, $\min_{[t_{n-1},t_{n+1}]}\widehat{S}^c(t) \mid \mathcal{F}_l = \min_{t\in[t_{n-1},t_{n+1}]}\widehat{S}^c(t) \mid \widehat{S}^c(t_n)\ ,\widehat{S}^c(\zeta_i)\,, \widehat{S}^c(t_{n+1}-)$ by the Markov property. Hence we have to calculate the Brownian interpolant $\widehat{S}^c(\zeta_i)$ by (3.4) and then given $\mathcal{F}_l$, $\min_{t\in[t_{n-1},t_{n+1}]}\widehat{S}^c(t) = \min\left(\min_{[t_{n-1},t_n]}\widehat{S}^c(t), \min_{[t_n,t_{n+1}]}\widehat{S}^c(t)\right)$.

We have

$$\min_{[t_{n-1},t_n]} \widehat{S}^c(t) = \tfrac{1}{2}\left(\widehat{S}^c\left(t_{n-1}\right) + \widehat{S}^c(\zeta_i) - \sqrt{\left(\widehat{S}^c(\zeta_i) - \widehat{S}^c\left(t_n\right)\right)^2 - 2\left(b_n^c\right)^2 \left(\zeta_i - t_{n-1}\right)\log U_n^f}\right),$$

$$\min_{[t_n,t_{n+1}]} \widehat{S}^c(t) = \tfrac{1}{2}\left(\widehat{S}^c(\zeta_i) + \widehat{S}^c(t_{n+1}-) - \sqrt{\left(\widehat{S}^c(t_{n+1}-) - \widehat{S}^c\left(\zeta_i\right)\right)^2 - 2\left(b_{n+1}^c\right)^2 \left(t_{n+1} - \zeta_i\right)\log U_{n+1}^f}\right).$$

where $b_n^c = b_{n+1}^c \equiv b(\widehat{S}_n^c, t_n)$. $\min_{[t_{n-1},t_{n+1}]} \widehat{S}^c(t)$ on level l has exactly the same distribution as $\min_{[t_{n-1},t_{n+1}]} \widehat{S}^f(t)$ on level $l-1$, since they are both based on the same Brownian interpolation (3.4), therefore equality (2.5) is satisfied.

Theorem 3.3.6. *Suppose that a lookback payoff is a uniform Lipschitz function of $S(t)$ and $\min_{0\le t\le T} S(t)$, and on level l, the minima of the fine and coarse paths $\widehat{S}^f_{min} = \min_{0\le n\le n_l-1} \widehat{S}_{n,min}$ and $\widehat{S}^c_{min} = \min_{0\le n\le n_{l-1}-1} \widehat{S}_{n,min}$ are generated by the previous method. Then the multilevel correction term of a lookback option has $V_l = \mathcal{O}\left(h_l^2 \left|\log h_l\right|^2\right)$.*

Proof. We have

$$\begin{aligned}\left|\widehat{S}^f_{min} - \widehat{S}^c_{min}\right| &\le \max_{0\le n\le n_l-1}\left|\widehat{S}^f_{n,min} - \widehat{S}^c_{n,min}\right| \\ &\le \max_{0\le n\le n_l-1}\left|\widehat{S}^f_n - \widehat{S}^c_n\right| + \max_{0\le n\le n_l-1}\left|\widehat{D}^f_n - \widehat{D}^c_n\right|,\end{aligned} \tag{3.9}$$

where

$$\widehat{D}^f_n = \tfrac{1}{2}\sqrt{\left(\widehat{S}^{f-}_{n+1} - \widehat{S}^f_n\right)^2 - 2\left(b_n^f\right)^2 h_n \log U_n}$$

and $\widehat{D}^c_n$ is defined with $b_n^c = b_{n-1}^c$ for $n\left(\zeta_i\right),\ i = 1, 2, ..., 2^{l-1}$. Now, using the inequality

$|\,|x|-|y|\,| \leq |x-y|$, calculation gives

$$\begin{aligned}
& \left|\widehat{D}^f_n - \widehat{D}^c_n\right| \\
= & \frac{\left|(\widehat{D}^f_n)^2 - (\widehat{D}^c_n)^2\right|}{\widehat{D}^f_n + \widehat{D}^c_n} \\
\leq & \frac{\left|(\widehat{S}^{f-}_{n+1} - \widehat{S}^f_n)^2 - (\widehat{S}^{c-}_{n+1} - \widehat{S}^c_n)^2\right|}{4(\widehat{D}^f_n + \widehat{D}^c_n)} + \frac{|(b^f_n)^2 - (b^c_n)^2|\; h_n\,|\log U_n|}{2(\widehat{D}^f_n + \widehat{D}^c_n)} \\
\leq & \tfrac{1}{2}\left|\,|\widehat{S}^{f-}_{n+1} - \widehat{S}^f_n| - |\widehat{S}^{c-}_{n+1} - \widehat{S}^c_n|\,\right| + \tfrac{1}{\sqrt{2}}\left|\,|b^f_n| - |b^c_n|\,\right| \sqrt{h_n\,|\log U_n|} \\
\leq & \tfrac{1}{2}\left(\left|\widehat{S}^{f-}_{n+1} - \widehat{S}^{c-}_{n+1}\right| + \left|\widehat{S}^f_n - \widehat{S}^c_n\right|\right) + \tfrac{1}{\sqrt{2}}\,|b^f_n - b^c_n|\,\sqrt{h_n\,|\log U_n|}. \qquad (3.10)
\end{aligned}$$

Hence

$$\begin{aligned}
& \left|\widehat{S}^f_{min} - \widehat{S}^c_{min}\right|^2 \\
\leq & \frac{27}{4}\max_{0\leq n\leq n_l-1}\left|\widehat{S}^f_n - \widehat{S}^c_n\right|^2 + \frac{3}{4}\max_{0\leq n\leq n_l-1}\left|\widehat{S}^{f-}_n - \widehat{S}^{c-}_n\right|^2 + \frac{3}{2}h_l\max_{0\leq n\leq n_l-1}|b^f_n - b^c_n|^2\,|\log U_n|^2.
\end{aligned}$$

Since $\widehat{S}^{f-}_{n+1}$ and $\widehat{S}^{c-}_{n+1}$ can be seen as KP interpolants of $\widehat{S}_n$, i.e. $\widehat{S}^{f-}_{n+1} = \widehat{S}^f_{KP}(t_{n+1}-)$, by Theorem 3.3.1, we have

$$\begin{aligned}
\mathbb{E}\left[\max_{0\leq n\leq n_l-1}\left|\widehat{S}^{f-}_n - \widehat{S}^{c-}_n\right|^2\right] &= \mathcal{O}(h_l^2); \\
\mathbb{E}\left[\max_{0\leq n\leq n_l-1}\left|\widehat{S}^f_n - \widehat{S}^c_n\right|^2\right] &= \mathcal{O}(h_l^2).
\end{aligned}$$

For $n \neq n(\zeta_i)$, Assumption A1 implies $|b^f_n - b^c_n| \leq K_1|\widehat{S}^f_n - \widehat{S}^c_n|$; for n where $\exists i$ s.t. $\zeta_i = t_n$, the observation shows that

$$\begin{aligned}
|b^f_n - b^c_n|^2 &= |b^f_n - b^c_{n-1}| \\
&\leq 2\,|b^f_n - b^f_{n-1}|^2 + 2|b^f_{n-1} - b^c_{n-1}|^2 \\
&\leq 2K_1|\widehat{S}^f_n - \widehat{S}^f_{n-1}|^2 + 2K_1|\widehat{S}^f_{n-1} - \widehat{S}^c_{n-1}|^2.
\end{aligned}$$

In

$$\widehat{S}^f_n - \widehat{S}^f_{n-1} = a_{n-1}h_{n-1} + b_{n-1}\Delta W_{n-1} + \frac{1}{2}\,b'_{n-1}b_{n-1}\,(\Delta W^2_{n-1} - h_{n-1}),$$

since $b_{n-1}\Delta W_{n-1}$ is the dominant term, we only analyse it. By Jensen's inequality,

(G denotes the filtration $\sigma\{N_t,\ 0 \le t \le T\}$)

$$\begin{aligned}
\mathbb{E}\left[\max_{1\le n\le n_l-1} |b_{n-1}\Delta W_{n-1}|^2\right] &\le \mathbb{E}\left[\max_{1\le n\le n_l-1} b_{n-1}^4\right]^{1/2} \mathbb{E}\left[\mathbb{E}\left[\max_{1\le n\le n_l-1} |\Delta W_{n-1}|^4 \mid G\right]\right]^{1/2} \\
&\le K\mathbb{E}\left[h_l^2 \left|\log\left(N_T + h_l^{-1}\right)\right|^2\right]^{1/2} \\
&\le K\mathbb{E}\left[h_l^2 \left(\log\left(N_T+1\right) + |\log h_l|\right)^2\right]^{1/2} \\
&= \mathcal{O}\left(h_l \left|\log h_l\right|\right).
\end{aligned}$$

The following results come from extreme value theory which determine the limiting distribution of the maximum of a large set of i.i.d. random variables; cf. [21].

Lemma 3.3.7. *If $U_n, n = 1, \ldots, N$ are independent samples from a uniform distribution on the unit interval $[0,1]$, then for any positive integer m*

$$\mathbb{E}\left[\max_{1\le n\le N} |\log U_n|^m\right] = \mathcal{O}((\log N)^m), \ \text{as } N \to \infty. \tag{3.11}$$

By Lemma 3.3.7,

$$\mathbb{E}\left[\max_{0\le n\le n_l-1} |\log U_n|^2\right] = \mathcal{O}\left(|\log h_l|^2\right).$$

Since U_n are independent of $|b_n^f - b_n^c|^2|$, and the payoff is a Lipschitz function of $\widehat{S}_{min}$, we conclude that

$$V_l = \mathcal{O}\left(h_l^2 \left|\log h_l\right|^2\right).$$

□

For the weak convergence of the scheme, we have

Theorem 3.3.8. *The Brownian interpolated approximation for a lookback option $\widehat{P} = \exp(-rT)\left(\widehat{S}(T) - \min_{0\le t\le T} \widehat{S}(t)\right)$ satisfies that $\left|\mathbb{E}[\widehat{P}_l^f - P]\right| = \mathcal{O}(h\log h)$.*

Proof. Note that

$$\left|\mathbb{E}\left[\min_{0\le t\le T} \widehat{S}(t) - \min_{0\le t\le T} S(t)\right]\right| \le \mathbb{E}\left[\sup_{[0,T]} \left|\widehat{S}(t) - S(t)\right|\right] = \mathcal{O}(h\log h).$$

where the last inequality comes from the $n = 1$ case of Theorem 3.3.1 and Theorem 3.3.2. □

3.3.6 Extreme paths

The analysis of the barrier and digital payoffs will use an extreme path argument that certain "extreme" paths make a negligible contribution to the overall expectation. This argument will be employed in this chapter based on the following two lemmas taken from [34].

Lemma 3.3.9. *If X_l is a scalar random variable defined on level l of the multilevel analysis, and for each positive integer m, $\mathbb{E}[\,|X_l|^m] \leq C_m$ is bounded, then, for any $\delta > 0$,*

$$\mathbb{P}[\,|X_l| > h_l^{-\delta}] = o(h_l^p), \qquad \forall p > 0.$$

Proof. It follows from Corollary A.2.8 that

$$\begin{aligned} \mathbb{P}[\,|X_l| > h_l^{-\delta}] &= \mathbb{P}[\,|X_l|^m > h_l^{-m\delta}] \\ &\leq h_l^{m\delta}\mathbb{E}[\,|X_l|^m] \\ &\leq h_l^{m\delta}C_m. \end{aligned}$$

By choosing $m > p/\delta$, $\forall p > 0$,

$$\mathbb{P}[\,|X_l| > h_l^{-\delta}] = o(h_l^p).$$

□

Lemma 3.3.10. *If Y_l is a scalar random variable on level l, $\mathbb{E}[\,Y_l^2]$ is uniformly bounded, and for each $p_1 > 0$, the indicator function $\mathbf{1}_{E_l}$ on level l (which takes value 1 or 0 depending whether or not a path lies within some set E_l) satisfies*

$$\mathbb{E}[\mathbf{1}_{E_l}] = o(h_l p_1),$$

then for each $p > 0$,

$$\mathbb{E}[\,|Y_l|\,\mathbf{1}_{E_l}] = o(h_l^p).$$

Proof. By Cauchy-Schwartz inequality,

$$\mathbb{E}[\,|Y_l|\mathbf{1}_{E_l}] \leq \left(\mathbb{E}[\,Y_l^2]\right)^{1/2} \left(\mathbb{E}[\,\mathbf{1}_{E_l}]\right)^{1/2}.$$

By choosing $p_1 = 2p$ we conclude. □

In the proofs of the main analysis, Lemma 3.3.9 will be used to establish the pre-conditions for Lemma 3.3.10, from which it can be concluded, by choosing p sufficiently large, that the contribution of the extreme paths is negligible compared to the paths that are not extreme.

The following lemma explains why the set of extreme paths has a small probability. There is some difference between the definition here and the one in [34]. Here we have to emphasize that for the extreme path, the Brownian increments exceeds certain level of the square root of its corresponding timestep. This is required for the analysis of digital options.

Lemma 3.3.11. *In a jump-adapted time grid $\mathbb{T}_l = \{0 = t_0 < t_1 < t_2 < \ldots < t_M = T\}$ with the maximum timestep h and prescribed jump times set $\mathbb{J} = \{\tau_1, \tau_2, \ldots, \tau_m\}$, for any $\gamma > 0$, the probability that a Brownian path $W(t)$, its increments $\Delta W_n \equiv W(t_{n+1}) - W(t_n)$, and the corresponding SDE solution $S(t)$ and its fine and coarse path approximations $\widehat{S}^f_n$ and $\widehat{S}^c_n$ satisfy any of the following extreme conditions*

$$\begin{aligned} \exists n,\ \max(|S(t_n)|,\ |\widehat{S}^f_n|,\ |\widehat{S}^c_n|) &> h^{-\gamma} \\ \exists n,\ \max\left(\left|\widehat{S}^{f-}_n - \widehat{S}^{c-}_n\right|, |S(t_n)-\widehat{S}^c_n|,\ |S(t_n)-\widehat{S}^f_n|,\ |\widehat{S}^f_n-\widehat{S}^c_n|\right) &> h^{1-\gamma} \\ \exists n,\ |\Delta W_n| &> h_n^{1/2-\gamma} \end{aligned}$$

where $h_n = t_{n+1} - t_n$, is $o(h^p)$ for all $p>0$.

Furthermore, there exist constants c_1, c_2, c_3, c_4 such that if none of these conditions is satisfied, and $\gamma < \frac{1}{2}$, then

$$\forall n,\ |\widehat{S}^{f-}_n - \widehat{S}^f_{n-1}| \ \le\ c_1\, h_n^{1/2-2\gamma} \tag{3.12}$$

$$\forall n,\ |b^f_n - b^f_{n-1}| \ \le\ c_2\, h_n^{1/2-2\gamma} \tag{3.13}$$

$$\forall n,\ \max(|b^f_n|, |b^c_n|) \ \le\ c_3\, h^{-\gamma} \tag{3.14}$$

$$\forall n,\ |b^f_n - b^c_n| \ \le\ c_4\, h^{1/2-2\gamma} \tag{3.15}$$

where $\widehat{S}^{f-}_n = \widehat{S}^f(t_n-)$, $b^c_n = b^c_{n-1}$ if $\exists i$ s.t. $\zeta_i = t_n$ (i.e. $[t_{n-1}, t_{n+1}]$ is a midpoint interval).

Proof. Note that $\widehat{S}^{f-}_n$ and $\widehat{S}^{c-}_n$ can be seen as a KP interpolant, i.e. $\widehat{S}^{f-}_n = \widehat{S}^f_{KP}(t_n-)$. Due to Theorems 3.2.1 and 3.3.1 and Lemma 3.3.9, the probability of event set which satisfies the first two extreme conditions is $o(h^p)$ for all $p>0$. In addition,

$$\begin{aligned}\mathbb{P}\left(\exists n, |\Delta W_n| > h_n^{1/2-\gamma}\right) &= \mathbb{E}\left[\mathbb{P}\left(\max_{0\le n\le N+N_T-1} |\Delta W_n| > h_n^{1/2-\gamma} \mid G\right)\right] \\ &\le \mathbb{E}\left[\sum_{0\le n\le N+N_T-1} \mathbb{P}\left(|\Delta W_n| > h_n^{1/2-\gamma} \mid G\right)\right],\end{aligned}$$

Due to Lemma 3.3.9, $\mathbb{P}\left(|\Delta W_n| > h_n^{1/2-\gamma} \mid G\right) \le C_m h^{m\gamma}$ for all $m>0$. By choosing $m > (p+1)/\gamma$, we have $\mathbb{P}\left(\sup_n |\Delta W_n| > h_n^{1/2-\gamma}\right) = o(h^p)$ for all $p>0$.

If none of the extreme conditions is satisfied, then the linear growth bound gives

$$|\widehat{S}^{f-}_{n+1} - \widehat{S}^f_n| < K_2 h_n(1+h_n^{-\gamma}) + K_2(1+h_n^{-\gamma})h_n^{1/2-\gamma} + \tfrac{1}{2}K_2(1+h_n^{-\gamma})(h_n^{1-2\gamma}+h_n).$$

Since $h_n \le T$, we can then construct a constant c_1 so that (3.12) is satisfied.

(3.13) follows as a consequence of Assumptions A1 and A3. (3.14) is obtained from Assumption A2 and the bound on $|\widehat{S}^f_n|$ and $|\widehat{S}^c_n|$.

If $n \ne n(\zeta_i)$, i.e. t_n is not a midpoint, the bound in (3.15) follows from Assumption A1 and the bound on $|\widehat{S}^f_n - \widehat{S}^c_n|$; for n where $\exists i$ s.t. $\zeta_i = t_n$, then

$$|b^f_n - b^c_n| \;=\; |b^f_n - b^c_{n-1}| \;\le\; |b^f_n - b^f_{n-1}| + |b^f_{n-1} - b^c_{n-1}|.$$

and it then follows from (3.13) and the bound for $|\widehat{S}^f_{n-1} - \widehat{S}^c_{n-1}|$. □

3.3.7 Barrier options

We consider a down-and-out call barrier option for which the discounted payoff is

$$P = \exp(-rT)\,(S(T)-K)^+\,\mathbf{1}_{\{M_T>B\}},$$

where $M_T = \min_{0\le t\le T} S(t)$.

A jump-adapted Milstein with a Brownian interpolant approximation gives

$$\widehat{P} = \exp(-rT)\,(\widehat{S}(T)-K)^+\,\mathbf{1}_{\left\{\widehat{M}_T>B\right\}},$$

where $\widehat{M}_T = \min_{0\le t\le T} \widehat{S}(t)$.

Let $f(S) = (S-K)^+$. We first prove the weak convergence of the estimator.

Theorem 3.3.12. *Provided that the* $\inf_{0\le t\le T} S(t)$ *has bounded density in the neighbourhood of* B*, the Brownian interpolated approximation for a barrier option has*

$$\left|\mathbb{E}\left[\widehat{P}_l^f - P\right]\right| = o(h_l^{1-\delta})$$

for any $\delta > 0$.

Proof. The difference between the approximated and true payoff can be represented as

$$\mathbb{E}\left[\widehat{P}_l^f - P\right] = \frac{1}{2}\mathbb{E}\Bigg[\left(f\left(\widehat{S}_{n_T}^f\right) - f(S_T)\right)\left(\mathbf{1}_{\{M_T>B\}} + \mathbf{1}_{\{\widehat{M}_T>B\}}\right) \tag{3.16}$$
$$+\left(f\left(\widehat{S}_{n_T}^f\right) + f(S_T)\right)\left(\mathbf{1}_{\{M_T>B\}} - \mathbf{1}_{\{\widehat{M}_T>B\}}\right)\Bigg].$$

This is derived by using the equality

$$f_1\, g_1 - f_2\, g_2 = \tfrac{1}{2}(f_1 - f_2)(g_1 + g_2) + \tfrac{1}{2}(f_1 + f_2)(g_1 - g_2). \tag{3.17}$$

The first part of the RHS of (3.16) is $\mathcal{O}(h_l)$ by Theorem 3.3.1.

Note that

$$\left|\mathbf{1}_{\{M_T>B\}} - \mathbf{1}_{\{\widehat{M}_T>B\}}\right| = 1 \text{ or } 0.$$

For sufficiently small h_l, "$= 1$" holds only if either M_T is close to the barrier or the difference between continuous and discretely monitored maximum $\left|M_T - \widehat{M}_T\right|$ is large. More precisely,

$$\left\{\left|\mathbf{1}_{\{M_T>B\}} - \mathbf{1}_{\{\widehat{M}_T>B\}}\right| = 1\right\} \subset F_l \cup H,$$

where for a $\gamma > 0$ to be determined,

$$F_l \; : = \left\{|M_T - B| \le h_l^{1-2\gamma}\right\},$$
$$H \; : = \left\{\left|M_T - \widehat{M}_T\right| > h_l^{1-\gamma}\right\}.$$

Due to the bounded density of M_T in the neighbourhood of B, $\mathbb{P}(F_l) = \mathcal{O}\left(h_l^{1-2\gamma}\right)$.

We define the extreme path set E_l to be the paths satisfying either the conditions of Lemma 3.3.11 or belonging to H. We shall explain the reason for considering paths belonging to H to be extreme paths. Note that $\left|M_T - \widehat{M}_T\right| \le \sup_{[0,T]}\left|\widehat{S}(t) - S(t)\right|$. Theorems 3.3.1 and 3.3.2 imply a uniform bound for

$$\mathbb{E}\left[h_l^{-m}\sup_{0\le t\le T}\left|\widehat{S}^f(t) - S(t)\right|^m\right],$$

for all l, and it follows that $\mathbb{P}\left(\left|M_T - \widehat{M}_T\right| > h_l^{1-\gamma}\right) = o(h_l^p)$ for all $p>0$. By the linear growth property of f and Theorems 3.3.1, $\mathbb{E}\left[f\left(\widehat{S}^f_{n_T}\right)^4\right]$ are uniformly bounded for all l. Hence by Lemma 3.3.10, for all $p>0$

$$\mathbb{E}\left[\left(f\left(\widehat{S}^f_{n_T}\right) + f\left(S_T\right)\right)(\mathbf{1}_{\{M_T>B\}} - \mathbf{1}_{\{\widehat{M}_T>B\}})\mathbf{1}_{E_l}\right] = o(h_l^p).$$

Hence we can focus on the contribution from E_l^c. In E_l^c, $f\left(\widehat{S}^f_{n_T}\right) + f\left(S_T\right)$ is bounded by $h^{-\gamma}$ by definition. The second part of the RHS of (3.16) is bounded by

$$h^{-\gamma}\mathbb{E}\left[\left|\mathbf{1}_{\{M_T>B\}} - \mathbf{1}_{\{\widehat{M}_T>B\}}\right|\right].$$

Hence

$$\begin{aligned}
& \mathbb{E}\left[\left|\mathbf{1}_{\{M_T>B\}} - \mathbf{1}_{\{\widehat{M}_T>B\}}\right|\mathbf{1}_{E_l^c}\right] \\
= \; & \mathbb{E}[\left|\mathbf{1}_{\{M_T>B\}} - \mathbf{1}_{\{\widehat{M}_T>B\}}\right|\mathbf{1}_{F_l}\mathbf{1}_{E_l^c}] + \mathbb{E}[\left|\mathbf{1}_{\{M_T>B\}} - \mathbf{1}_{\{\widehat{M}_T>B\}}\right|\mathbf{1}_{F_l^c}\mathbf{1}_{E_l^c}] \\
= \; & \mathcal{O}(h_l^{1-2\gamma}).
\end{aligned}$$

Thus

$$\left|\mathbb{E}\left[\widehat{P}^f_l - P\right]\right| = \mathcal{O}(h_l^{1-3\gamma}).$$

The result follows by taking $\gamma = \delta/3$ and $p \geq 1$. □

Taking the standard variance reduction simulation approach to continuously monitored barrier crossings (see section 6.4 in [39]), we can construct an estimator using conditional expectation. For the fine path

$$\widehat{P}^f_l = f(\widehat{S}^f_{n_T}) \prod_{n=0}^{n_T-1}(1-\widehat{p}^f_n),$$

where

$$\widehat{p}^f_n = \exp\left(\frac{-2\,(\widehat{S}^f_n - B)^+(\widehat{S}^{f-}_{n+1} - B)^+}{(b^f_n)^2\,h_l}\right).$$

To prove the convergence of variance for this estimator, there are some gaps if we follow the approach in [34]. The problem is that we can not have a good estimate of the difference between crossing probabilities on fine and coarse level since the presence of jumps causes arbitrarily small timesteps. Here, instead, we prove the convergence of variance result for the estimator without taking the conditional expectation to calculate the crossing probability. Since the variance is reduced by using conditional expectation, as a corollary we conclude the convergence of variance of the ML estimator using conditional expectation.

Theorem 3.3.13. *Provided* $\inf_{0\leq t\leq T} S(t)$ *has a bounded density in the neighbourhood of* B*, then the multilevel estimator for a down-and-out barrier option has variance* $V_l = o(h_l^{1-\delta})$ *for any* $\delta > 0$*.*

Proof. The proof will use a similar approach to the weak convergence analysis.

The ML estimator for barrier option is

$$\begin{aligned} \widehat{P}_l^f &= f(\widehat{S}_{n_T}^f)\mathbf{1}_{\left\{\widehat{S}_{min}^f > B\right\}}; \\ \widehat{P}_{l-1}^c &= f(\widehat{S}_{n_T}^c)\mathbf{1}_{\left\{\widehat{S}_{min}^c > B\right\}}. \end{aligned}$$

First we define the extreme path set E_l to be the path which fulfills any of the conditions of Lemma 3.3.11 for $\gamma < \frac{1}{2}$ or either of the following conditions:

$$\left|M_T - \widehat{S}_{min}^f\right| > h_l^{1-\gamma},$$

$$\sup_n |\log U_n| > h_l^{-\gamma}.$$

The probability of the last condition being satisfied is $o(h^p)$ for all $p > 0$ because of the inequality

$$\begin{aligned} \mathbb{P}\left(\sup_n |\log U_n| > h_l^{-\gamma}\right) &= \mathbb{E}\left[\mathbb{P}\left(\max_{0\leq n\leq N+N_T-1} |\log U_n| > h_l^{-\gamma} \mid G\right)\right] \\ &\leq \mathbb{E}\left[\sum_{0\leq n\leq N+N_T-1} \mathbb{P}\left(|\log U_n| > h_l^{-\gamma}\right)\right]. \end{aligned}$$

together with Lemma 3.3.9 and the observation that $\mathbb{E}[\,|\log U_n|^m] = m!$. In the proof of the last theorem we have considered the probability of

$$\left|M_T - \widehat{S}_{min}^f\right| > h_l^{1-\gamma},$$

being satisfied is $o(h^p)$ for all $p > 0$. Hence it follows that $\mathbb{P}(E_l) = o(h_l^p)$ for all $p > 0$. By the linear growth property of f and Theorems 3.3.1 and 3.2.1, $\mathbb{E}\left[\left(\widehat{P}_l^f\right)^4\right]$, $\mathbb{E}\left[\left(\widehat{P}_{l-1}^c\right)^4\right]$ and also $\mathbb{E}\left[(\widehat{P}_l^f - \widehat{P}_{l-1}^c)^4\right]$ are uniformly bounded for all l. Hence by Lemma (3.3.10), for all $p > 0$

$$\mathbb{E}\left[(\widehat{P}_l^f - \widehat{P}_{l-1}^c)^2\mathbf{1}_{E_l}\right] = o(h_l^p).$$

Now we focus on the contribution from non-extreme paths. We again define the paths where M_T is close to the barrier by

$$F_l = \left\{|M_T - B| \leq h_l^{1-3\gamma}\right\}.$$

Note that in the proof of the multilevel estimator for lookback option, (3.9) and (3.10) show that

$$\begin{aligned}
\left|\widehat{S}^f_{min} - \widehat{S}^c_{min}\right| &\leq \max_{0\leq n\leq n_l-1}\left|\widehat{S}^f_n - \widehat{S}^c_n\right| + \max_{0\leq n\leq n_l-1}\left|\widehat{D}^f_n - \widehat{D}^c_n\right| \\
&\leq \frac{3}{2}\max_{0\leq n\leq n_l-1}\left|\widehat{S}^f_n - \widehat{S}^c_n\right| + \frac{1}{2}\max_{0\leq n\leq n_l-1}\left|\widehat{S}^{f-}_{n+1} - \widehat{S}^{c-}_{n+1}\right| \\
&\quad +\tfrac{1}{\sqrt{2}}\max_{0\leq n\leq n_l-1}\left|b^f_n - b^c_n\right|\sqrt{|\log U_n|}.
\end{aligned}$$

The definition of extreme paths, inequality (3.12) and the bound on $|\log U_n|$ give $\left|\widehat{S}^f_{min} - \widehat{S}^c_{min}\right|\mathbf{1}_{E^c_l} = \mathcal{O}(h_l^{1-\frac{5}{2}\gamma})$.

In F^c_l, for sufficiently small h_l, non-extreme paths where $M_T < B - h_l^{1-3\gamma}$ satisfy that $\widehat{S}^f_{min},\ \widehat{S}^c_{min} < B$. Therefore for those paths $\widehat{P}^f_l - \widehat{P}^c_{l-1} = 0$. For non-extreme paths where $M_T > B + h_l^{1-3\gamma}$ we have $\widehat{P}^f_l - \widehat{P}^c_{l-1} = f(\widehat{S}^f_{n_T}) - f(\widehat{S}^c_{n_T}) = \mathcal{O}(h_l^{1-\gamma})$. Hence

$$\mathbb{E}[(\widehat{P}^f_l - \widehat{P}^c_{l-1})^2\mathbf{1}_{F^c_l}\mathbf{1}_{E^c_l}] = \mathcal{O}(h_l^{2-2\gamma}).$$

In F_l, by definition of extreme paths we have $(\widehat{P}^f_l - \widehat{P}^c_{l-1})^2\mathbf{1}_{F_l}\mathbf{1}_{E^c_l} = \mathcal{O}(h_l^{-2\gamma})$. Due to the assumption that $\inf_{0\leq t\leq T} S(t)$ has a bounded density in the neighbourhood of B,

$$\mathbb{E}[(\widehat{P}^f_l - \widehat{P}^c_{l-1})^2\mathbf{1}_{F_l}\mathbf{1}_{E^c_l}] = \mathcal{O}(h_l^{1-5\gamma})$$

Therefore for sufficiently small h_l,

$$\mathbb{E}\left[(\widehat{P}^f_l - \widehat{P}^c_{l-1})^2\mathbf{1}_{E^c_l}\right] = \mathcal{O}(h_l^{1-5\gamma}).$$

The results follows by choosing $\gamma < \min(\frac{1}{2}, \delta/5)$ and $p \geq 1$. □

3.3.8 Digital options

A digital option has a payoff which is a discontinuous function of the value of the underlying asset at maturity. Here we consider the European digital call option

$$P = \mathbf{1}_{\{S(T)>K\}}.$$

Following the previous convention, we can discretise the path using a jump-adapted Milstein scheme, and then we have a simple estimator:

$$\widehat{P}_l^{naive} := \mathbf{1}_{\left\{\widehat{S}^f_{n_T} > K\right\}}.$$

where n_T is the index of the fine grid at the expiry date.

In order to reduce the ML variance we use conditional expectation taken at the penultimate timestep. Note that

$$\mathbb{E}[\widehat{P}_l^{naive}] = \mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\left\{\widehat{S}^f_{n_T} > K\right\}} \mid \widehat{S}^f_{n_T-1}\right]\right]$$

where $\widehat{S}^f_{n_T-1}$is the value of the fine path approximation one timestep before T. The idea is to use $\widehat{P}_l^f = \mathbb{E}\left[\mathbf{1}_{\left\{\widehat{S}^f_{n_T} > K\right\}} \mid \widehat{S}^f_{n_T-1}\right]$ as the estimator instead of $\widehat{P}_l^{naive}$. If we approximate the process within the last timestep as a diffusion process with constant drift and volatility as in (3.4), we have an analytic expression for $\widehat{P}_l^f$:

$$\widehat{P}_l^f = \Phi\left(\frac{\widehat{S}^f_{n_T-1} + a^f_{n_T-1} h - K}{\left|b^f_{n_T-1}\right| \sqrt{h}}\right).$$

The expressions of $\widehat{P}_l^f$ in a jump diffusion model are in Section 2.4.2.5 of Chapter 2. This approach also brings extra benefits to smooth the payoff for computing Greeks.

The weak convergence of the approximation is guaranteed by the following theorem:

Theorem 3.3.14. *Provided $S(T)$ has a bounded density in the neighborhood of K, then the conditional expectation estimator for a digital option satisfies $\left|\mathbb{E}[\widehat{P}_l^f - P]\right| = o(h_l^{1-\delta})$ for any $\delta > 0$.*

Proof. The deviation we are about to analyse is

$$\mathbb{E}[\widehat{P}_l^f - P] = \mathbb{E}[\mathbf{1}_{\left\{\widehat{S}^f_{n_T} > K\right\}} - \mathbf{1}_{\{S(T) > K\}}].$$

We follow the approach in the barrier case- first we define extreme paths to satisfy the conditions in Lemma 3.3.11.

To bound the contribution from the paths ending around K, we define the set

$$F_l := \left\{|S_T - K| \le h_l^{1-2\gamma}\right\}.$$

For an extreme path in E_l, $\mathbb{E}[(\widehat{P}_l^f - P)1_{E_l}] = o(h_l^p)$, for all $p > 0$.

For a non-extreme path in F_l^c, $\left|S_T - \widehat{S}^f_{n_T}\right| \leq h_l^{1-\gamma}$. Thus for sufficiently small h_l, $\left|\widehat{S}^f_{n_T} - K\right| > h_l^{1-2\gamma}$ and the contribution of those paths

$$\mathbb{E}[\left(\widehat{P}^f_l - P\right) \mathbf{1}_{F_l^c} \mathbf{1}_{E_l^c}] = 0.$$

For a non-extreme path in F_l, the contribution of those paths to the $\mathbb{E}[\widehat{P}^f_l - P]$ is $\mathcal{O}(h_l^{1-2\gamma})$ due to the bounded density of $S(T)$ in the neighborhood of K, i.e.

$$\mathbb{E}[\left(\widehat{P}^f_l - P\right) \mathbf{1}_{F_l} \mathbf{1}_{E_l^c}] = \mathcal{O}(h_l^{1-2\gamma}).$$

Hence by choosing $\gamma < \min(\frac{1}{2}, \delta/3)$ and a $p \geq 1$ we conclude that

$$\left|\mathbb{E}[\widehat{P}^f_l - P]\right| = o(h_l^{1-\delta}).$$

□

A bound on the variance of the multilevel estimator is given by the following result:

Theorem 3.3.15. *Provided $b(K,T) \neq 0$, and $S(t)$ has a bounded density in the neighborhood of K, then the multilevel estimator for a digital option has variance $V_l = o(h_l^{3/2-\delta})$ for any $\delta > 0$.*

Proof. First we start with decomposing the variance into three parts

$$\mathbb{V}[\widehat{P}^f_l - \widehat{P}^c_{l-1}] \;\leq\; \mathbb{E}[(\widehat{P}^f_l - \widehat{P}^c_{l-1})^2 \mathbf{1}_{A_1}] + \mathbb{E}[(\widehat{P}^f_l - \widehat{P}^c_{l-1})^2 \mathbf{1}_{A_2}] + \mathbb{E}[(\widehat{P}^f_l - \widehat{P}^c_{l-1})^2 \mathbf{1}_{A_3}],$$

where A_i denotes the event sets of three cases, that is

$$\begin{aligned} A_1 &= 1_{\{J \leq T-2h\}}; \\ A_2 &= 1_{\{T-h < J \leq T\}}; \\ A_2 &= 1_{\{T-2h < J < T-h\}}. \end{aligned}$$

As in the proof of Theorem 3.3.6, we can then deal with the cases of extreme and non-extreme path.

For the case 1 and 3, the analysis is the same as in [34] despite the fact that the timestep is no longer uniform, which does no harm to the original argument.

For the case 2, we again use E_l to represent the set of extreme paths which satisfy one or more of the conditions in Lemma 3.3.11, with strictly positive $\gamma < \frac{1}{4}$ to be chosen later. $\mathbb{E}[(\widehat{P}_l^f - \widehat{P}_{l-1}^c)^2 \mathbf{1}_{E_l}]$ is again $o(h_l^p)$ for all $p > 0$. So we can concentrate on the contribution from non-extreme paths:

$$\mathbb{E}[(\widehat{P}_l^f - \widehat{P}_{l-1}^c)^2 \mathbf{1}_{A_2} \mathbf{1}_{E_l^c}].$$

For a non-extreme path, by Assumption A1 and (3.13), with $\gamma < \frac{1}{4}$.

Since $|\Phi(x_1) - \Phi(x_2)| \le |x_1 - x_2|$,

$$\begin{aligned}
\left|\widehat{P}_l^f - \widehat{P}_{l-1}^c\right| &= \left| \Phi\left(\frac{\widehat{S}_{n_T-1}^f + a_{n_T-1}^f h_j - K}{\left|b_{n_T-1}^f\right| \sqrt{h_j}} \right) - \Phi\left(\frac{\widehat{S}_{n_T-1}^c + a_{n_T-1}^c h_j - K}{\left|b_{n_T-1}^c\right| \sqrt{h_j}} \right) \right| \\
&\le \left| \frac{\widehat{S}_{n_T-1}^f + a_{n_T-1}^f h_j - K}{\left|b_{n_T-1}^f\right| \sqrt{h_j}} - \frac{\widehat{S}_{n_T-1}^c + a_{n_T-1}^c h_j - K}{\left|b_{n_T-1}^c\right| \sqrt{h_j}} \right|.
\end{aligned}$$

Let $s = h_j = T - J$. By the definition of the Milstein scheme,

$$\begin{aligned}
&\frac{\widehat{S}_{n_T-1}^f + a_{n_T-1}^f h_j - K}{\left|b_{n_T-1}^f\right|} \\
&\quad = \frac{\widehat{S}_{n_T}^f - K}{\left|b_{n_T-1}^f\right|} - \frac{b_{n_T-1}^f}{\left|b_{n_T-1}^f\right|} \left(\Delta W_s + \frac{1}{2} (b')_{n_T-1}^f \left(\Delta W_s^2 - s\right) \right),
\end{aligned}$$

where $\Delta W_s = \Delta W_{n_T-1}$ is the Brownian increment of the last timestep. We have a similar expression for the coarse estimator. Hence

$$\begin{aligned}
\left|\widehat{P}_l^f - \widehat{P}_{l-1}^c\right| \le& \left| \frac{\widehat{S}_{n_T}^f - K}{\left|b_{n_T-1}^f\right|} - \frac{\widehat{S}_{n_T}^c - K}{\left|b_{n_T-1}^c\right|} \right| \frac{1}{\sqrt{s}} + \frac{1}{\sqrt{s}} \left| \frac{b_{n_T-1}^f}{\left|b_{n_T-1}^f\right|} \left(\Delta W_s + \frac{1}{2} (b')_{n_T-1}^f \left(\Delta W_s^2 - s\right) \right) \right. \\
&\left. - \frac{b_{n_T-1}^c}{\left|b_{n_T-1}^c\right|} \left(\Delta W_s + \frac{1}{2} (b')_{n_T-1}^c \left(\Delta W_s^2 - s\right) \right) \right|.
\end{aligned}$$

Using the identity (3.17) we obtain

$$\begin{aligned}
\frac{\widehat{S}_{n_T}^f - K}{\left|b_{n_T-1}^f\right|} - \frac{\widehat{S}_{n_T}^c - K}{\left|b_{n_T-1}^c\right|} &= \frac{1}{2} \left(\widehat{S}_{n_T}^f - \widehat{S}_{n_T}^c\right) \left(\frac{1}{|b_{n_T-1}^f|} + \frac{1}{|b_{n_T-2}^c|} \right) \\
&\quad + \frac{1}{2} \left(\widehat{S}_{n_T}^f + \widehat{S}_{n_T}^c - 2K\right) \left(\frac{|b_{n_T-2}^c| - |b_{n_T-1}^f|}{|b_{n_T-1}^f| \, |b_{n_T-2}^c|} \right).
\end{aligned}$$

If follows from the bounds provided in Lemma 3.3.11 that

$$\frac{\widehat{S}^f_{n_T} - K}{\left|b^f_{n_T-1}\right|} - \frac{\widehat{S}^c_{n_T} - K}{\left|b^c_{n_T-1}\right|} \le C_1 h^{1-5\gamma}.$$

Since

$$b^f_{n_T-1} - b(K,T) = (b^f_{n_T-1} - b^f_{n_T}) + (b^f_{n_T} - b(K,T)),$$

using Assumption A1 and (3.13) with $\gamma < \frac{1}{4}$ we can conclude that for sufficiently small h_l, $|b^f_{n_T-1} - b(K,T)| < \frac{1}{2}|b(K,T)|$. In particular $b^f_{n_T-1}$ is non-zero and of the same sign as $b(K,T)$. So is $b^c_{n_T-1}$ and therefore

$$\begin{aligned}
&\left| \frac{b^f_{n_T-1}}{\left|b^f_{n_T-1}\right|} \left(\Delta W_s + \frac{1}{2}(b')^f_{n_T-1}\left(\Delta W_s^2 - s\right)\right) - \frac{b^c_{n_T-1}}{\left|b^c_{n_T-1}\right|} \left(\Delta W_s + \frac{1}{2}(b')^c_{n_T-1}\left(\Delta W_s^2 - s\right)\right) \right| \\
= \; & \frac{1}{2}\left|(b')^f_{n_T-1} - (b')^c_{n_T-1}\right| \left|\Delta W_s^2 - s\right| \\
\le \; & \frac{1}{2} K_1 s^{1-2\gamma}.
\end{aligned}$$

The last inequality is by Assumption A2 and Lemma 3.3.11. In sum,

$$\left|\widehat{P}^f_l - \widehat{P}^c_{l-1}\right| \le C_1 h^{1-5\gamma} \frac{1}{\sqrt{s}} + \frac{1}{2} K_1 s^{1/2-2\gamma}.$$

On the other hand it is obvious that

$$\left|\widehat{P}^f_l - \widehat{P}^c_{l-1}\right| \le 1.$$

Looking backwards, the last jump can be seen as the first jump. If we take the expectation conditional on $N_t = n$, n jump times $\tau_1, \ldots, \tau_n \equiv T - s$ have the same distribution as n ordered independent uniform $[0,T]$ random variables; thus $s, T - \tau_{n-1}, \ldots, T - \tau_1$ have the same distribution as n ordered uniform $[0,T]$ random variables as well. Therefore s satisfies an exponential distribution with parameter λ, namely $s \sim \text{Exp}(\lambda)$. Using the two estimates in $[0, \varepsilon]$ and $[\varepsilon, h]$, where ε is a scale to be determined, we have

$$\begin{aligned}
\mathbb{E}[(\widehat{P}^f_l - \widehat{P}^c_{l-1})^2 \mathbf{1}_{A_2} \mathbf{1}_{E^c_l}] &= \mathbb{E}[\lambda \int_0^h (\widehat{P}^f_l - \widehat{P}^c_{l-1})^2 e^{-\lambda s} \mathrm{d}s \mathbf{1}_{E^c_l}] \\
&= \lambda \mathbb{E}[(\int_0^\varepsilon (\widehat{P}^f_l - \widehat{P}^c_{l-1})^2 e^{-\lambda s} \mathrm{d}s \; + \int_\varepsilon^h (\widehat{P}^f_l - \widehat{P}^c_{l-1})^2 e^{-\lambda s} \mathrm{d}s) \; \mathbf{1}_{E^c_l}] \\
&\leq \lambda \int_0^\varepsilon e^{-\lambda s} \mathrm{d}s + C_1 h^{2-10\gamma} \int_\varepsilon^h \frac{1}{s} e^{-\lambda s} \mathrm{d}s + \frac{1}{4} K_1^2 \int_\varepsilon^h s^{1-4\gamma} e^{-\lambda s} \mathrm{d}s \\
&\leq 1 - e^{-\lambda\varepsilon} + C_1 h^{2-10\gamma} \int_\varepsilon^h \frac{1}{s} \mathrm{d}s + \frac{1}{4} K_1^2 \int_\varepsilon^h s^{1-4\gamma} \mathrm{d}s \\
&\leq 1 - e^{-\lambda\varepsilon} + C_1 h^{2-10\gamma} \log \frac{h}{\varepsilon} + \frac{K_1^2}{4\,(2-4\gamma)} h^{2-4\gamma}.
\end{aligned}$$

By choosing ε to be h^2/T, we obtain that

$$\mathbb{E}[(\widehat{P}^f_l - \widehat{P}^c_{l-1})^2 \mathbf{1}_{A_2} \mathbf{1}_{E^c_l}] \leq 1 - e^{-\lambda h^2/T} + C_1 h^{2-10\gamma} \log \frac{T}{h} + \frac{K_1^2}{4\,(2-4\gamma)} h^{2-4\gamma}.$$

According to [34], for $i = 1, 3$, by choosing $p \geq 2$, we obtain

$$\mathbb{E}[(\widehat{P}^f_l - \widehat{P}^c_{l-1})^2 \mathbf{1}_{A_i}] = \mathcal{O}(h_l^{3/2-13\gamma}).$$

This is the dominant compared with the contribution from A_2. Hence by choosing $\gamma < \min(\frac{1}{4}, \delta/13)$ and $p \geq 2$ we conclude the proof. □

3.4 Analysis of the state-dependent intensity setting

In this section we discuss the state-dependent rate setting where $\lambda(S(t-), t)$ in (3.1) is a function satisfying assumption A5.

A natural approach to handle the bounded state-dependent rate is the thinning method. The idea is to construct a Poisson process with a constant rate which is an upper bound of the state-dependent rate, and then accept candidate jump times according to the probability that is the ratio of the updated intensity to the upper bound. At a jump time τ, we can consider that the superposition jump process comes from two sources :

1. The desired process with rate $\lambda(S(\tau-), \tau)$;

2. The remainder: with rate $\lambda - \lambda(S(\tau-), \tau)$.

To distinguish them we can have an acceptance-rejection procedure. The probability that the candidate jump belongs to the desired process within an infinitesimal interval $[t, t+dt]$ is exactly the ratio of the current intensity to the upper bound.

Mathematically, it transcribes this state-dependent random measure representation into a constant-intensity Poisson random measure one:

$$\begin{aligned} \mathrm{d}S(t) = \quad & a(S(t-),t)\mathrm{d}t + b(S(t-),t)\mathrm{d}W(t) \\ + \quad & \int_0^1 \int_{z\in E} 1_{\left\{u<\frac{\lambda(S(t-),t)}{\lambda_{\sup}}\right\}} c(S(t-),t,z) p_{\lambda_{\sup}}(\mathrm{d}u \times \mathrm{d}z, \mathrm{d}t),\ 0 \le t \le T. \end{aligned} \tag{3.18}$$

To simulate the general processes we use the jump-adapted Milstein discretisation scheme. Suppose we have the superposition of fixed time grid $t_i' = i \times \frac{T}{N}$ and the jump time grid $\mathbb{J}$, that is a jump-adapted grid $\mathbb{T} = \{0 = t_0 < t_1 < t_2 < \ldots < t_M = T\}$, a jump-adapted Milstein scheme with the thinning procedure formally is

$$\begin{aligned} \widehat{S}_{n+1}^{-} &= \widehat{S}_n + a_n\, h_n + b_n\, \Delta W_n + \tfrac{1}{2}\, b_n' b_n\, (\Delta W_n^2 - h_n), \\ \widehat{S}_{n+1} &= \widehat{S}_{n+1}^{-} + \int_{z\in E} 1_{\left\{\frac{\lambda(\widehat{S}_{n+1}^{-}, t_{n+1})}{\lambda_{\sup}} > U_i\right\}} c(\widehat{S}_{n+1}^{-}, t_{n+1}, z) p_{\lambda_{\sup}}(\mathrm{d}z, t_{n+1}), \text{ if } t_{n+1} = \tau_i. \end{aligned}$$

where τ_i, $i = 1, \ldots, N_T$ are jump times, $N_T \sim Poi\left(\lambda_{\sup}\right)$, U_i are independent uniform $[0,1]$ random variables.

For the path-dependent payoffs, we follow equation (3.3.2) and the previous section to construct Brownian-interpolant based estimators in the thinning context.

3.4.1 Multilevel treatment

If we construct a multilevel estimator based on the scheme above, different acceptance probabilities for fine and coarse level lead to samples in which a jump candidate is accepted on the fine path, but not on the coarse path, or vice versa. On level l, at jump time τ, let $p_\tau^f = \dfrac{\lambda(\widehat{S}^f(\tau-), \tau)}{\lambda_{\sup}}$, $p_\tau^c = \dfrac{\lambda(\widehat{S}^c(\tau-), \tau)}{\lambda_{\sup}}$. Due to first order strong convergence of jump-adapted Milstein, $\lambda(\widehat{S}^f(t-), t) - \lambda(\widehat{S}^c(t-), t)$ would be $\mathcal{O}(h_l)$. Thus the proportion of paths where the jump candidates are accepted differently would be $\mathcal{O}(h_l)$, i.e. for the set $E_l = \left\{\tau \in \text{jump times} : (p_\tau^f - U_\tau)(p_\tau^c - U_\tau) < 0\right\}$, $\mathbb{P}\left(E_l\right) = \mathcal{O}(h_l)$. Further since $\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^2 | E_l\right]$ is $\mathcal{O}(1)$, it follows that $V_l = \mathcal{O}(h_l)$.

To improve the variance convergence rate V_l, we use a change of measure so that the acceptance probability is the same for both fine and coarse paths. This is achieved by taking the expectation with respect to a new measure $\mathbb{Q}$:

$$\mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}] = \mathbb{E}_Q[\widehat{P}_l \prod_\tau R^f_\tau - \widehat{P}_{l-1} \prod_\tau R^c_\tau]. \tag{3.19}$$

where τ are the jump times. The acceptance probability for a candidate jump under the measure $\mathbb{Q}$ is $\frac{1}{2}$ for both coarse and fine paths, instead of $p_\tau = \lambda(S(\tau-), \tau)\,/\,\lambda_{\text{sup}}$. The corresponding Radon-Nikodym derivatives are

$$R^f_\tau = \begin{cases} 2p^f_\tau, & \text{if } U_\tau < \frac{1}{2} \quad ; \\ 2(1-p^f_\tau), & \text{if } U_\tau \geq \frac{1}{2} \quad , \end{cases} \qquad R^c_\tau = \begin{cases} 2p^c_\tau, & \text{if } U_\tau < \frac{1}{2} \quad ; \\ 2(1-p^c_\tau), & \text{if } U_\tau \geq \frac{1}{2} \quad , \end{cases}$$

where U_τ are independent $(0,1)$ uniform random variables.

On the right hand side of (3.19), under measure Q, $\widehat{S}^f_i$ and $\widehat{S}^c_i$ are simulated by the scheme:

$$\begin{aligned} \widehat{S}^-_{n+1} &= \widehat{S}_n + a_n\, h_n + b_n\, \Delta W_n + \tfrac{1}{2}\, b'_n b_n\, (\Delta W_n^2 - h_n), \\ \widehat{S}_{n+1} &= \widehat{S}^-_{n+1} + \int_{z\in E} 1_{\{\frac{1}{2} > U_i\}} c(\widehat{S}^-_{n+1}, t_{n+1}, z) p_{\lambda_{\text{sup}}}(\mathrm{d}z, t_{n+1}), \text{ if } t_{n+1} = \tau_i. \end{aligned} \tag{3.20}$$

where τ_i, $i = 1, \ldots, N_T$ are jump times, $N_T \sim Poi\,(\lambda_{\text{sup}})$.

This is the jump-adapted Milstein scheme with the thinning procedure for

$$\begin{aligned} \mathrm{d}S(t) &= a(S(t-), t)\mathrm{d}t + b(S(t-), t)\mathrm{d}W(t) \\ &\quad + \int_0^1 \int_{z\in E} 1_{\{u<\frac{1}{2}\}} c(S(t-), t, z) p_{\lambda_{\text{sup}}}(\mathrm{d}u \times \mathrm{d}z, \mathrm{d}t), \quad 0 \leq t \leq T. \end{aligned}$$

which is equivalent to a constant $\lambda_{\text{sup}}/2$ intensity jump-diffusion SDE. Assumption A2-A4 are satisfied for this SDE, so Theorem 3.2.1, 3.3.1 and 3.3.2 hold for the numerical solution.

To analyse the variance of the estimator $Y = \widehat{P}_l \prod_{i=1}^{N_T} R^f_i - \widehat{P}_{l-1} \prod_{i=1}^{N_T} R^c_i$, using (3.17) iteratively we split Y into two parts

$$\begin{aligned} Y &= \widehat{P}_l \prod_{i=1}^{N_T} R^f_i - \widehat{P}_{l-1} \prod_{i=1}^{N_T} R^c_i \\ &= \widehat{P}_l \sum_{j=1}^{N_T} \left(R^f_j - R^c_j\right) \prod_{i=1}^{j-1} R^f_i \prod_{i=j+1}^{N_T} R^c_i \\ &\quad + \left(\widehat{P}_l - \widehat{P}_{l-1}\right) \prod_{i=1}^{N_T} R^c_i. \end{aligned} \tag{3.21}$$

Thus

$$\begin{aligned}\mathbb{E}\left[Y^2\right] &\leq 2\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^2 \left(\prod_{i=1}^{N_T} R_i^c\right)^2\right] \\ &\quad + 2\mathbb{E}\left[\widehat{P}_l \left(\sum_{j=1}^{N_T} \left(R_j^f - R_j^c\right) \prod_{i=1}^{j-1} R_i^f \prod_{i=j+1}^{N_T} R_i^c\right)^2\right].\end{aligned}$$

By Hölder's inequality and using the result for the constant rate cases, for arbitrary $1/p + 1/q = 1$, where $p, q > 1$,

$$\text{The first part} \leq 2\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^{2p}\right]^{1/p} \left(\mathbb{E}\left[\left(\prod_{i=1}^{N_T} R_i^c\right)^{2q}\right]\right)^{1/q}.$$

Due to the fact that $R_i^c \leq 2$,

$$2\left(\mathbb{E}\left[\left(\prod_{i=1}^{N_T} R_i^c\right)^{2q}\right]\right)^{1/q} \leq 2\left(\mathbb{E}\left[2^{2qN_T}\right]\right)^{1/q} = 2\exp\left(\lambda T \frac{4^q - 1}{q}\right) \equiv C_1(p).$$

Hence we have for arbitrary $p > 1$,

$$\text{The first part} \leq C_1(p) \left(\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^{2p}\right]\right)^{1/p}.$$

We emphasise that $\widehat{P}_l$ and $\widehat{P}_{l-1}$ are constructed using the jump-adapted Milstein scheme with the thinning procedure for a constant $\lambda_{\sup}/2$ intensity jump-diffusion SDE.

The second part can also be bounded. By Hölder's inequality, for arbitrary $1/p+$

$1/q = 1,\ p, q > 1$, we have

$$
\begin{aligned}
&\text{The second part} \\
\leq\ & 2\left(\mathbb{E}\left[\left|\widehat{P}_l\right|^{2p}\right]\right)^{1/p}\left(\mathbb{E}\left[\left|\sum_{j=1}^{N_T}\left(R_j^f - R_j^c\right)\prod_{i=1}^{j-1}R_i^f\prod_{i=j+1}^{N_T}R_i^c\right|^{2q}\right]\right)^{1/q} \\
\leq\ & 2C\left(p,T\right)\left(\mathbb{E}\left[\left|\sum_{j=1}^{N_T}\left|p_j^f - p_j^c\right|2^{N_T}\right|^{2q}\right]\right)^{1/q} \\
\leq\ & 2C\left(p,T\right)\frac{1}{\lambda_{\text{sup}}^2}L_\lambda\left(\mathbb{E}\left[\left|\sum_{j=1}^{N_T}\left|\widehat{S}_j^f - \widehat{S}_j^c\right|2^{N_T}\right|^{2q}\right]\right)^{1/q} \\
\leq\ & 2C\left(p,T\right)\frac{1}{\lambda_{\text{sup}}^2}L_\lambda\left(\mathbb{E}\left[\sup_{1\leq j\leq N_T+2^l}\left|\widehat{S}_j^f - \widehat{S}_j^c\right|^{4q}\right]\right)^{1/2q}\left(\mathbb{E}\left[\left|2^{N_T}N_T\right|^{4q}\right]\right)^{1/2q} \\
\leq\ & C_2\left(p,T,\lambda_{\text{sup}},L_\lambda\right)h_l^2,
\end{aligned}
$$

where we use the boundness of any moments of the interpolated numerical solution, boundness of R_i, and Lipschitz condition on $\lambda\left(\cdot,t\right)$ where L_λ is the Lipschitz constant. Theorem 3.3.1 is used to bound the maximum of the numerical solution in the last step. Therefore we conclude that for arbitrary $p_1, p_2 > 1$,

$$
\mathbb{E}\left[Y^2\right] \leq C_1(p_1)\left(\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^{2p_1}\right]\right)^{1/p_1} + C_2\left(p_2,T,\lambda_{\text{sup}},L_\lambda\right)h_l^2,
$$

which means that the variance of the ML correction term using a change of measure in the state-dependent rate case can be decomposed into a ML correction term with constant intensity and a second order term.

The analysis in the constant rate case can be carried through to obtain the bound of $\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^{2p_1}\right]$. For any $\delta > 0$, we can choose p_1 sufficiently close to 1 s.t. previous analysis can be extended to prove

$$
\left(\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^{2p_1}\right]\right)^{1/p_1} = \begin{cases} o\left(h^{2-\delta}\right), & \text{Lipschitz, Asian and Lookback payoff;} \\ o\left(h^{3/2-\delta}\right), & \text{digital payoff;} \\ o\left(h^{1-\delta}\right), & \text{Barrier payoff.} \end{cases}
$$

We have the results for $\mathbb{E}\left[Y^2\right]$ accordingly.

To summarise:

Theorem 3.4.1. *For arbitrary $p_1, p_2 > 1$, assuming conditions A1-A5 on both (3.1) and the SDE with λ_{sup} constant intensity, and conditions on the local boundness of the* *density of S_t or $\sup_{[0,t]} S_s$ as specified in Theorem 3.3.13 and 3.3.15, for Lipschitz, Asian, lookback, barrier and digital option we have ML correction term estimators constructed based on (3.4) satisfying*

$$V_l \leq C_1(p_1)\left(\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^{2p_1}\right]\right)^{1/p_1} + C_2\left(p_2, T, \lambda_{\text{sup}}, L_\lambda\right) h_l^2$$

where $C_1(p_1)$ is a constant increasing in p_1, and $C_2\left(p_2, T, \lambda_{\text{sup}}, L_\lambda\right)$ is a constant. The convergence rate of V_l is $o\left(h^{2-\delta}\right)$ for Lipschitz, Asian and Lookback payoff; $o\left(h^{3/2-\delta}\right)$ for digital payoff; $o\left(h^{1-\delta}\right)$ for Barrier payoff.

We have obtained the general variance convergence rate using a jump-adapted Milstein scheme with the thinning procedure and a change of measure.

The weak convergence of the jump-adapted discretisation with the thinning procedure is proved in [41], for a class of payoffs on which they impose to 4th order differentiability and uniformly bounded partial derivatives up to 4th order. The drawback of this approach is that you have to construct a series of payoffs in that class which converges to the target payoff, which could be difficult in some discontinuous or path-dependent payoffs.

The alternative method is to use the same decomposition (3.21), decomposing the estimator into the constant rate ML correction part and the difference of fine and coarse Radon-Nikdym derivatives part. Assuming the Lipschitz condition on λ, we can obtain the weak convergence of the estimators for various payoffs, circumventing the difficulties caused by discontinuous payoffs. Another advantage of this argument is that it can reduce the analysis to the constant rate case so that the proofs for various payoffs are simplified.

Hence by similar arguments we conclude

Theorem 3.4.2. *Provided assumptions A1-A5 are satisfied, we have*

$$\left|\mathbb{E}[\widehat{P}_l - P]\right| \leq C_1(p_1)\left(\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^{p_1}\right]\right)^{1/p_1} + C_2\left(p_2, T, \lambda_{\text{sup}}, L_\lambda\right) h_l$$

for arbitrary $p_1, p_2 > 1$, where $C_1(p_1)$ is a constant increasing in p_1, and $C_2\left(p_2, T, \lambda_{\text{sup}}, L_\lambda\right)$ is a constant. In particular we can take $p_1 = 2$ to use the previous constant intensity case results.

3.5 Conclusions

In this chapter we present numerical analysis for MLMC applied to option pricing in scalar jump-diffusion models. We discuss European call, Asian, lookback, barrier and digital options in the constant jump rate case. We also extend the analysis to the case of bounded state-dependent intensities. In more detail, we have proved that the estimator constructed using a Brownian interpolant is convergent weakly to the analytic solution: Theorems 3.3.1 and 3.3.2 together give $\mathcal{O}(h)$ weak convergence for the Lipschitz and Asian options. For the lookback, digital and barrier options, the weak order of convergence is $o(h^{1-\delta})$ for any $\delta > 0$ due to Theorems 3.3.8, 3.3.12 and 3.3.14.Therefore in all cases, the weak convergence rate satisfies the inequality $\alpha \geq \frac{1}{2}$ required by Theorem 1.3.1. The variance of the multilevel estimator is $\mathcal{O}(h_l^2)$ for Lipschitz and Asian options, $\mathcal{O}(h_l^2 \, |\log h_l|^2)$ for lookback options, and $o(h_l^{3/2-\delta})$ for digital , and $o(h_l^{1-\delta})$ for barrier for options, for any $\delta > 0$. In the bounded state-dependent rate case, using thinning with change of measure, results for weak convergence and convergence of variance of the multilevel estimator are established as well. Consequently, the same $O(\epsilon^{-2})$ computational complexity is proved for all cases except for the barrier option for which the best that can currently be proved is a near first order convergence of the variance, leading to the computational complexity being $o(\epsilon^{-2-\delta})$ for any strictly positive δ. These analyses support the numerical results in chapter 2.

Chapter 4

Multilevel Greeks in Jump-diffusion Models

In the last two chapters we have established how to use Multilevel (ML) to efficiently estimate expectations of various payoffs. This chapter addresses the problem of estimating sensitivities of expectations with respect to parameters. In financial applications, sensitivities of derivative prices are usually referred to as Greeks. The first importance of computing Greeks is due to hedging. In the continuous trading Black-Scholes model, in order to hedge the risk of contingent claims, dynamic balancing a position with a certain amount of underlying assets can offset the risk of the position in derivatives. For example for a European call option, in the Black-Scholes model, the uncertainty of the price of the derivative in the future can be completely hedged by a delta-hedging strategy, where delta refers to the partial derivative of the option price with respect to the underlying price. In jump-diffusion models, the market is incomplete, i.e. the risk of derivatives cannot be fully eliminated by holding the underlying and other tradable assets. Nevertheless, according to different hedging criteria, the risk of derivative securities can be managed by hedging the position on the underlying or other tradable assets, where the amount is calculated through the sensitivity of option prices with respect to the current underlying price and other parameters, i.e. the Greeks. Another aspect of computing Greeks lies in the fact that they are not quotable in the market- they can only be calculated according to the model. The computation of Greeks also brings lots of challenges which are not present in the price estimation case, especially for discontinuous payoffs. Last but not least, Greeks are useful in the calibration of models, since the computation of price sensitivities is necessary for solving the optimisation problem to find out optimal model parameters. The objective of the optimisation problem is to minimise the calibration error defined as a certain distance between model prices and market prices.

In the following sections we first review two major methods of sensitivity estimation. Then we recall the jump-diffusion model, and the jump-adapted Milstein scheme for computing numerical path value of the model. In the constant jump rate case, for delta and vega, we mainly follow the multilevel pathwise approach developed in [7]. For the sensitivity to the rate, we find that it consists of two parts: we derive a likelihood ratio estimator for the part coming from perturbation of the rate to the number of jumps in the path, and two estimators for the part caused by perturbation of the intensity compensator in the drift term of the SDE. In the state-dependent rate case, we use the thinning procedure for path simulations, and corresponding estimators of Greeks are derived.

4.1 Review of Greeks computation by Monte Carlo

Two major estimation methods for Greeks differ in whether we differentiate the path value or the transition density. Here we briefly introduce those two methods under our notation and model setting. See chapter 7 of [39] for detailed discussions.

In the constant jump intensity setting, assuming there is no jump risk premium, the jump-diffusion SDE in the risk-neutral world is

$$\mathrm{d}S(t) = (r - c(S(t-),t)\lambda m)\,\mathrm{d}t + b(S(t-),t)\mathrm{d}W(t) + c(S(t-),t)\mathrm{d}J(t), \quad 0 \le t \le T, \tag{4.1}$$

where r is risk-free rate, $J(t)$ is a compound Poisson process $\sum_{i=1}^{N_t}(Y_i - 1)$, the jump magnitude Y_i has a prescribed distribution which has finite mean, the compensator rate $m = \mathbb{E}[Y] - 1$, and N_t is a Poisson process with intensity λ, independent of the Brownian motion.

We use a jump-adapted numerical discretisation scheme with 2^l uniform timesteps to calculate the numerical approximation of $\mathrm{d}S(t)$ up to time $T : \widehat{S}_0, \widehat{S}_1, \widehat{S}_2, \ldots, \widehat{S}_{2^l+N_T}$.

Suppose we have to estimate the partial derivative of a payoff on the numerical path value $\widehat{P}_l = f(\widehat{S}\,(\theta))$ with respect to a model parameter θ :

$$\frac{\partial}{\partial\theta}\mathbb{E}\left[\widehat{P}_l\right], \tag{4.2}$$

where the payoff is evaluated by: $\widehat{P}_l = f(\widehat{S}) = f(\widehat{S}_0, \widehat{S}_1, \widehat{S}_2, \ldots, \widehat{S}_{2^l+N_T})$. $\widehat{S}$ depends on the number of jumps N_T, jump times τ_i, $i = 1 \ldots N_T$, Brownian increments $W = (W_1, W_2, \ldots, W_{2^l+N})$ and jump magnitudes $Y = (Y_1, Y_2, \ldots, Y_{2^l+N})$.

Pathwise sensitivities

If we informally write the expectation (4.2) as an integral with respect to the density of W, N_T, jump times τ_i, jump sizes Y_i, i.e.

$$\mathbb{E}\left[\widehat{P}_l\right] = \int f(\widehat{S}\,(\theta))\; p(W, N_T, \tau_i, Y_i)\; \mathrm{d}W\; \mathrm{d}N_T\; \Pi_{i=1}^{N_T} \mathrm{d}\tau_i\; \mathrm{d}Y_i,$$

under regularity conditions which will be mentioned later, it is legitimate to interchange the order of differentiation and integration, i.e.

$$\begin{aligned} & \frac{\partial}{\partial\theta} \int f(\widehat{S}\,(\theta))\; p(W, N_T, \tau_i, Y_i)\; \mathrm{d}W\; \mathrm{d}N_T\; \mathrm{d}Y \\ = & \int \frac{\partial}{\partial\theta} f(\widehat{S}\,(\theta))\; p(W, N_T, \tau_i, Y_i)\; \mathrm{d}W\; \mathrm{d}N_T\; \Pi_{i=1}^{N_T} \mathrm{d}\tau_i\; \mathrm{d}Y_i. \end{aligned}$$

If the payoff is smooth enough, for example if it is differentiable, we may further write

$$\frac{\partial}{\partial\theta} f(\widehat{S}\,(\theta)) = \left(\frac{\partial f}{\partial \widehat{S}}\right)^T \cdot \frac{\partial \widehat{S}}{\partial\theta} = \sum_i \frac{\partial f}{\partial \widehat{S}_i} \frac{\partial \widehat{S}_i}{\partial\theta},$$

that is, the sensitivity equals derivative of payoff with respect to the path value multiplied by derivative of path value with respect to the desired parameter.

This approach is called pathwise sensitivity analysis. The advantage of this method is that its variance is relatively stable when the number of discretisation timesteps goes up (p410 of [39]). It also saves lots of computational costs when one wants multiple sensitivities for the same payoff if we use the adjoint method [35]. The drawback is that it requires the payoff to be smooth. But this obstacle can be coped with by smoothing the payoff (p399 of [39]) or a combination with the Likelihood ratio method, called Vibrato Monte Carlo [33].

Conditions for an unbiased pathwise estimator

To validate the correctness of estimators derived in the jump-diffusion setting, we need to review conditions needed for an unbiased estimator.

In [39], sufficient conditions that

$$\frac{\partial}{\partial\theta} \mathbb{E}\left[\widehat{P}_l\right] = \mathbb{E}\left[\frac{\partial}{\partial\theta} \widehat{P}_l\right]$$

holds are given as

1. $\frac{\partial}{\partial\theta}\widehat{P}_l$ exists with probability 1;

2. For any sufficiently small h, the payoff and underlying dynamics satisfies

$$\left|\frac{f(\widehat{S}(\theta+h))-f(\widehat{S}(\theta))}{h}\right| \leq H,$$

where H is a random variable with finite moment. .

Under those conditions, the interchange of expectation and differentiation is valid thanks to the dominated convergence theorem. It can be verified that the Lipschitz and European call payoffs in the numerical results section satisfy those conditions. For more details about alternatives for these exchangeability conditions, we refer to Section 7.2.2 of [39].

Likelihood ratio method (LRM)

Suppose $\widehat{S}$ has a probability density function (p.d.f.) depending on $\theta : p\left(\widehat{S},\theta\right)$. We can write the expected payoff (4.2) as an integral with respect to the density of $\widehat{S}$:

$$\mathbb{E}\left[\widehat{P}_l\right] = \int f\left(\widehat{S}\right) p\left(\widehat{S},\theta\right) \mathrm{d}\widehat{S}.$$

As discussed in p407 of [39], in the applications we are concerned with, the differentiation and integration is interchangeable:

$$\begin{aligned}
\frac{\partial}{\partial\theta}\mathbb{E}\left[\widehat{P}_l\right] &= \int f\left(\widehat{S}\right)\frac{\partial p\left(\widehat{S},\theta\right)}{\partial\theta}\mathrm{d}\widehat{S} \\
&= \int f\left(\widehat{S}\right)\frac{\partial\left(\log p\left(\widehat{S},\theta\right)\right)}{\partial\theta} p\left(\widehat{S},\theta\right)\mathrm{d}\widehat{S} \\
&= \mathbb{E}\left[f\left(\widehat{S}\right)\frac{\partial\left(\log p\left(\widehat{S},\theta\right)\right)}{\partial\theta}\right].
\end{aligned}$$

This is called the Likelihood ratio method, which transfers the differentiation to the density of $\widehat{S}$. The scope of the LRM is broader- its unbiasedness is to do with the smoothness of the density rather than that of the payoff, and in the applications we are concerned with, the density satisfies the condition according to [39]. The drawback is that in the multidimensional cases, e.g. for path-dependent payoffs, the variance of the LRM estimator increases with the number of discretisation steps. This limitation motivates people to use the LRM in the last timestep for the digital payoff, e.g. in Vibrato Monte Carlo [33].

As in the previous chapter, in the following we first discuss the case of constant-rate jump-diffusion models. Then the state-dependent case will be demonstrated.

4.2 Constant jump rate case

4.2.1 Delta and vega by Pathwise methods

If we use the jump-adapted Milstein scheme for path simulation, pathwise estimators for delta and vega in the diffusion setting can be simply extended to the jump-diffusion setting. The difference is to add jumps to path simulation.

Recall that using the jump-adapted Milstein scheme, the numerical approximation $\widehat{S}$ would be

$$\begin{aligned}
\widehat{S}_{n+1}^{-} &= \widehat{S}_n + a_n\, h_n + b_n\, \Delta W_n + \tfrac{1}{2}\, b_n' b_n\, (\Delta W_n^2 - h_n) := \widehat{S}_n + D_n, \\
\widehat{S}_{n+1} &= \widehat{S}_{n+1}^{-} + c(\widehat{S}_{n+1}^{-}, t_{n+1}, z)(Z_{n+1} - 1); \\
\widehat{S}_0 &= S_0,
\end{aligned}$$

where $Z_n = Y_i 1_{\{t_n=\tau_i\}} + 1_{\{t_n \neq \tau_i\}}. n = 1, \ldots, N_T + 2^l$. τ_i is the ith jump time.

By differentiating both sides of the scheme, $\frac{\partial \widehat{S}_i}{\partial \theta}$ is computed recursively by:

$$\begin{aligned}
\frac{\partial \widehat{S}_{n+1}^{-}}{\partial \theta} &= \frac{\partial \widehat{S}_n}{\partial \theta} + \frac{\partial D_n}{\partial \theta}, \\
\frac{\partial \widehat{S}_{n+1}}{\partial \theta} &= \frac{\partial \widehat{S}_{n+1}^{-}}{\partial \theta} + \frac{\partial}{\partial \theta} c(\widehat{S}_{n+1}^{-}, t_{n+1}, z)(Z_{n+1} - 1); \\
\frac{\partial \widehat{S}_0}{\partial \theta} &= \frac{\partial S_0}{\partial \theta},
\end{aligned} \tag{4.3}$$

where $\theta = S_0$ or σ.

4.2.2 The sensitivity to the rate

For the sensitivity to the jump rate, since the perturbation of the rate leads to different jump times, it is not obvious how to derive the recursive equation of $\frac{\partial \widehat{S}_i}{\partial \lambda}$ like (4.3).

A way to derive the estimator of the sensitivity to the jump intensity is by conditioning on the number of jumps:

$$\begin{aligned}
\frac{\partial}{\partial\lambda}\mathbb{E}[\widehat{P}_l] &= \frac{\partial}{\partial\lambda}\sum_{n=0}^{\infty} e^{-\lambda T}\frac{(\lambda T)^n}{n!}\mathbb{E}[\widehat{P}_l \mid N_T = n] \\
&= \sum_{n=0}^{\infty} e^{-\lambda T}\frac{(\lambda T)^n}{n!}\left[\left(\frac{n}{\lambda} - T\right)\mathbb{E}[\widehat{P}_l \mid N_T = n] + \frac{\partial}{\partial\lambda}\mathbb{E}[\widehat{P}_l \mid N_T = n]\right]
\end{aligned}$$

Since given the number of jumps up to time T, the distribution of jump times is independent of λ (actually they are uniformly distributed on $[0, T]$), $\mathbb{E}[\widehat{P}_l \mid N_T = n]$ depends on λ only because the drift term of the SDE includes the compensator of jumps. Hence it can be written as

$$\mathbb{E}[\widehat{P}_l \mid N_T = n] = \int f(\widehat{S}(\lambda))\; p(W, Y_i, \tau_i)\; \mathrm{d}W\; \Pi_{i=1}^{n}\mathrm{d}\tau_i\; \mathrm{d}Y_i$$

where τ_i are jump times and Y_i are jump magnitudes. It then makes sense to interchange the order of differentiation and integration.

$$\begin{aligned}
\frac{\partial}{\partial\lambda}\mathbb{E}[\widehat{P}_l] &= \sum_{n=0}^{\infty} e^{-\lambda T}\frac{(\lambda T)^n}{n!}\left[\left(\frac{n}{\lambda} - T\right)\mathbb{E}[\widehat{P}_l \mid N_T = n] + \mathbb{E}[\frac{\partial\widehat{P}_l}{\partial\lambda} \mid N_T = n]\right] \\
&= \mathbb{E}\left[\left(\frac{N_T}{\lambda} - T\right)\widehat{P}_l + \frac{\partial\widehat{P}_l}{\partial\lambda}\right]. \qquad (4.4)
\end{aligned}$$

This is an estimator of the sensitivity to rate. The first part is an LRM estimator in the sense that the density of N_T is differentiated, representing the contribution from the perturbation of the rate to the jump times on the path, and the second part comes from the perturbation of the rate to the jump compensator in the drift term.

The second part $\mathbb{E}\left[\frac{\partial\widehat{P}_l}{\partial\lambda}\right]$ can be computed pathwise, when the payoff is differentiable:

$$\frac{\partial\widehat{P}_l}{\partial\lambda} = \frac{\partial f}{\partial\widehat{S}} \cdot \frac{\partial\widehat{S}}{\partial\lambda}.$$

On the other hand, for the sake of avoiding differentiating the payoff, we derive another estimator for $\mathbb{E}\left[\frac{\partial\widehat{P}_l}{\partial\lambda}\right]$. Proposition 3.6 of [15] can be generalised without difficulty to our time-dependent drift and diffusion coefficients and state-dependent jump coefficient setting, using the same change of measure argument.

Proposition 4.2.1. *Suppose $b(S(t-), t) > 0$ for $0 \le t \le T$ in the jump-diffusion SDE (4.1). Let $S_n = S(t_n)$ for $0 \le t_i \le T$, $i = 1 \ldots n$.*

We have

$$
\begin{aligned}
&\frac{\partial}{\partial\lambda}\mathbb{E}[f(S_0,S_1,S_2,\ldots,S_n)]\\
=\ &\mathbb{E}\left[f(S_0,S_1,S_2,\ldots,S_n)\left(\frac{N_T}{\lambda}-T-m\int_0^T b^{-1}(S(t-),t)c\,(S(t-),t)\,\mathrm{d}W(t)\right)\right].
\end{aligned}
$$

Applying this proposition to the discretised version of the jump-diffusion SDE and comparing with (4.4), we have

$$
\mathbb{E}\left[\frac{\partial\widehat{P}_l}{\partial\lambda}\right]=-m\mathbb{E}\left[\widehat{P}_l\sum_{n=1}^{2^l+N_T}b_n^{-1}c_n\,(W_n-W_{n-1})\right]. \tag{4.5}
$$

In the Merton model where we test the numerical result, $b_n^{-1}c_n=\sigma^{-1}$ is a constant. Therefore $\mathbb{E}\left[\frac{\partial\widehat{P}_l}{\partial\lambda}\right]=-m\sigma^{-1}\mathbb{E}\left[\widehat{P}_lW_T\right]$ which is easy to simulate.

4.2.3 Numerical results

We again use a Merton model to do numerical tests in the constant rate case. The jump-diffusion SDE under the risk-neutral measure is

$$
\frac{\mathrm{d}S(t)}{S(t-)}=(r-\lambda m)\,\mathrm{d}t+\sigma\,\mathrm{d}W(t)+\mathrm{d}J(t),\quad 0\le t\le T,
$$

where $J(t)$ is a compound Poisson process $\sum_{i=1}^{N(t)}(Y_i-1)$, the jump magnitude Y_i has a log normal distribution, i.e. $\log Y\sim N(a,b^2)$. $N(t)$ is a Poisson process with a constant intensity λ, independent of the Brownian motion. r is the risk-free interest rate, σ is the volatility, and $m=\mathbb{E}[Y_i]-1$ is the compensator to ensure the discounted asset price is a martingale. All of the simulations in this section use the parameter values $S_0=100$, $K=100$, $T=1$, $r=0.05$, $\sigma=0.2$, $a=-0.1$, $b=0.2$, $\lambda=1$.

The multilevel implementation of the estimator is similar to the procedure in previous pricing chapters. We modify Giles's Matlab code used for [31] to produce performance figures.

The numerical results of variance convergence order for Greeks of European call option are summarised in Table 4.1. For the implementation of other path-dependent payoffs, we refer to [7].

Figure 4.1 shows the numerical results for the delta of European call option with a payoff $\exp(-rT)\,(S(T)-K)^+$, where $(x)^+\equiv\max(x,0)$ and strike $K=100$.

	Jump-adapted Milstein	
Greeks	weak	variance
delta	$\mathcal{O}(h)$	$\mathcal{O}(h)$
vega	$\mathcal{O}(h)$	$\mathcal{O}(h)$
rate via pathwise	$\mathcal{O}(h)$	$\mathcal{O}(h^{1.5})$
rate via change of measure	$\mathcal{O}(h)$	$\mathcal{O}(h^{1.75})$

Table 4.1: Orders of convergence for V_l as observed numerically for Greeks of European call option. $h \equiv h_l$.

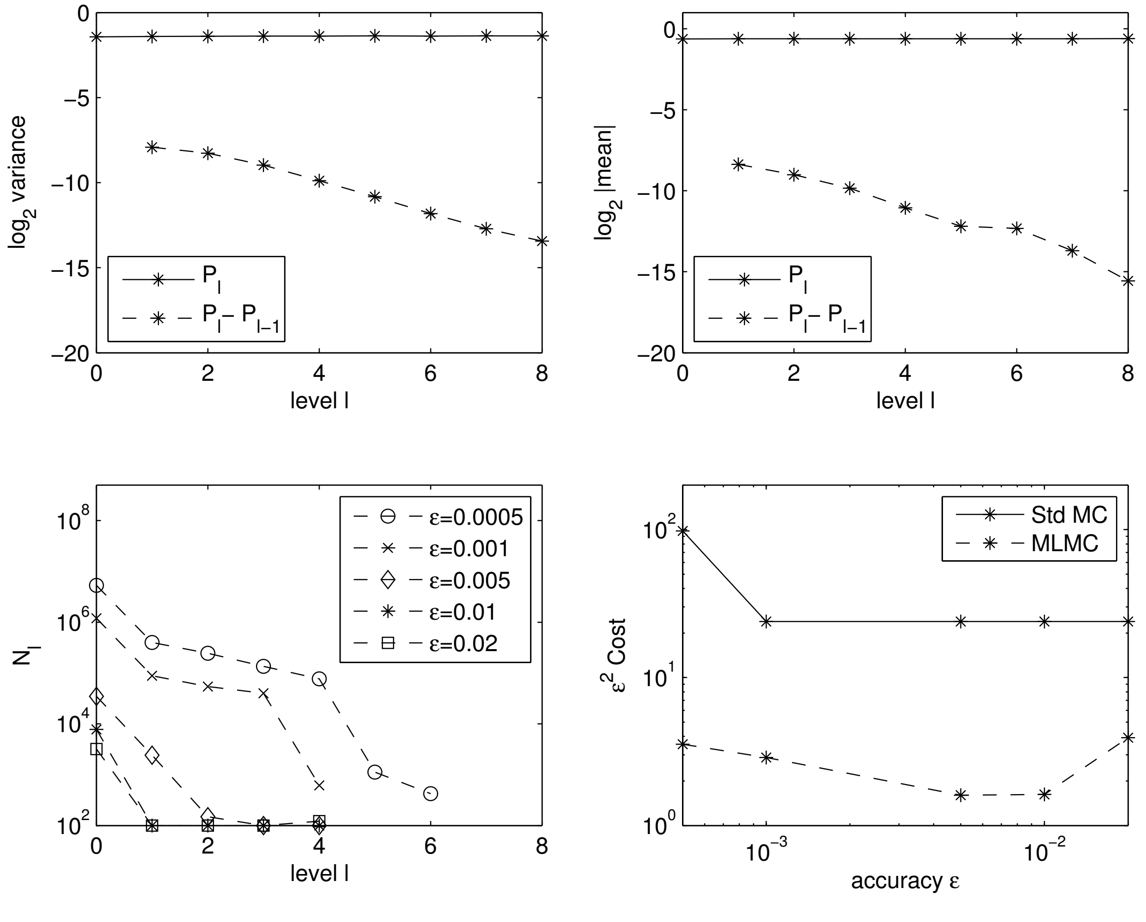

Figure 4.1: Delta of European call option

The top left plot shows the behavior of the variance of both $\frac{\partial \widehat{P}_l}{\partial S_0}$ and the multilevel correction $\frac{\partial\left(\widehat{P}_l - \widehat{P}_{l-1}\right)}{\partial S_0}$, estimated using 10^6 samples so that the Monte Carlo sampling error is negligible. The slope of the MLMC line indicates that $V_l \equiv Var\left[\frac{\partial\left(\widehat{P}_l - \widehat{P}_{l-1}\right)}{\partial S_0}\right] = O(h_l)$, corresponding to $\beta = 1$ in condition $iii)$ of Theorem 1.3.1. The top right plot

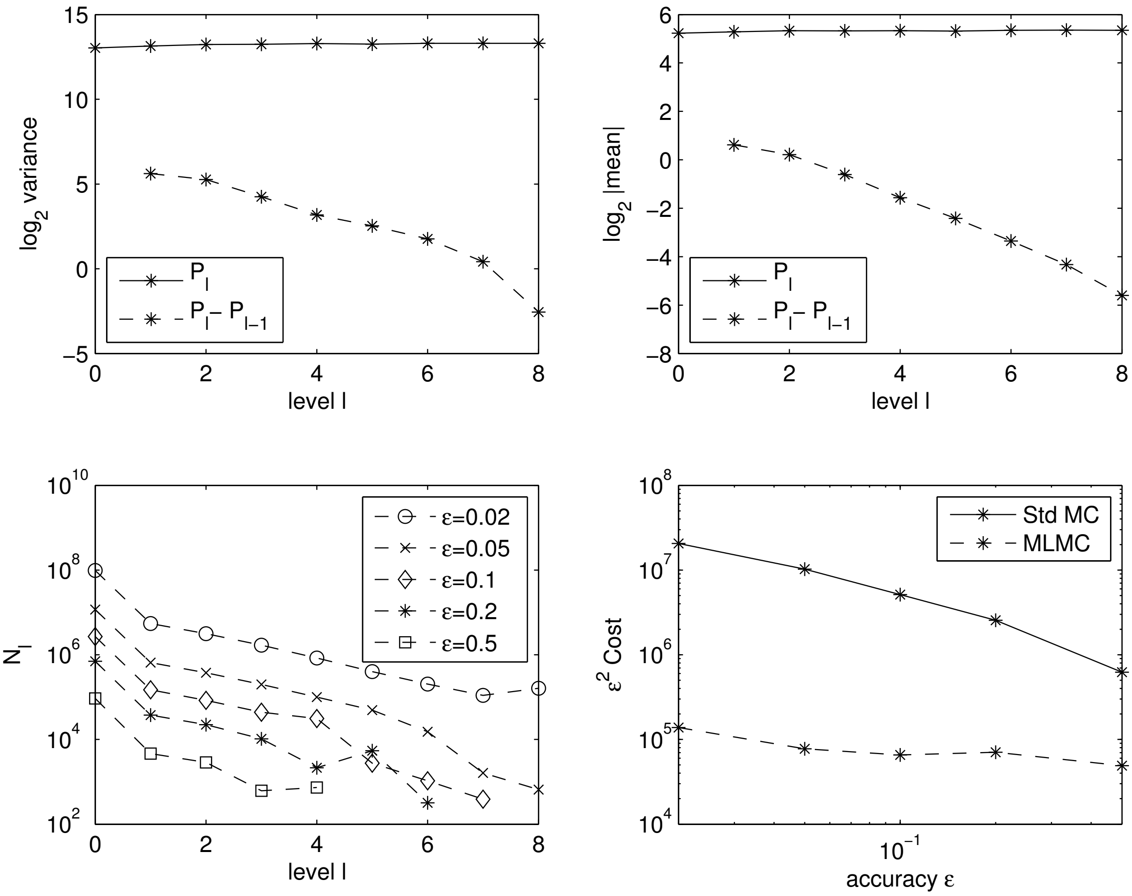

Figure 4.2: Vega of European call option

shows that $\mathbb{E}\left[\frac{\partial\left(\widehat{P}_l - \widehat{P}_{l-1}\right)}{\partial S_0}\right]$ is approximately $O(h_l)$, corresponding to $\alpha=1$ in condition i). Noting that the derivative of the payoff is actually an indicator function, the numerical result corresponds to the case of estimating the value of digital options without using conditional expectation. The numerical result is consistent with the weak convergence result proved in the previous chapter.

The bottom two plots correspond to five different multilevel calculations with different user-specified accuracies to be achieved. These use the numerical algorithm given in [32] to determine the number of grid levels, and the optimal number of samples on each level, which are required to achieve the desired accuracy. The left plot shows that in each case many more samples are used on level 0 than on any other level, with very few samples used on the finest level of resolution. The right plot shows that the multilevel cost is approximately proportional to ϵ^{-2}, which agrees with the computational complexity bound in Theorem 1.3.1 for the $\beta = 1$ case.

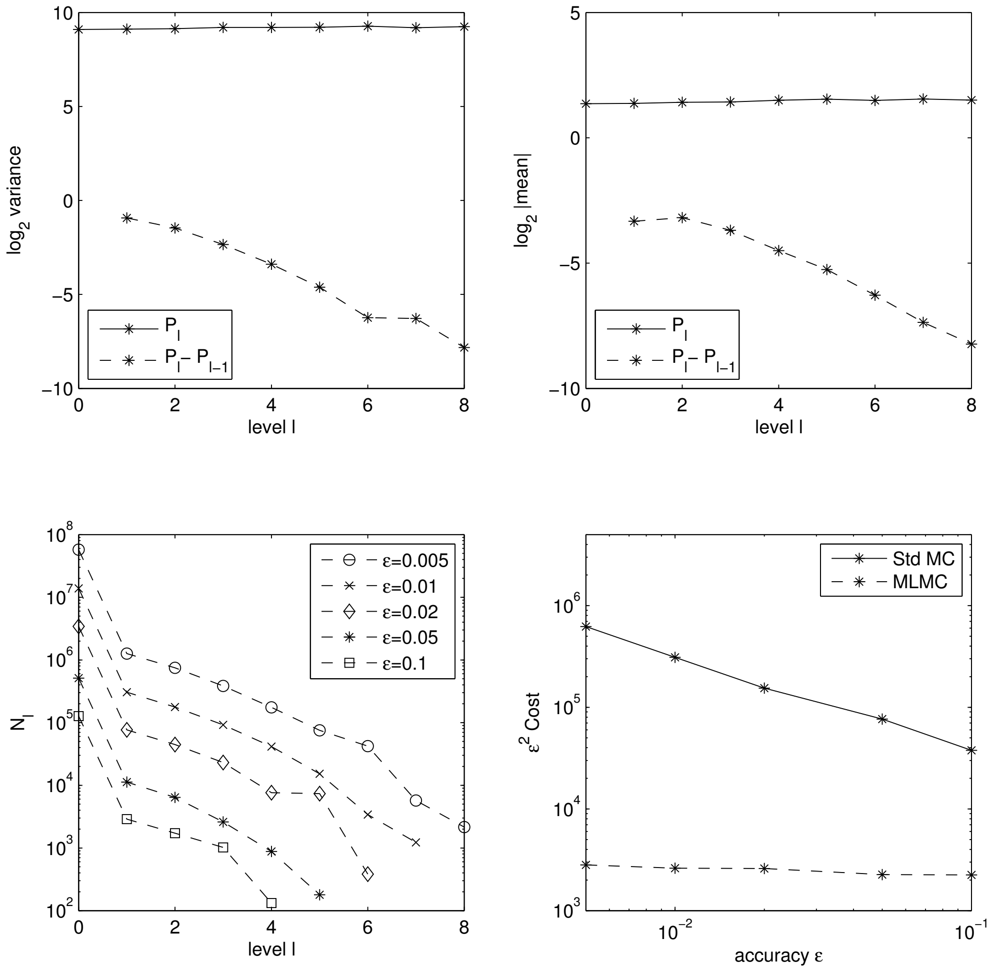

Figure 4.3: Sensitivity to the rate of European call options by pathwise method

Figure 4.2 shows the numerical results for the vega of European call option. We have similar explanations on the performance.

The sensitivity to the rate is

$$\frac{\partial}{\partial\lambda}\mathbb{E}[\widehat{P}_l] = \mathbb{E}\left[\left(\frac{N_T}{\lambda} - T\right)\widehat{P}_l\right] + \mathbb{E}\left[\frac{\partial\widehat{P}_l}{\partial\lambda}\right].$$

Figure 4.3 shows the numerical results for the sensitivity to the jump rate of European call options where $\mathbb{E}\left[\frac{\partial\widehat{P}_l}{\partial\lambda}\right]$ is computed by the pathwise sensitivities method.

Figure 4.4 shows the numerical results for the sensitivity to the jump rate of European call options where $\mathbb{E}\left[\frac{\partial\widehat{P}_l}{\partial\lambda}\right]$ is computed by a change of measure estimator

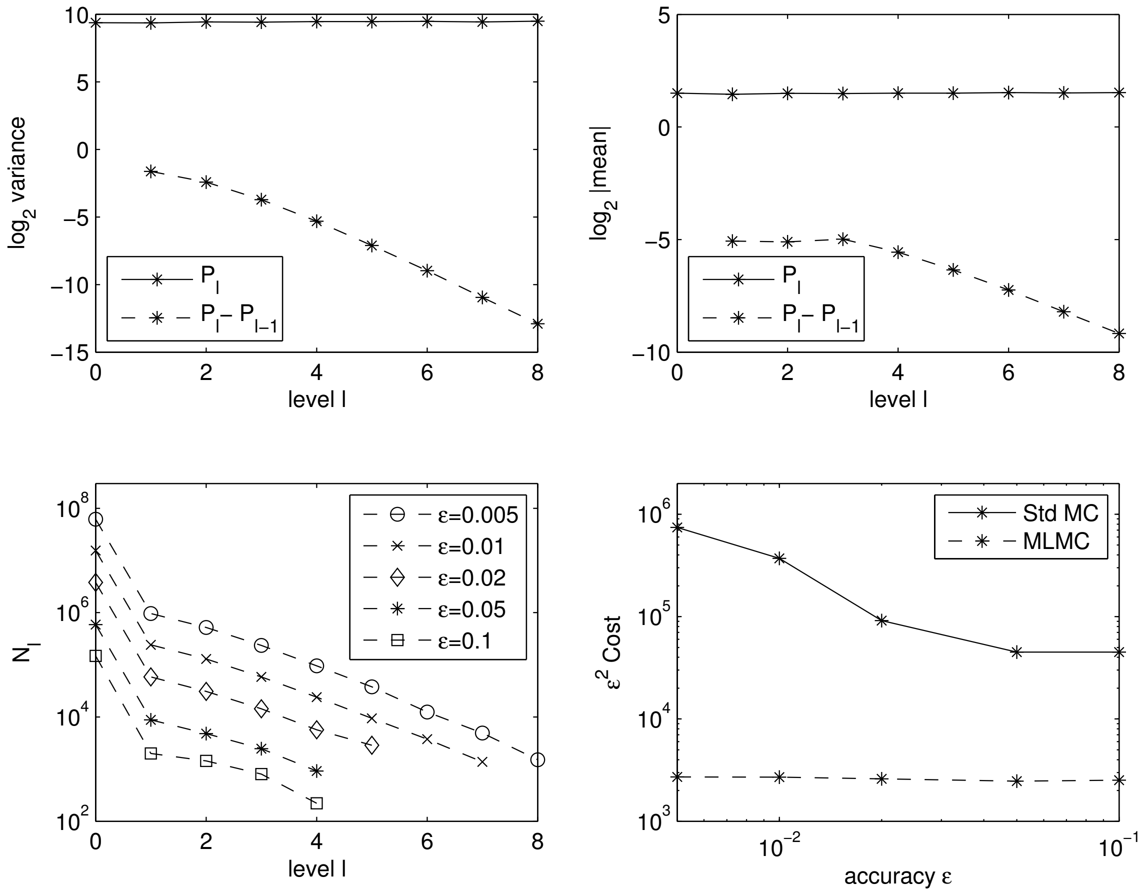

Figure 4.4: Sensitivity to the rate of European call options by change of measure

$-m\sigma^{-1}\mathbb{E}\left[\widehat{P}_l W_T\right]$.

Benchmark

By differentiating the series expansion of the option value, we obtain analytic values of greeks as a validation benchmark for numerical results generated by multilevel computation:

According to [39], the European call option price $V_{merton} = \mathbb{E}\left[e^{-rT}\left(S(T)-K\right)^+\right]$ can be evaluated by

$$V_{merton} = \sum_{n=0}^{\infty} e^{-\lambda' T}\frac{(\lambda' T)^n}{n!} V_{BS}(S,K,T,\sigma_n,r_n),$$

where $\lambda' = \lambda(1+m)$, $\sigma_n^2 = \sigma^2 + \frac{nb^2}{T}$, $r_n = r - \lambda m + \frac{n\log(1+m)}{T}$, and $V_{BS}(S,K,T,\sigma_n,r_n)$ is the Black-Scholes price with stock price S, strike K, maturity T, volatility σ_n,

	delta (using $\epsilon = 0.0005$)	vega (using $\epsilon = 0.02$)	rate (using $\epsilon = 0.005$)
analytic	0.6531	27.14	2.882
pathwise	0.6533	27.10	2.877
change of measure			2.881

Table 4.2: Numerical results of multilevel Greeks for the Merton model compared to the benchmark

risk-neutral rate r_n.

The BS Greeks (delta and vega) can be calculated by

$$\frac{\partial V_{merton}}{\partial \theta} = \sum_{n=0}^{\infty} e^{-\lambda' T} \frac{(\lambda' T)^n}{n!} \frac{\partial V_{BS}}{\partial \theta}(S, K, T, \sigma_n, r_n),$$

The sensitivity to λ is

$$\frac{\partial V_{merton}}{\partial \lambda} = \sum_{n=0}^{\infty} e^{-\lambda' T} \frac{(\lambda' T)^n}{n!} \left((\frac{n}{\lambda} - T) V_{BS}(S, K, T, \sigma_n, r_n) - m \frac{\partial V_{BS}}{\partial r}(S, K, T, \sigma_n, r_n) \right),$$

where the two parts are contributions from the jump term and compensator, respectively.

To make sure the calculations are correct, we compare numerical results with the benchmark in Table 4.2.

4.3 Path-dependent rate case

In the following, we assume the dynamics of the state-dependent jump-diffusion SDE can be written as

$$\mathrm{d}S(t) = a(S(t-),t)\mathrm{d}t + b(S(t-),t)\mathrm{d}W(t) + \int_{z\in E} c(S(t-),t,z)\mu(\mathrm{d}z,\mathrm{d}t), \quad 0 \le t \le T.$$

where $\mu(\mathrm{d}z,\mathrm{d}t)$ is a Poisson random measure. The intensity of μ is defined as $\varphi(\mathrm{d}z)\mathrm{d}t = \lambda(S(t-),t)g(z)\mathrm{d}z\mathrm{d}t$. For all bounded $\mathbb{P}$-measurable function f, we have

$$\int_0^t \int_{z\in E} f(z,s)\mu(\mathrm{d}z,\mathrm{d}s) - \int_0^t \int_{z\in E} f(z,s)\varphi(\mathrm{d}z)\mathrm{d}s$$

is a martingale with respect to $\mathcal{F}_t$.

We again make the assumptions on the coefficients of our jump-diffusion SDEs as listed in Section 3.2.2 of Chapter 3.

4.3.1 Multilevel treatment

Recall the multilevel estimators on level l and $l-1$ are $\widehat{P}_l = f(\widehat{S}^f(\cdot))$, $\widehat{P}_{l-1} = f(\widehat{S}^c(\cdot))$. Using the usual thinning method, the jump times are sampled using rate $\lambda_{\sup}$ and then selected by acceptance probability $p_\tau = \lambda(\widehat{S}(\tau-),\tau)/\lambda_{\sup}$. Due to different acceptance probabilities for coarse and fine paths, the variance of multilevel correction terms is large. Following the previous change of measure approach, to reduce the variance, we introduce a new measure $\mathbb{Q}$ where the acceptance probability for a candidate jump time is $\frac{1}{2}$ for both coarse and fine paths :

$$\mathbb{E}[\widehat{P}_l - \widehat{P}_{l-1}] = \mathbb{E}_{\mathbb{Q}}[\widehat{P}_l \prod_{i=1}^{N_T^{\lambda_{\sup}}} R_i^f - \widehat{P}_{l-1} \prod_{i=1}^{N_T^{\lambda_{\sup}}} R_i^c] := \mathbb{E}_{\mathbb{Q}}[\widehat{Q}^f - \widehat{Q}^c],$$

The corresponding Radon-Nikodym derivatives are: (U_i ~Uniform$[0,1]$ are i.i.d. samples for acceptance-rejection.)

$$R_i = \begin{cases} 2\frac{\lambda(\widehat{S}_i^-,t_i)}{\lambda_{\sup}}, & \text{if } U_i < \frac{1}{2}; \\ 2(1-\frac{\lambda(\widehat{S}_i^-,t_i)}{\lambda_{\sup}}), & \text{if } U_i \ge \frac{1}{2}. \end{cases}$$

Under the measure $\mathbb{Q}$, the numerical sample path $\widehat{S}$ is computed using the following jump-adapted Milstein scheme with the thinning procedure:

$$\begin{aligned}
\widehat{S}_{n+1}^{-} &= \widehat{S}_n + a_n\, h_n + b_n\, \Delta W_n + \tfrac{1}{2}\, b_n' b_n\, (\Delta W_n^2 - h_n) := \widehat{S}_n + D_n, \qquad (4.6)\\
\widehat{S}_{n+1} &= \widehat{S}_{n+1}^{-} + \int_{z\in E} 1_{\{\frac{1}{2}>U_i\}} c(\widehat{S}_{n+1}^{-}, t_{n+1}, z) p_{\lambda_{\text{sup}}}(\mathrm{d}z, t_{n+1}), \text{ if } t_{n+1} = \tau_i.
\end{aligned}$$

where τ_i, $i = 1, \ldots, N_T^{\lambda_{\text{sup}}}$ are jump times; $n = 1, \ldots, N = N_T^{\lambda_{\text{sup}}} + \frac{T}{h}$, and h is the uniform timestep.

4.3.2 Greeks

Recall that under the measure $\mathbb{Q}$, $\widehat{S}$ is simulated using a fixed jump rate λ_{sup} and $\frac{1}{2}$ acceptance probability for a candidate jump time, so the perturbation to the rate parameter does not change the jump times. Define

$$\widehat{Q} = \widehat{P} \prod_{i=1}^{N_T^{\lambda_{\text{sup}}}} R_i.$$

Its sensitivity with respect to some parameter θ, by the chain rule is

$$\frac{\partial \widehat{Q}}{\partial \theta} = \frac{\partial \widehat{P}}{\partial \widehat{S}} \frac{\partial \widehat{S}}{\partial \theta} \prod_{i=1}^{N_T^{\lambda_{\text{sup}}}} R_i + \widehat{P} \sum_{i=1}^{N_T^{\lambda_{\text{sup}}}} \frac{\partial R_i}{\partial \theta} \prod_{j\neq i} R_j, \qquad (4.7)$$

where the sensitivities of the Radon-Nikodym derivatives are

$$\frac{\partial R_i}{\partial \theta} = (-1)^{1_{\{U_i \geq \frac{1}{2}\}}} \frac{2}{\lambda_{\text{sup}}} \frac{\partial \lambda(\widehat{S}_i^{-}, t_i)}{\partial \theta}.$$

$\frac{\partial \lambda(\widehat{S}_i^{-}, t_i)}{\partial \theta}$ may be computed directly from the pathwise derivatives $\frac{\partial \widehat{S}}{\partial \theta}$ if the functional form of λ is explicit. By recalling that D_n is defined in (4.6), the pathwise sensitivity of the path value follows:

$$\begin{aligned}
\frac{\partial \widehat{S}_{n+1}^{-}}{\partial \theta} &= \frac{\partial \widehat{S}_n}{\partial \theta} + \frac{\partial D_n}{\partial \theta},\\
\frac{\partial \widehat{S}_{n+1}}{\partial \theta} &= \frac{\partial \widehat{S}_{n+1}^{-}}{\partial \theta} + \frac{\partial}{\partial \theta} \int_{z\in E} 1_{\{\frac{1}{2}>U_i\}} c(\widehat{S}_{n+1}^{-}, t_{n+1}, z) p_{\lambda_{\text{sup}}}(\mathrm{d}z, t_{n+1}), \text{ if } t_{n+1} = \tau_i.
\end{aligned}$$

Remark: In the implementation of the algorithm, we compute

$$\sum_{i=1}^{N_T} \frac{\partial R_i^f}{\partial \theta} \prod_{j\neq i} R_j^f$$

as a sum of ratios:

$$\prod_{i=1}^{N_T} R_i^f \sum_{i=1}^{N_T} \frac{\partial R_i^f/\partial\theta}{R_i^f}.$$

Note that we don't have a particular formula for the sensitivity to the rate since the rate is not specificed in the general model.

4.3.3 An example of the model

In an example of the model, we demonstrate the steps in computing the Greeks. The underlying dynamics under the risk-neutral measure are:

$$\frac{\mathrm{d}S(t)}{S(t-)} = r\,\mathrm{d}t + \sigma\,\mathrm{d}W(t) + \int_{z\in E} z\mu(\mathrm{d}z,\mathrm{d}t) - \int_{z\in E} zf(z)\mathrm{d}z\lambda\mathrm{d}t, \quad 0 \le t \le T,$$

where the random measure μ has compensator λ dt with the intensity function

$$\lambda = \tfrac{\lambda_0}{1+CS(t-)^2},\ \lambda_0 = 1,\ C = 10^{-4},\ S_0 = 100.$$

$f(z)$ is the p.d.f. of a log normal random variable Y, i.e. $\log Y \sim N(a, b^2)$. $m = E[Y] - 1 = \exp(a + b^2/2) - 1$.

For multilevel implementation, we simulate the payoff using the thinning rate $\lambda_{\text{sup}} = \lambda_0$ with $\frac{1}{2}$ acceptance probability for a candidate jump time. In the scheme (4.6),

$$\int_{z\in E} 1_{\{\frac{1}{2}>U_i\}} c(\widehat{S}_{n+1}^-, t_{n+1}, z) p_{\lambda_0}(\mathrm{d}z, t_{n+1}) = \widehat{S}_{n+1}^- \left(1_{\{\frac{1}{2}>U_i\}}(Y_i - 1)\right).$$

The path calculation would be

$$\begin{aligned}\widehat{S}_{n+1}^- &= \widehat{S}_n + \widehat{S}_n\left(r - \lambda(\widehat{S}_n^-, t_n)m\right) h_n + \widehat{S}_n\sigma\,\Delta W_n + \widehat{S}_n\tfrac{1}{2}\sigma^2\left(\Delta W_n^2 - h_n\right) := \widehat{S}_n F_n,\\ \widehat{S}_{n+1} &= \widehat{S}_{n+1}^- + \widehat{S}_{n+1}^- 1_{\{\frac{1}{2}>U_i\}}(Y_i - 1) \text{ if } t_{n+1} = \tau_i.\end{aligned}$$

For $\theta \neq \lambda_0$, from (4.7) we have

$$\frac{\partial \mathbb{E}\left[\widehat{Q}\right]}{\partial\theta} = \mathbb{E}\left[\frac{\partial\widehat{Q}}{\partial\theta}\right] = \mathbb{E}\left[\frac{\partial\widehat{P}}{\partial\widehat{S}}\frac{\partial\widehat{S}}{\partial\theta}\prod_{i=1}^{N_T^{\lambda_{\text{sup}}}} R_i + \widehat{P}\sum_{i=1}^{N_T^{\lambda_{\text{sup}}}} \frac{\partial R_i}{\partial\theta}\prod_{j\neq i} R_j\right].$$

For the sensitivity to the thinning rate, which is still a measure to the effect of the change of the rate to the price

$$\begin{aligned}\frac{\partial}{\partial\lambda_0}\mathbb{E}[\widehat{P}\prod_{i=1}^{N_T^{\lambda_0}} R_i] &= \mathbb{E}\left[\left(\frac{N_T^{\lambda_0}}{\lambda_0}-T\right)\widehat{P}\prod_{i=1}^{N_T^{\lambda_0}} R_i\right]+\mathbb{E}\left[\frac{\partial\widehat{P}}{\partial\lambda_0}\right] \\ &\quad +\mathbb{E}\left[\widehat{P}\sum_{i=1}^{N_T^{\lambda\sup}}\frac{\partial R_i}{\partial\theta}\prod_{j\neq i}R_j\right].\end{aligned}$$

The pathwise sensitivities are

$$\begin{aligned}\frac{\partial\widehat{S}_{n+1}^-}{\partial\theta} &= \frac{\partial\widehat{S}_n}{\partial\theta}F_n+\widehat{S}_n\frac{\partial F_n}{\partial\theta}, \\ \frac{\partial\widehat{S}_{n+1}}{\partial\theta} &= \frac{\partial\widehat{S}_{n+1}^-}{\partial\theta}+\frac{\partial\widehat{S}_{n+1}^-}{\partial\theta}1_{\{\frac{1}{2}>U_i\}}(Y_i-1)\text{ if }t_{n+1}=\tau_i.\end{aligned}$$

In the cases of delta, vega and the sensitivity to the rate, we have

$$\frac{\partial F_n}{\partial\theta}=\begin{cases}-\frac{\partial\lambda(\widehat{S}_n^-,t_n)}{\partial S_0}m\,h_n, & \theta=S_0\ ; \\ -\frac{\partial\lambda(\widehat{S}_n^-,t_n)}{\partial\sigma}m\,h_n+\Delta W_n+\sigma(\Delta W_n^2-h_n), & \theta=\sigma\ ; \\ -\frac{\partial\lambda(\widehat{S}_n^-,t_n)}{\partial\lambda_0}\ mh_n, & \theta=\lambda_0.\end{cases}$$

According to the intensity function form,

$$\begin{aligned}\frac{\partial\lambda(\widehat{S}_n^-,t_n)}{\partial S_0} &= -\frac{2\lambda_0 C\widehat{S}_n^-}{\left(1+C\left(\widehat{S}_n^-\right)^2\right)^2}\frac{\partial\widehat{S}_n^-}{\partial S_0}; \\ \frac{\partial\lambda(\widehat{S}_n^-,t_n)}{\partial\sigma} &= -\frac{2\lambda_0 C\widehat{S}_n^-}{\left(1+C\left(\widehat{S}_n^-\right)^2\right)^2}\frac{\partial\widehat{S}_n^-}{\partial\sigma}; \\ \frac{\partial\lambda(\widehat{S}_n^-,t_n)}{\partial\lambda_0} &= \frac{1}{1+C\left(\widehat{S}_n^-\right)^2}-\frac{2\lambda_0 C\widehat{S}_n^-}{\left(1+C\left(\widehat{S}_n^-\right)^2\right)^2}\frac{\partial\widehat{S}_n^-}{\partial\lambda_0},\end{aligned}$$

where $\frac{\partial\widehat{S}_n}{\partial\lambda_0}$ are computed pathwise.

For the Radon-Nikodym derivatives,

$$\frac{\partial R_i}{\partial\theta}=\begin{cases}\frac{2}{\lambda_0}\frac{\partial\lambda(\widehat{S}_i^-,t_i)}{\partial\theta}, & \text{if } U_i<\frac{1}{2}\quad ; \\ -\frac{2}{\lambda_0}\frac{\partial\lambda(\widehat{S}_i^-,t_i)}{\partial\theta}, & \text{if } U_i\geq\frac{1}{2}\quad .\end{cases},$$

To compute $\partial R_i^f/\partial\theta$, we recall previous expressions of $\frac{\partial\lambda(\widehat{S}_i^-,t_i)}{\partial\theta}$.

$$\frac{\partial R_i^f/\partial\theta}{R_i^f} = \begin{cases} \frac{\partial\lambda(\widehat{S}_i^-,t_i)}{\partial\theta}\lambda(\widehat{S}_i^-,t_i)^{-1}, & \text{if } U_i < \frac{1}{2} \quad ; \\ -\frac{\partial\lambda(\widehat{S}_i^-,t_i)}{\partial\theta}\left(\lambda_0 - \lambda(\widehat{S}_i^-,t_i)\right)^{-1}, & \text{if } U_i \geq \frac{1}{2} \quad . \end{cases}$$

This example tells us that using the thinning with a change of measure, the perturbation of the intensity function is transferred to the Radon-Nikodym derivatives.

4.4 Variance reduction in the presence of jumps

4.4.1 Importance sampling for particular payoffs

Suppose we have a deep out-the-money (OTM) digital put option, e.g. underlying $S = 100$ with strike $K = 30$. In a Merton model with parameters $T=1$, $r=0.05$, $\sigma=0.2$, $a=-0.1$, $b=0.2$, $\lambda=0.1$, which means on average every 10 samples there is a sample having a jump within one-year maturity. Here the parameters are chosen such that the compensator $m = \mathbb{E}[Y_i] - 1 = \exp(a + b^2/2) - 1 < 0$. It turns out that a large portion of samples have zero contribute to the payoff if we simulate the path directly, leading to a relatively large variance of the estimator. In this specific case, we have two potential importance sampling treatments to reduce variance:

- Change the size of a jump so that once a jump happens, the sample is more likely to be below the strike. In particular we change the drift parameter a of the log-normal distribution of the jump magnitude.

  Assume A_i is the Normal random variable generated for the jump amplitude at time τ_i, the Radon-Nikodym derivative of this approach is the product of

$$\begin{aligned} R_i &= \exp\left(-\frac{(A_i - a)^2}{2b^2}\right) / \exp\left(-\frac{(A_i - a')^2}{2b^2}\right) \\ &= \exp\left(-\frac{(2\,(a - a')\,A_i + a'^2 - a^2)}{2b^2}\right). \end{aligned}$$

  where a' is the new drift parameter.

- Change the frequency of jumps so that the cumulative effect of jumps could eventually causes the option to be in-the-money at maturity.

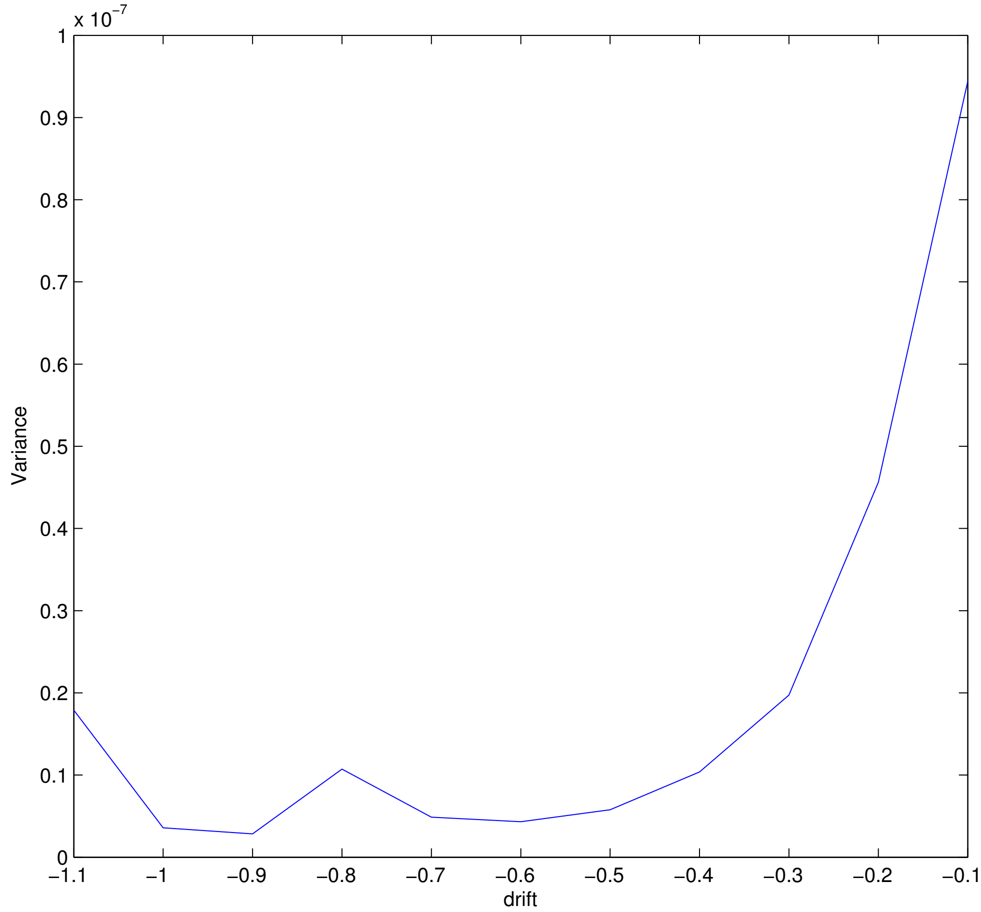

Figure 4.5: Change of jump drift for deep OTM digital put

The Radon-Nikodym derivative of this approach is

$$R = \left(\frac{\lambda}{\lambda_1}\right)^{N_T} \exp\left(\lambda_1 N_T - \lambda N_T\right).$$

where λ_1 is the new intensity of jumps.

4.4.1.1 Numerical result

By numerical experiments, we find the optimal a' and λ_1 so that the variance of new estimator is minimised.

The variance of the estimator is estimated by using 10^6 samples.

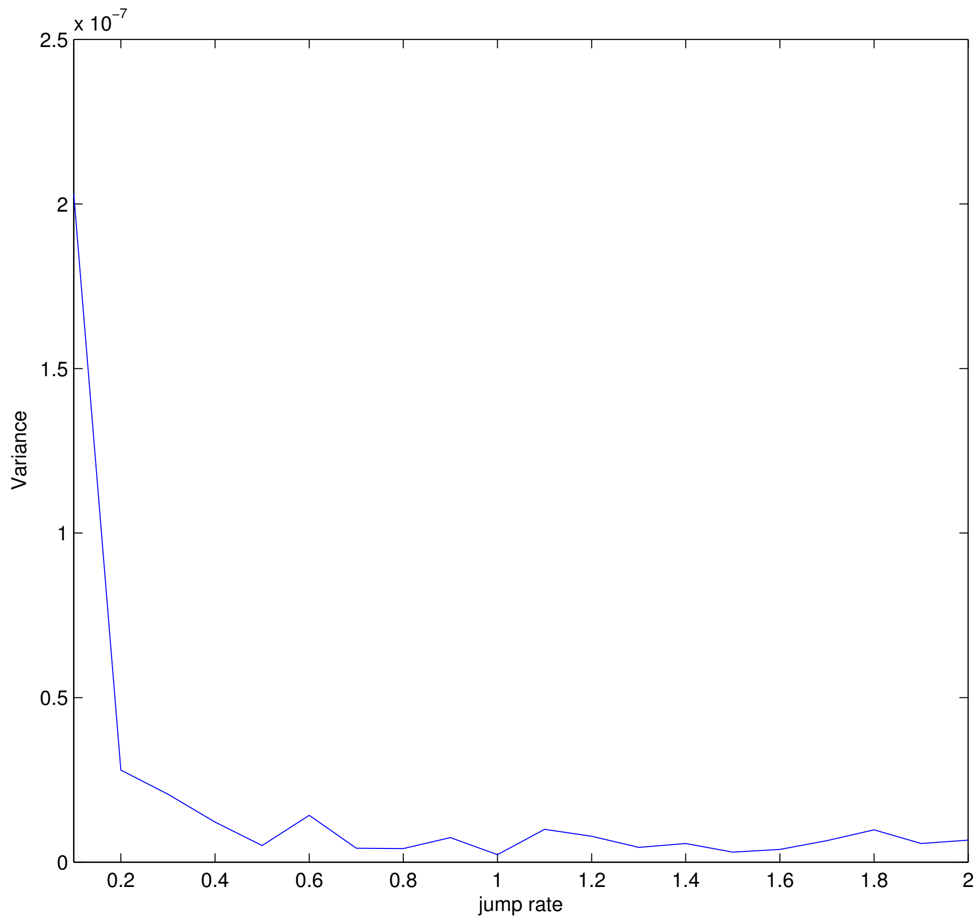

Figure 4.6: Change of intensity for deep OTM digital put

Figure 4.5 shows how the variance of the estimator changes as we adjust the target drift a'. It can be seen that as long as the drift is shifted below -0.5, the variance attains approximate minimum of 3×10^{-8}, equivalent to a saving factor of about 30.

Figure 4.6 shows that by increasing the jump rate the accuracy of estimator is greatly improved. It can be seen that as soon as the rate is increased above 0.4, the variance is reduced by a factor of about 20.

Chapter 5

Multilevel Monte Carlo Option Pricing In Exponential Lévy Models

5.1 Introduction to Lévy processes

We have already discussed using multilevel methods for computing option values in finite-activity jump-diffusion models. Apart from variable jump rate, another direction to generalise the constant rate jump-diffusion model is to allow infinite activity of jumps. The Lévy process is the building block to study such processes. Lévy processes are a class of stationary independent-increment stochastic processes whose paths are continuous in probability. In the following we use notations in [52].

Definition 5.1.1. *X_t is a Lévy process if it is an adapted stochastic process on $(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}$ with $X_0 = 0$ a.s., and has the following properties:*

1. Independent increments: for any $t_0 < t_1 < \cdots < t_n$, $X_{t_0},\ X_{t_1} - X_{t_0},\ \cdots,\ X_{t_n} - X_{t_{n-1}}$ are independent.
2. Stationarity: for arbitrary $h > 0$, $X_{t+h} - X_t$ has the same law for all t.
3. Continuity in probability: for any ε, $\lim_{h\to 0} \mathbb{P}\left(|X_{t+h} - X_t| > \varepsilon\right) = 0$.

We will only study its cádlág modification, of which the existence and uniqueness are proved in [52].

In terms of the empirical fitting of financial data time series, Lévy processes include a large class of distributions with flexibility to choose the number of parameters.

According to [14], Lévy type models can fit empirical properties of stock returns like heavy tails and asymmetric distribution of increments which are not true for a diffusion model.

5.1.1 Characteristics of Lévy processes

According to [52], Lévy decomposition states that each Lévy process can be represented as the sum of a deterministic drift, a Brownian Motion and a pure-jump process independent of the Brownian motion. In order to demonstrate this result, we will introduce the notation of random measure in the context of Lévy processes. Following Protter's notation, the jump of the cádlág Lévy process X_t is denoted by

$$\Delta X_t = X_t - X_{t-}.$$

For a Borel set E such that $0 \notin \overline{E}$, we can define the arrival time of jumps of X_t falling in E :

$$\begin{aligned} 0 &= T_0 < T_1 < T_2 < \dots \\ T_{n+1}(E) &= \inf \{ t > T_n(E) : \Delta X_t \in E \}. \end{aligned}$$

Then the counting process $\mu(\omega; E, t)$ associated with stopping times $T_n(E)$ is defined by

$$\mu(\omega; E, t) = \sum_{n \geq 1} 1_{\{T_n(E) \leq t\}}.$$

It counts the number of jumps within E up to time t. For each $\omega \in \Omega$, $\mu(\omega; \cdot, \cdot)$ is a Radon measure, i.e. σ-finite measure on $(E \times [0, \infty],\ \mathcal{B}(E \times [0, \infty]))$.

For any finite Borel function f, we have the integral with respect to μ:

$$\int_{E \times [0,t]} f(z, s) \mu(\omega; \mathrm{d}z, \mathrm{d}s) = \sum_{0 < s \leq t} f(\Delta X_s, s) 1_{\{\Delta X_s \in E\}}.$$

The Lévy measure ν is defined by

$$\nu(E) = \mathbb{E}\left[\int_0^1 \int_{z \in E} \mu(\mathrm{d}z, \mathrm{d}s)\right] = \mathbb{E}\left[\sum_{0 < s \leq 1} 1_{\{\Delta X_s \in E\}}\right].$$

In the time-homogeneous f case where $f(z,s) = f(z)$, $\int_0^t \int_{z\in E} f(z,s)\mu(\mathrm{d}z,\mathrm{d}s) = \int_{z\in E} f(z)\mu_t(\mathrm{d}z)$; it has an important property that for any bounded measurable function f,

$$\int_0^t \int_{z\in E} f(z,s)\mu(\mathrm{d}z,\mathrm{d}s) - \int_0^t \int_{z\in E} f(z,s)\nu(\mathrm{d}z)\mathrm{d}s$$

is a martingale with respect to $\mathcal{F}_t$.

Now we present the decomposition theorem.

Theorem 5.1.2 (decomposition theorem)**.** *If X_t is a scalar Lévy process, there exists a triple (m,σ,ν) where $m \in \mathbf{R}$, B_t is a standard Brownian Motion and ν is a Radon measure called the Lévy measure, satisfying the integrability condition:*

$$\int \left(|z|^2 \wedge 1\right)\nu(\mathrm{d}z) < \infty.$$

$$X_t = mt + \sigma B_t + \int_0^t \int_{|z|\geq 1} z\mu(\mathrm{d}z,\mathrm{d}s) + \int_0^t \int_{|z|<1} z\left(\mu(\mathrm{d}z,\mathrm{d}s) - \nu(\mathrm{d}z)\mathrm{d}s\right).$$

Moreover, for any Borel set E such that $0 \notin \overline{E}$, B_t is independent of the jump process $\mu(E,[0,t]) = \int_0^t \int_{z\in E} \mu(\mathrm{d}z,\mathrm{d}s)$.

This states that Lévy processes can be decomposed into three parts: a drifted Brownian Motion, a finite activity compound Poisson process with large jumps and an infinite activity compensated jump process which comes from the small jumps. It is worth mentioning that the infinite jump part of the decomposition is a martingale.

The proof of the decomposition theorem is due to a theorem concerning characteristic functions of Lévy processes:

Theorem 5.1.3 (Lévy-Khintchine representation). *If X_t is a scalar Lévy process, there exists a unique triple (m,σ,ν) where $m \in \mathbf{R}$, ν is a Lévy measure satisfying the integrability condition, s.t.*

$$\phi_t(u) := \mathbb{E}\left[\exp\left(iuX_t\right)\right] = \exp\left(t\left(ium - \frac{\sigma^2}{2}u^2 + \int (e^{iuz} - 1 - iuz1_{|z|\leq 1})\nu(\mathrm{d}z)\right)\right).$$

The triple (m,Σ,ν) is called the Lévy characteristic triple. If the process has finite variation, i.e. $\int_{|x|\leq 1} |x|\nu(\mathrm{d}x) < \infty$, the representation degenerates to

$$\phi(u) = \exp\left(ium_0 - \frac{\sigma^2}{2}u^2 + \int_R (e^{iuz} - 1)\nu(\mathrm{d}z)\right).$$

where $m_0 = m - \int_{|z|\leq 1} iuz\ \nu(\mathrm{d}z)$.

Remark 5.1.4. *The Lévy-Khintchine representation can be interpreted by observing the characteristic functions of the compound Poisson process and Brownian motion.*

For a compound Poisson process $X_t = \sum_{i=1}^{N(t)} Y_i$*,*

$$\mathbb{E}[e^{iuX_t}] = \exp\left(\lambda t \int_R \left(e^{iux} - 1\right) f(\mathrm{d}z)\right),$$

where λ *is the Poisson jump rate and* f *is the p.d.f. of* Y*.*

For a Brownian motion B_t*,*

$$\mathbb{E}[e^{iuB_t}] = e^{-\frac{1}{2}u^2 t}.$$

Remark 5.1.5. *The Lévy-Khintchine representation also unveils another property of Lévy processes: infinite divisibility.*

Definition 5.1.6. *A random variable* X *is infinitely divisible if for any integer* n *it can be split as*

$$X = Y_1 + Y_2 + \cdots + Y_n,$$

where the Y_i *are i.i.d. random variables. Equivalently, its characteristic function satisfies*

$$\phi(u) = \mathbb{E}[e^{iuX}] = (\psi(u))^n,$$

where $\psi(u)$ *is also a characteristic function.*

Remark 5.1.7. *The financial interpretation of infinite divisibility is that the fluctuation of an asset can be ascribed to the behavior of many informed traders who share the same information, which can be modeled as i.i.d. random variables.*

We now introduce the index of jump activity, the Blumenthal-Getoor (BG) index of X_t :

$$\beta \equiv \inf\left\{\alpha > 0 : \int_{|x|<1} |x|^\alpha \, \nu(\mathrm{d}x) < \infty\right\}$$

There is extensive research on this topic. From an econometric perspective, [1] does statistical inference of a jump activity index generalized from the BG index for a scalar semimartingale from high-frequency data, in which there is a very clear demonstration of the BG index. The BG index ranges from 0 to 2. A finite jump activity process has $\beta = 0$. For a finite variation process, $\beta \leq 1$; for an infinite variation process, $\beta \geq 1$. The diffusion process has $\beta = 2$.

5.1.2 Examples of Lévy processes

5.1.2.1 Variance Gamma (VG) processes

The Variance Gamma (VG) process with a parameter set (σ, θ, κ) (note that this is not a Lévy triplet) is the Lévy process where X_t has the characteristic function

$$\mathbb{E}[\exp(iuX_t)] = (1 - iu\theta\kappa + \frac{1}{2}\sigma^2 u^2 \kappa)^{-t/\kappa}.$$

The Lévy measure of the VG process is ([14] p117)

$$\nu(x) = \frac{1}{\kappa |x|} e^{A - B|x|} \quad \text{with} \quad A = \frac{\theta}{\sigma^2} \quad \text{and} \quad B = \frac{\sqrt{\theta^2 + 2\sigma^2/\kappa}}{\sigma^2}.$$

One advantage of the Variance Gamma process is that its additional parameters enable fitting to the skewness and kurtosis of the stock returns ([14]).

The Variance Gamma process is a finite variation process with BG index $\beta = 0$. It can be represented as the difference of two increasing processes, more specifically, two independent Gamma processes:

$$X_t = G_t(C, M) - G_t(C, G).$$

where

$$\begin{aligned} C &= 1/\kappa > 0, \\ G &= \sqrt{\frac{1}{4}\theta^2\kappa^2 + \frac{1}{2}\sigma^2\kappa} - \frac{1}{2}\theta\kappa > 0, \\ M &= \sqrt{\frac{1}{4}\theta^2\kappa^2 + \frac{1}{2}\sigma^2\kappa} + \frac{1}{2}\theta\kappa > 0. \end{aligned}$$

A Gamma process $G_t(\lambda, c)$ has the Gamma distribution with density

$$p_t(x) = \frac{\lambda^{ct}}{\Gamma(ct)} x^{ct-1} e^{-\lambda t}.$$

We also have a subordinator representation of the VG process:

$$X_t = \theta G_t + \sigma B_{G_t},$$

where the subordinator G_t is a Gamma process with parameter $(1/\kappa, 1/\kappa)$. The subordinator representation corresponds to the Dubins-Schwarz Theorem which states that any continuous local martingale starting at 0 is a time-change Brownian motion (see [52] chapter II Theorem 42).

The simulation of the VG process is straightforward since sampling from the Gamma distribution is well established in software packages. The subordinator representation is used in our simulation algorithm, because the random number generator for the Normal distribution in the software package we use (Matlab) is about 20 times faster than the one for the Gamma distribution.

5.1.2.2 Normal Inverse Gaussian (NIG) processes

The Normal Inverse Gaussian (NIG) process with a parameter set (σ, θ, κ) is a Lévy process with the characteristic function

$$\mathbb{E}[\exp(iuX_t)] = \exp\left(\frac{t}{\kappa} - \frac{t}{\kappa}\sqrt{1 - 2iu\theta\kappa + \kappa\sigma^2 u^2}\right).$$

The Lévy measure of NIG process is

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

$K_n(x)$ is the modified Bessel function of the second kind ([14] p117). It has asymptotic expansion as $x \to 0$:

$$K_1(x) \sim \frac{1}{x} + \mathcal{O}(1). \tag{5.1}$$

Hence its BG index is $\beta = 1$. Also it has asymptotic expansion as $x \to \infty$:

$$K_1(x) \sim e^{-x}\sqrt{\frac{\pi}{2\,|x|}}\left(1 + \mathcal{O}\left(\frac{1}{|x|}\right)\right). \tag{5.2}$$

In terms of simulation, the NIG process can be represented by

$$X_t = \theta I_t + \sigma B_{I_t},$$

where the subordinator I_t is an Inverse Gaussian process with parameters $\lambda = \frac{1}{\kappa},\ \mu = 1$. An Inverse Gaussian process with (λ, μ) has a density of the form

$$p_t(x) = \sqrt{\frac{\lambda}{2\pi x^3}}\ t\ \exp\left(-\frac{\lambda\left(x - \mu t\right)^2}{2\mu^2 x}\right).$$

To generate its sample paths, we only have to generate samples of the Inverse Gaussian and Brownian motion. We use the algorithm in p184 of [14] to generate Inverse Gaussian samples.

5.1.2.3 α-stable and tempered α-stable processes

The scalar α-stable process has a Lévy measure of the form ([14]):

$$\nu(x) = \frac{A}{x^{\alpha+1}} 1_{\{x>0\}} + \frac{B}{|x|^{\alpha+1}} 1_{\{x<0\}}$$

for $0 < \alpha < 2$ and some non-negative number A and B.

We follow [14] to discuss another parameterisation of α-stable process with characteristic function

$$\begin{aligned} \mathbb{E}[\exp(iuX_t)] &= \exp\left\{-(A+B)^{\alpha}|u|^{\alpha}\left(1 - i\tfrac{A-B}{A+B}\operatorname{sgn}(u)\tan\tfrac{\pi\alpha}{2}\right) + i\mu u\right\}, \text{ if } \alpha \neq 1, \\ \mathbb{E}[\exp(iuX_t)] &= \exp\left\{-(A+B)|u|\left(1 - i\tfrac{A-B}{A+B}\tfrac{2}{\pi}\operatorname{sgn}(u)\log|u|\right) + i\mu u\right\}, \text{ if } \alpha = 1, \end{aligned} \tag{5.3}$$

where μ is the drift parameter which we set to be zero in the following, $\operatorname{sgn}(u) = |u|/u$ if $u \neq 0$ and $\operatorname{sgn}(0) = 0$; see [47]. Under this specification of parameterisation, according to [8], we can determine that only when $A = 0$ does the process have a finite exponential moment $\mathbb{E}[\exp(uX_t)]$. When $A = 0$, there are no positive jumps and the process is said to be spectrally negative.

Sometimes people consider tempered α-stable processes, whose Lévy measure is

$$\nu(x) = \frac{A}{x^{\alpha+1}} e^{-\lambda_+ x} 1_{\{x>0\}} + \frac{B}{|x|^{\alpha+1}} e^{\lambda_- x} 1_{\{x<0\}} \tag{5.4}$$

for $0 < \alpha < 2$ and some non-negative numbers $A,\ B,\ \lambda_+,\ \lambda_-$. This kind of process is included in the class of Lévy processes for which we do numerical analysis.

The BG index of the (tempered) α-stable process is equal to α. Hence it is a good example to examine the numerical behavior of the algorithm for various BG indices.

Generating sample paths of α-stable processes can be done by the algorithm in [9]. Here we adopt the implementation in MATLAB by Mark Veillette

(http://math.bu.edu/people/mveillet/html/alphastablepub.html). Direct simulation of increments of tempered α-stable processes is not known so we only include numerical results of α-stable processes.

5.2 Exponential Lévy models

In the literature ([54, 13]) a class of tractable models is constructed based on the assumption that asset returns follow a Lévy process. For a continuous-time model, it means that log asset price follows a Lévy process. Let X_t be a real-valued (m, σ, ν)-Lévy process. The asset price process

$$S_t = S_0 \exp(X_t) \tag{5.5}$$

is called an exponential Lévy model. Under the physical measure, parameters of a Lévy process can be fitted to asset return time series data. Statistical tests shows that the fitting has advantages compared with the diffusion model; cf. [54].

It's worth noting that this class of model is not always equivalent to the class of Lévy-driven SDEs studied in [17] and [16] if the Lévy processes has jumps less than or equal to -1. According to Section 4.2 in [56], the two classes are only equivalent under the condition that the Lévy measure has zero support on $(-\infty, -1)$. In our cases, none of the VG, NIG or spectrally negative α-stable process satisfies this condition.

5.2.1 Pricing in exponential Lévy models

Exponential Lévy models are used in option pricing. We suppose that $e^{-rt}S_t$ is a martingale under a risk-neutral measure $\mathbb{Q}$. Under measure $\mathbb{Q}$, X_t is an $(\tilde{m}, \sigma, \nu)$-Lévy process. According to [14], since the Predictable Representation Property only holds for diffusion or Poisson processes, the market is incomplete in a model driven by general Lévy processes. Hence there are infinitely many pricing measures under which the discounted asset price is a martingale. For details about hedging and pricing issues in the context of incomplete markets we refer to Chapter 8-9 of [14].

For the ease of computation, we follow the mean-correcting pricing measure in [54]. We shall derive the parameter relationship between the physical and mean-correcting pricing measures. By the martingale property we have

$$\mathbb{E}_{\mathbb{Q}}[e^{-rt}S_t] = S_0\mathbb{E}_{\mathbb{Q}}[e^{X_t - rt}] = S_0.$$

This implies

$$\tilde{m} - r + \frac{\sigma^2}{2} + \int \left(e^z - 1 - z1_{|z|\leq 1}\right) \nu(\mathrm{d}z) = 0.$$

On the other hand, recall the definition of characteristic function:

$$\phi_1(-i) = \exp\left(m + \frac{\sigma^2}{2} + \int \left(e^z - 1 - z1_{|z|\leq 1}\right) \nu(\mathrm{d}z)\right).$$

So,

$$\tilde{m} = m + r - \log \phi_1(-i)$$

is the new drift of X_t under the mean-correcting pricing measure Q. Schoutens suggests calculating $\tilde{m}$ by estimating parameters of the Lévy process assuming $m = 0$. Under such an assumption, according to the forms of characteristic functions of VG and NIG and Proposition 1 of [8], for VG, NIG and α-stable processes we have:

$$\begin{aligned} \tilde{m}_{VG} &= r + \frac{1}{\kappa}\log(1 + \theta\kappa - \frac{1}{2}\sigma^2\kappa). \\ \tilde{m}_{NIG} &= r - \frac{1}{\kappa} + \frac{1}{\kappa}\sqrt{1 + 2\theta\kappa - \kappa\sigma^2}. \\ \tilde{m}_{\alpha} &= r + (A+B)^{\alpha} \sec\frac{\pi\alpha}{2}. \end{aligned}$$

The option price is given by the expected value of the discounted payoff $P = f(S)$

$$\mathbb{E}_Q\left[e^{-rT}P\right].$$

The calibration performance of exponential Lévy models to S&P 500 option prices data can be found in [54] and [8]. The Fourier Transform method is one way to calibrate the model, see [14]. It provides an analytic formula for the European option so it is very efficient to use it as a calibration tool. However, for lookback and barrier options in models driven by VG, NIG and α-stable processes, there is no closed-form of the Wiener-Hopf factorisation so the Fourier Transform method is computationally expensive. This motivates the use of Monte Carlo for pricing those payoffs.

5.3 Multilevel Monte Carlo in exponential Lévy models

Although the increments of those Lévy processes used in practice can be directly sampled efficiently, for exotic options like Asian, lookback and barrier options, there is no simple exact payoff simulation technique. For the Asian payoff, conventional quadrature works well in the (jump-) diffusion setting. People have investigated various simulation methods to deal with lookback and barrier payoffs. For example, [23] develops an adaptive Monte Carlo method for functionals of killed Lévy processes with controlled bias. Small-time asymptotic expansions of the exit probability are given with a computable error bounds. For evaluating the exit probability when the barrier is close to the starting point of the process, e.g. $\mathbb{P}(\sup_{0\leq t\leq 1} X_t \leq 10^{-2}),\ X_0 = 0$, this algorithm outperforms uniform discretisation significantly; see the numerical example of a Cauchy process (symmetric $\alpha-$stable process with $\alpha = 1$) in Section 6 of [23]. On the other hand, [46] develops a novel Wiener-Hopf Monte-Carlo method to generate the joint distribution of $(S(T), \min_{0\leq t\leq T} S(t))$. This simulation technique is further extended to a multilevel setting in [22]. The method currently cannot be directly applied to VG, NIG and $\alpha-$stable processes. For the Lévy-driven SDEs model, [17, 16] show upper bounds on the worst case computational complexity of a class of MLMC algorithms for payoffs Lipschitz continuous with respect to the supremum norm. The BG index has an impact on the performance of this class of algorithms truncating small jumps or approximating them by diffusions. In [16], the bound on the the worst case multilevel computational complexity for any Lipschitz path-dependent payoffs is $\mathcal{O}\left(\epsilon^{-6}\right)$ when $\beta \to 2$. However, as $\beta \to 2$ the process gets closer to the diffusion process. In diffusion case, for a particular payoff like a Lipschitz function of the running maximum of the process (lookback type), a near optimal multilevel computational complexity $\mathcal{O}\left(\epsilon^{-2}\left(\log \epsilon\right)^2\right)$ is indicated by the first order variance convergence rate for a uniform discretisation. There should be a method that has similar performance to the diffusion case. Numerical experiments for lookback payoff using spectrally negative α-stable processes as $\alpha \to 2$ support this intuition.

In contrast to those advanced techniques, we take the discretely monitored maximum (uniform discretisation) of the Lévy process as the approximation of the running supremum. To reduce the bias caused by discretisation, people naturally want to extend the Brownian bridge sampling method. [53] argued to use a bridge sampling technique for the subordinator representation of VG and NIG processes, but this introduces an even larger bias due to discontinuities in the subordinator [5]. Nevertheless, the (uniformly) discretely monitored maximum serves as a good approximation. Moreover it is easy to implement with multilevel which improves its performance significantly. For evaluating the exit probability when the barrier is relatively far from the starting point of the process, e.g. $\mathbb{P}(\sup_{0\leq t\leq 1} X_t \leq 1),\ X_0 = 0$, to achieve the same RMS error of $\epsilon = 0.0001$, the multilevel cost is 1/30 of the cost of a single-level uniform discretisation. For the same problem, the computing time of the adaptive method in [23] just achieves a break-even point compared with a single-level uniform discretisation. In contrast to the general approach in [46], we aim for specific processes and get better complexity results for the specific processes we are looking at. This work differs from [17, 16] since there is an intrinsic difference between the

Lévy-driven SDEs model and the exponential Lévy model driven by a VG, NIG or α-stable process, as we addressed in the previous section.

We provide comparison of complexity between ours and methods in the literature. There are two ways to compare the complexity of algorithms. The first way presented in Theorem 1.3.1 estimates the expected computational cost to achieve a target RMS error. Another way analyses the convergence of the RMS error of an algorithm as a function of the cost. For comparison with [17, 16] and [46], we use the estimate of the RMS error ϵ in terms of expected value of the computational cost $\omega = \mathbb{E}[C]$ in Theorem 1.3.1. Compared to our previous notation $\mathcal{O}(\epsilon^{-r})$ computational complexity is equivalent to RMS error convergence of $\mathcal{O}(\omega^{-1/r})$.

[43] shows that for Lipschitz payoff depending on the final value of the solution of Lévy-driven SDEs, the RMS error is $\omega^{-\frac{3-\beta}{6-\beta}}$ if $\beta \geq \frac{3}{2}$. In the Lévy-driven SDEs model, if the Lévy process does not incorporate a Wiener process, [17] shows an upper bound $\omega^{-\min\left(\frac{1}{2},\left(\frac{1}{\beta}-\frac{1}{2}\right)\right)}$ and [16] shows an upper bound $\omega^{-\frac{4-\beta}{6\beta}}$ on the worst case RMS error of a class of MLMC algorithms for payoffs Lipschitz continuous with respect to the supremum norm.

For simulating a Lipschitz function of $(S(T), \min_{0\leq t\leq T} S(t))$, [22] obtains an RMS error of $\mathcal{O}\left(\omega^{-\frac{1}{4}}\right)$ for Lévy processes with infinite variation and $O\left(\omega^{-\frac{1}{3}}\right)$ for processes with bounded variation, robust in terms of BG index.

In the exponential Lévy model, using multilevel with uniform approximation simulating lookback payoff for a class of Lévy processes with the Lévy measure specified in Proposition 5.3.7, we have an RMS error of $\mathcal{O}\left(\omega^{-1/2}\right)$ when $\alpha \leq 1$. According to Proposition 5.3.7 and complexity Theorem 1.3.1, when $\alpha > 1$, our method has $\mathcal{O}\left(\epsilon^{-2-\frac{1-(1/\alpha)}{1/\alpha}}\right)$ computational complexity, equally $\mathcal{O}\left(\omega^{-\frac{1}{1+\alpha}}\right)$ RMS error convergence. For spectrally negative α-stable processes, we have $\mathcal{O}\left(\omega^{-\frac{1}{2}}\right)$ RMS error when $\alpha > 1$ by Proposition 5.3.12. Those comparisons are shown in Figure 5.1.

5.3.1 Numerical results

We have numerical results for lookback, barrier and Asian options for models using VG, NIG and α-stable processes. In the following $S_t = S_0 \exp(X_t)$, where X_t is a Lévy process. Under the mean-correcting pricing measure, calibrated model parameters we use for the VG process are $\sigma = 0.1213, \theta = -0.1436, \kappa = 0.1686$, according to the calibration to option price data in [48]. The parameters for the NIG process are

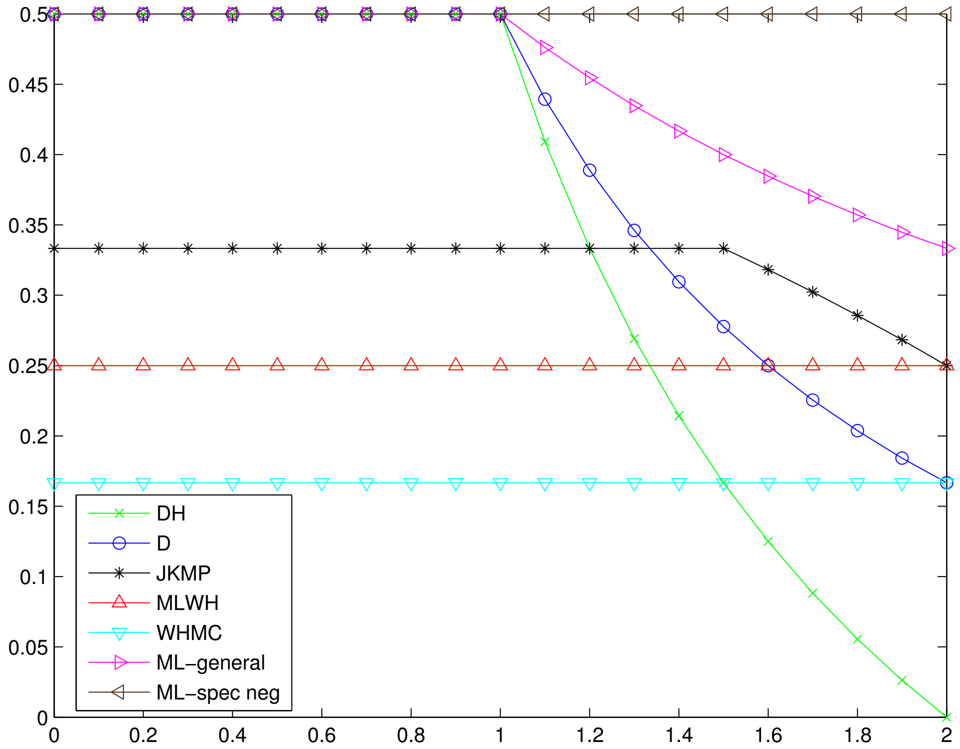

Figure 5.1: Order of RMS convergence with respect to the Blumenthal-Getoor index

$A = 6.1882, B = -3.8941, \delta = 0.1622$, according to the calibration on page 82 of [54]. The parameters for the α-stable process are $\alpha = 1.5597$, $A = 0$ and $B = 0.1486$ according to the calibration to option prices data on page 770 of [8]. Note that the α-stable process in the simulation is spectrally negative.

The option pricing parameters are chosen as $K = 100,\ T = 1,\ S_0 = 100,\ r = 0.05$ and the barrier $B = 115$. We modify Giles's matlab code used for [31] to produce performance figures:

- For the lookback option payoff, in the numerical experiment we test a call option with the payoff

$$P = \exp(-rT)\left(\max_{0\le t\le T} S(t) - K\right)^+.$$

We shall use the discretely monitored maximum as the approximation

$$\widehat{P} = \exp(-rT)\left(S_0 \max_{i=0,1,\dots,n} \exp\left(X_{\frac{i}{n}T}\right) - K\right)^+.$$

Suppose we use M fine timesteps per coarse timestep. In this work, $M = 2$. The ML estimate $\widehat{P}_l$ is given by setting $n = M^l$ in $\widehat{P}$. We denote $h_l = T/M^l$. The ML estimates of other payoffs are defined similarly.

Figures 1-3 show that we have computational savings of around a factor of 100 in the lookback case when RMS $\epsilon = 0.05$.

- For the Asian payoff, in the numerical test, we deal with an arithmetic Asian call option

  $$P = \exp(-rT)\ \max\left(0,\, \overline{S} - K\ \right).$$

  where

  $$\overline{S} = S_0\, T^{-1} \int_0^T \exp\left(X_t\right) \mathrm{d}t\ .$$

  Recall that in the diffusion case, $\int_0^T B_t\ \mathrm{d}t$ and B_T are jointly normal so $\int_0^T B_t\ \mathrm{d}t$ can be exactly simulated. However, for a general Lévy process it is not easy to directly sample the integral process. Hence we shall use the trapezoidal approximation

  $$\widehat{\overline{S}} := S_0 \sum_{j=0}^{n-1} \tfrac{1}{2}\, h\ \left(\exp\left(X_{jh}\right) + \exp\left(X_{(j+1)h}\right)\right),$$

  where $n = T/h$ is the number of timesteps. The ML estimate is defined by letting $n = 2^l$ in

  $$\widehat{P} = \exp(-rT)\ \max\left(0,\, \widehat{\overline{S}} - K\ \right).$$

  What the graphs indicate is that the trapezoidal approximation has a correction term $\widehat{P}_1 - \widehat{P}_0$ with a big variance compared to $\widehat{P}_0$, which leads to a reasonable savings factor of about 10 even if RMS $\epsilon = 0.005$.

- The barrier option we test is

  $$P = \exp(-rT)\, (S_T - K)^+\ \mathbb{1}_{\left\{\max_{0 \le t \le T} S(t) < B\right\}}.$$

  The discretely monitored approximation $\hat{m} = S_0 \max_{i=0,1,\ldots,n} \exp\left(X_{\frac{i}{n}T}\right)$ gives

  $$\widehat{P} = \exp(-rT)\, (S_T - K)^+\ \mathbb{1}_{\{\hat{m} < B\}}.$$

  The numerical result shows that the computational saving is substantial, with a factor of 10 if RMS $\epsilon = 0.005$. Since in the VG and NIG case, the variance of the correction term $\widehat{P}_l - \widehat{P}_{l-1}$ doesn't decrease when l is small, potentially its performance could be further enhanced by using an adaptive method [25]. In the α-stable model case, the variance of $\widehat{P}_l$ converges slower. Therefore the benefits from multilevel is limited.

The reason for the small fluctuation of numbers of samples N_l on levels greater than 5 is due to poor estimates of variance in the first step using a limited number of initial samples.

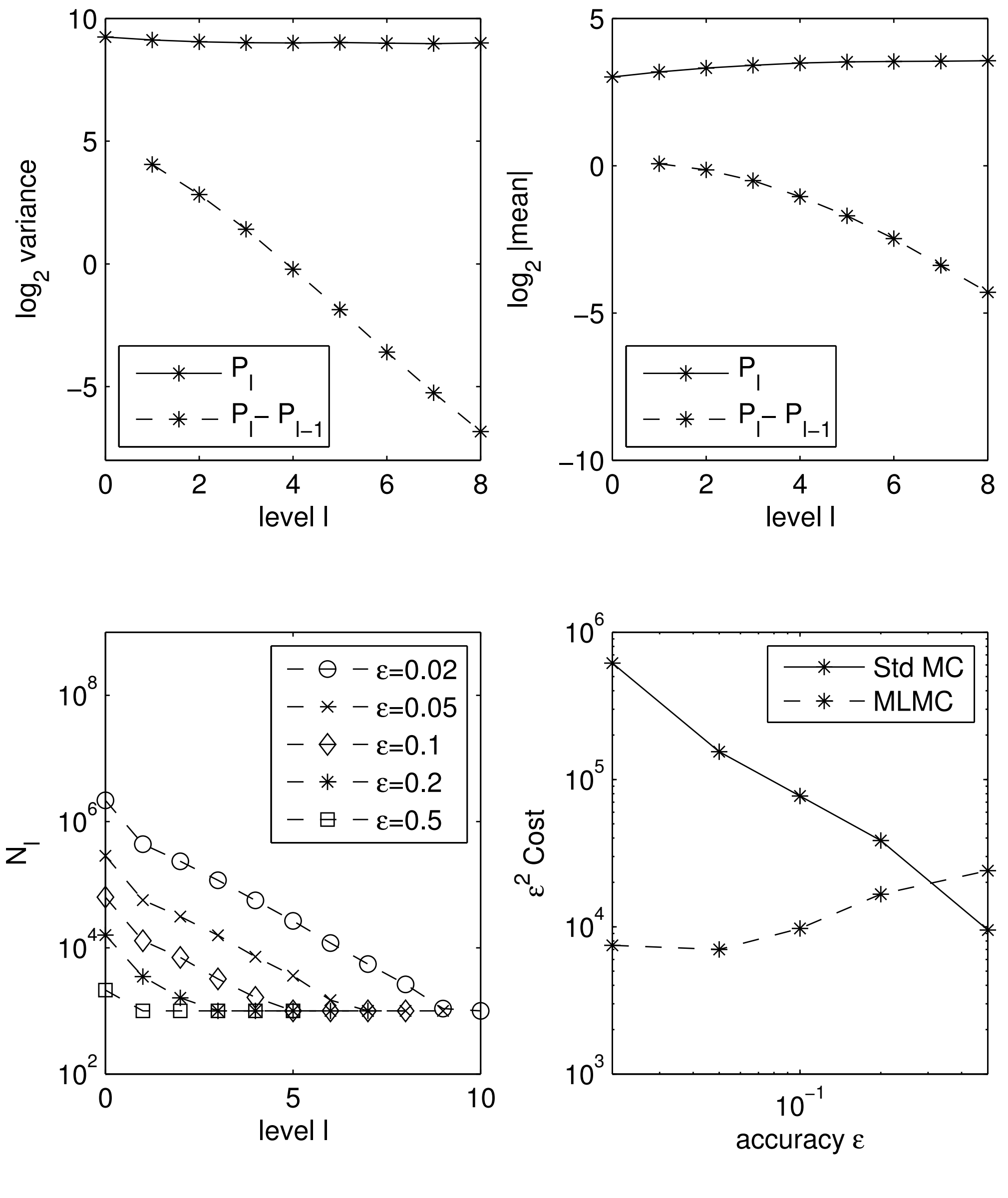

Figure 5.2: Lookback option in variance gamma model

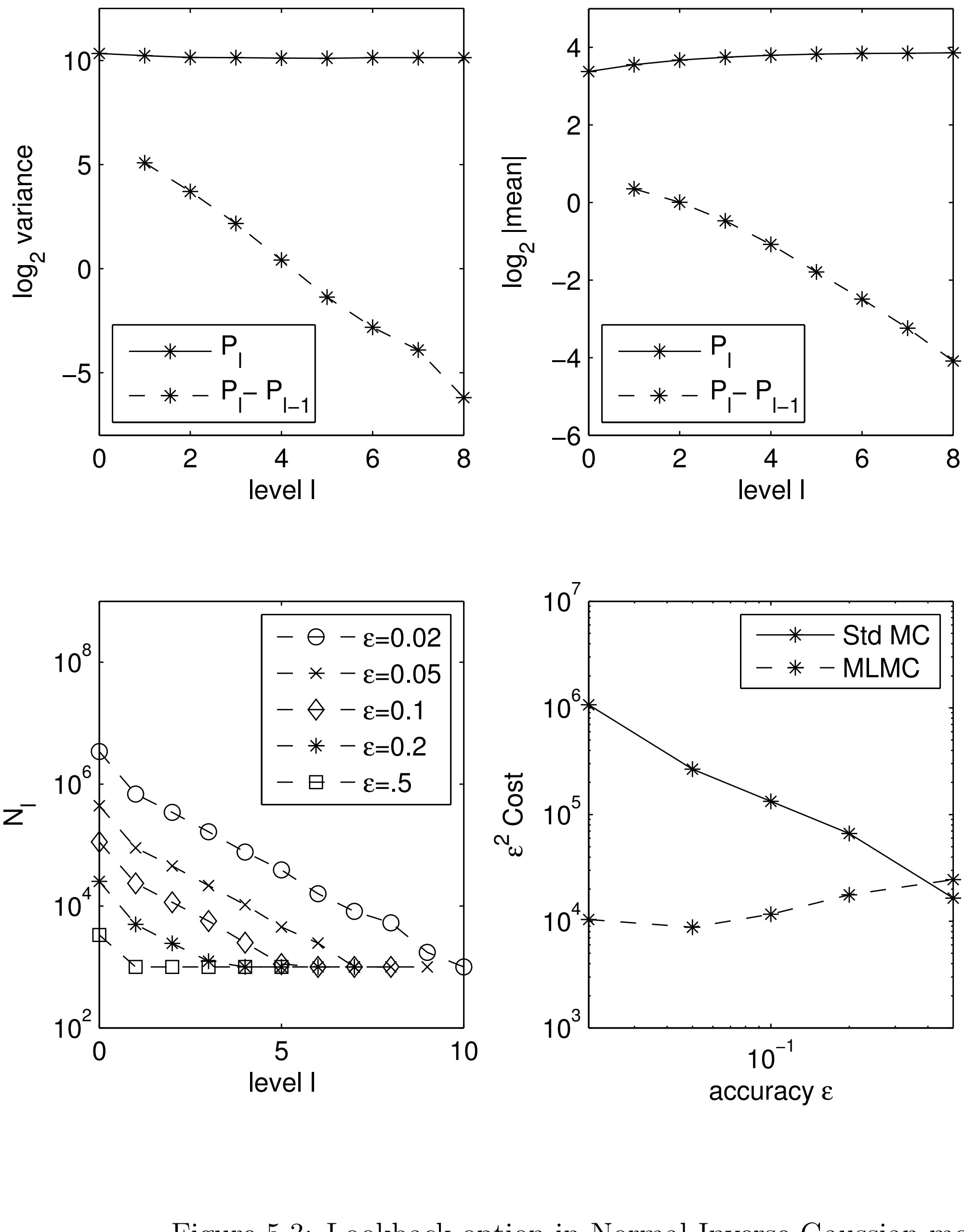

Figure 5.3: Lookback option in Normal Inverse Gaussian model

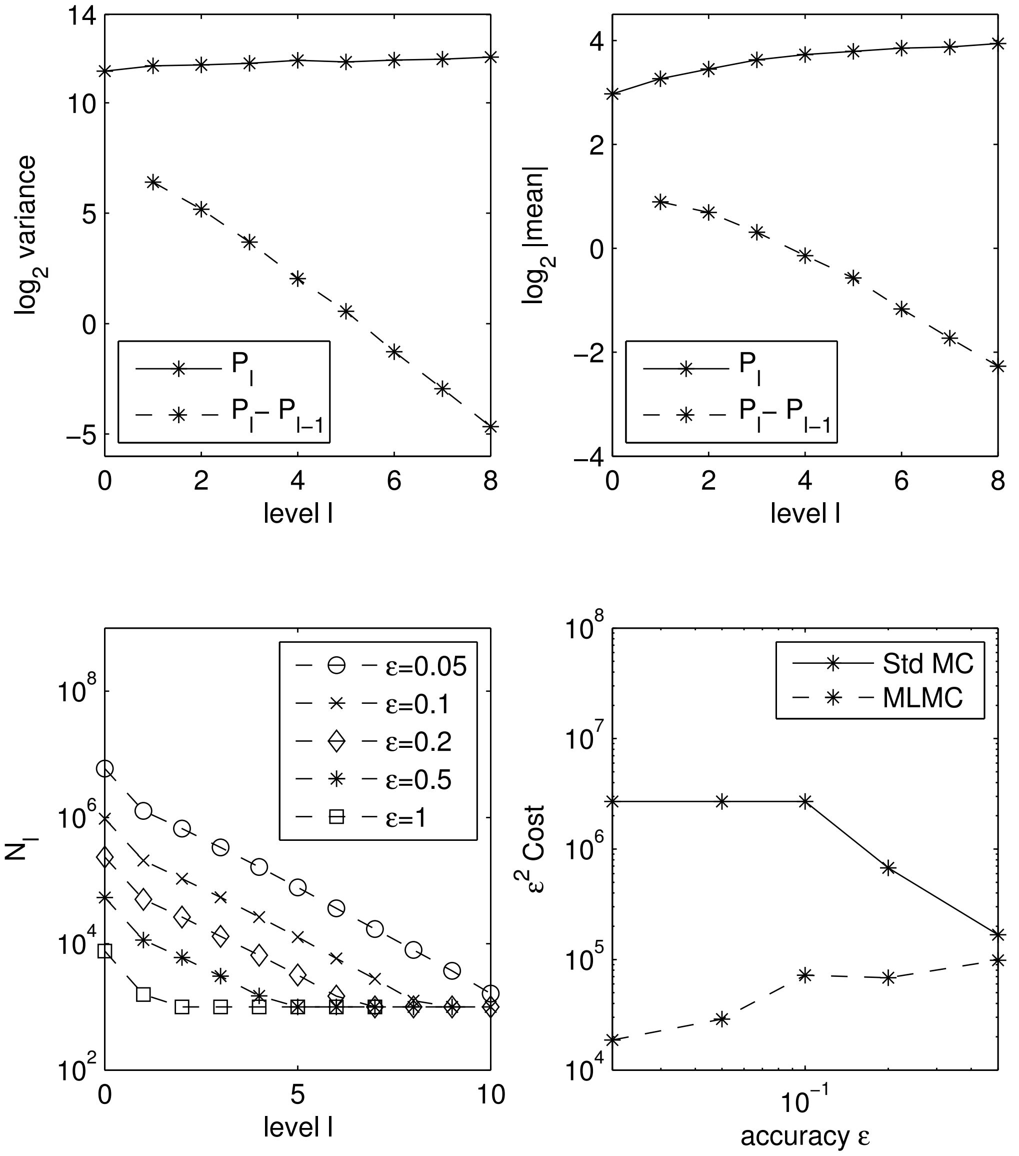

Figure 5.4: Lookback option in spectrally negative α-stable model

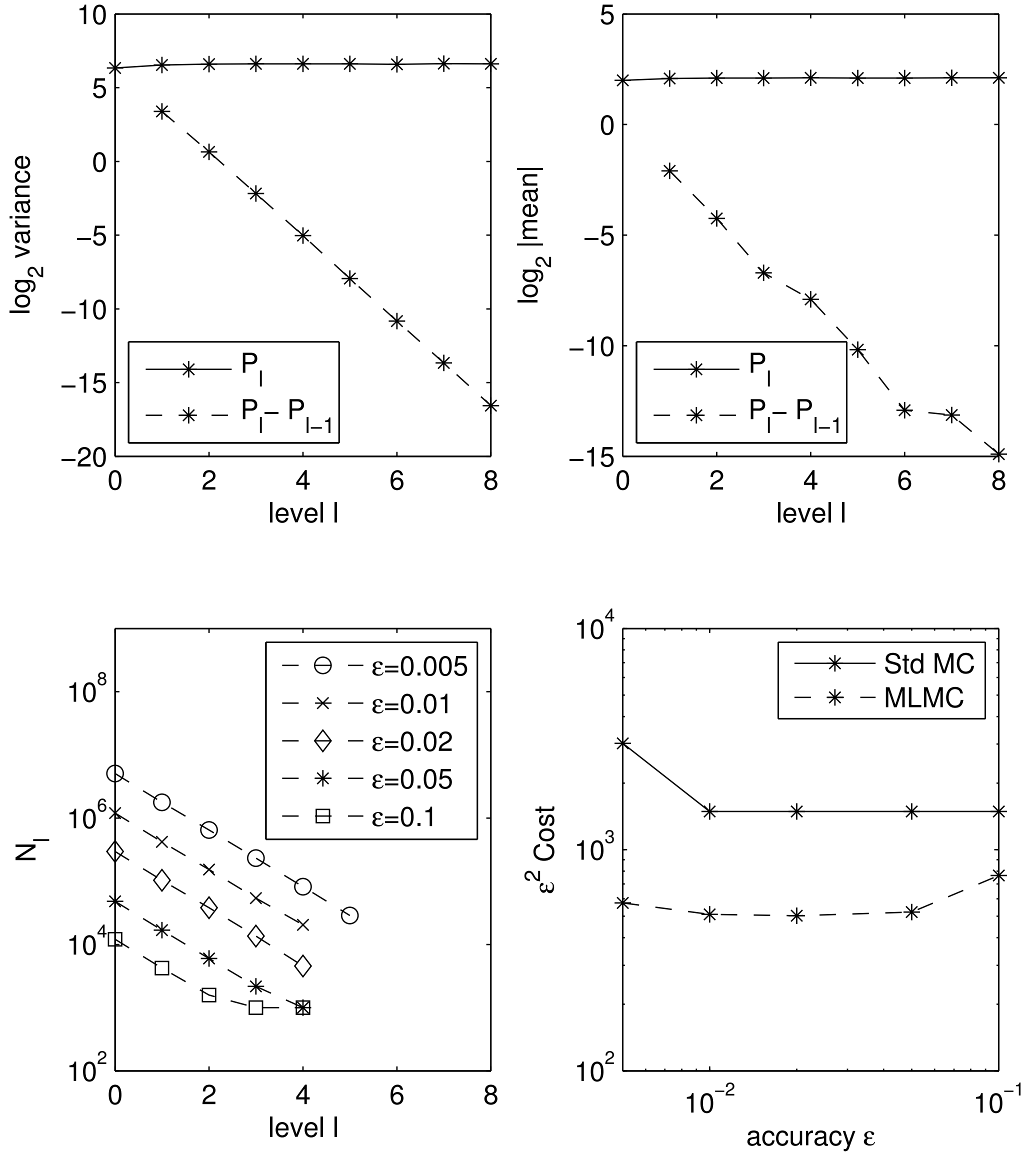

Figure 5.5: Asian option in variance gamma model

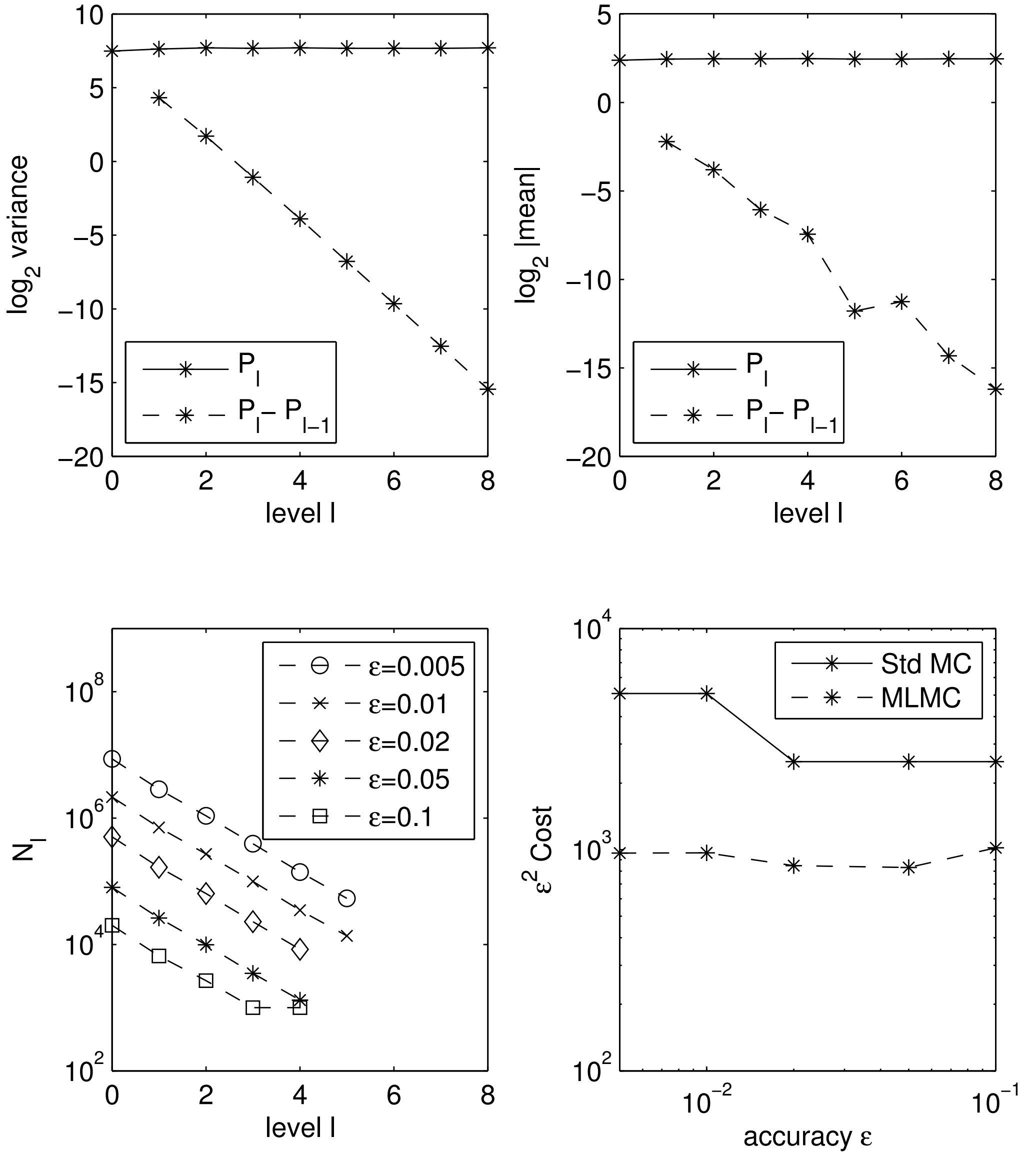

Figure 5.6: Asian option in Normal Inverse Gaussian model

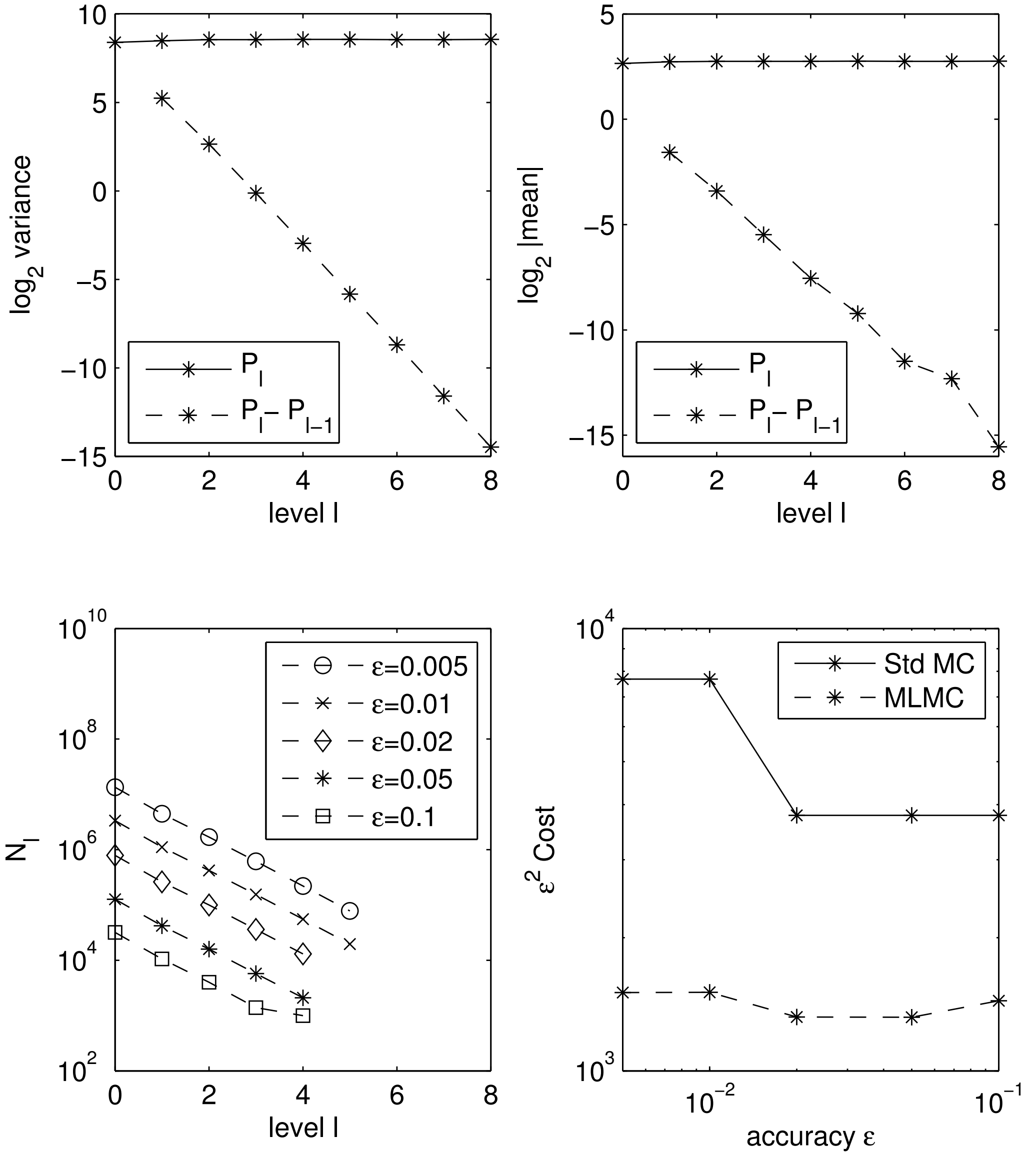

Figure 5.7: Asian option in spectrally negative α-stable model

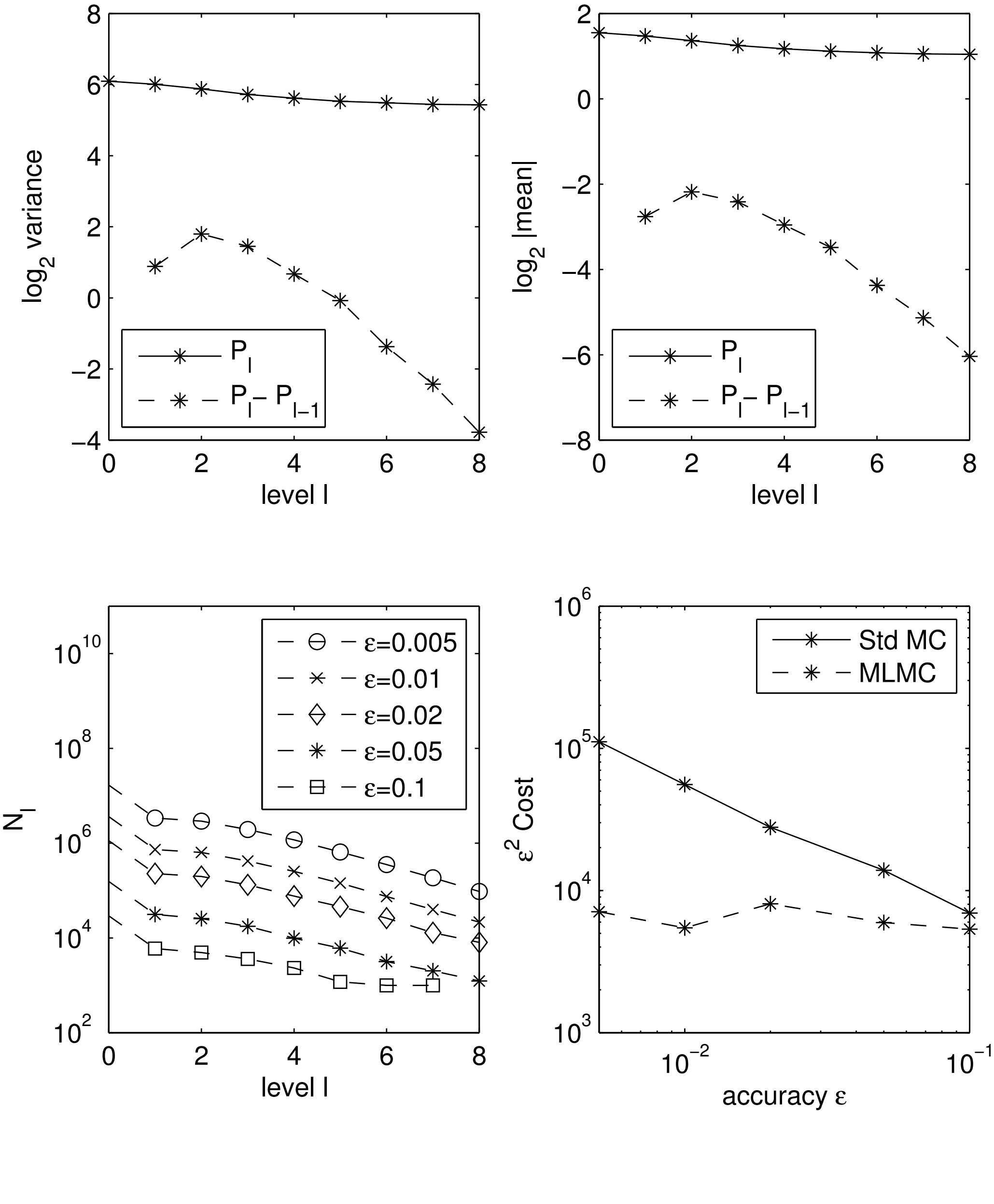

Figure 5.8: Barrier option in variance gamma model

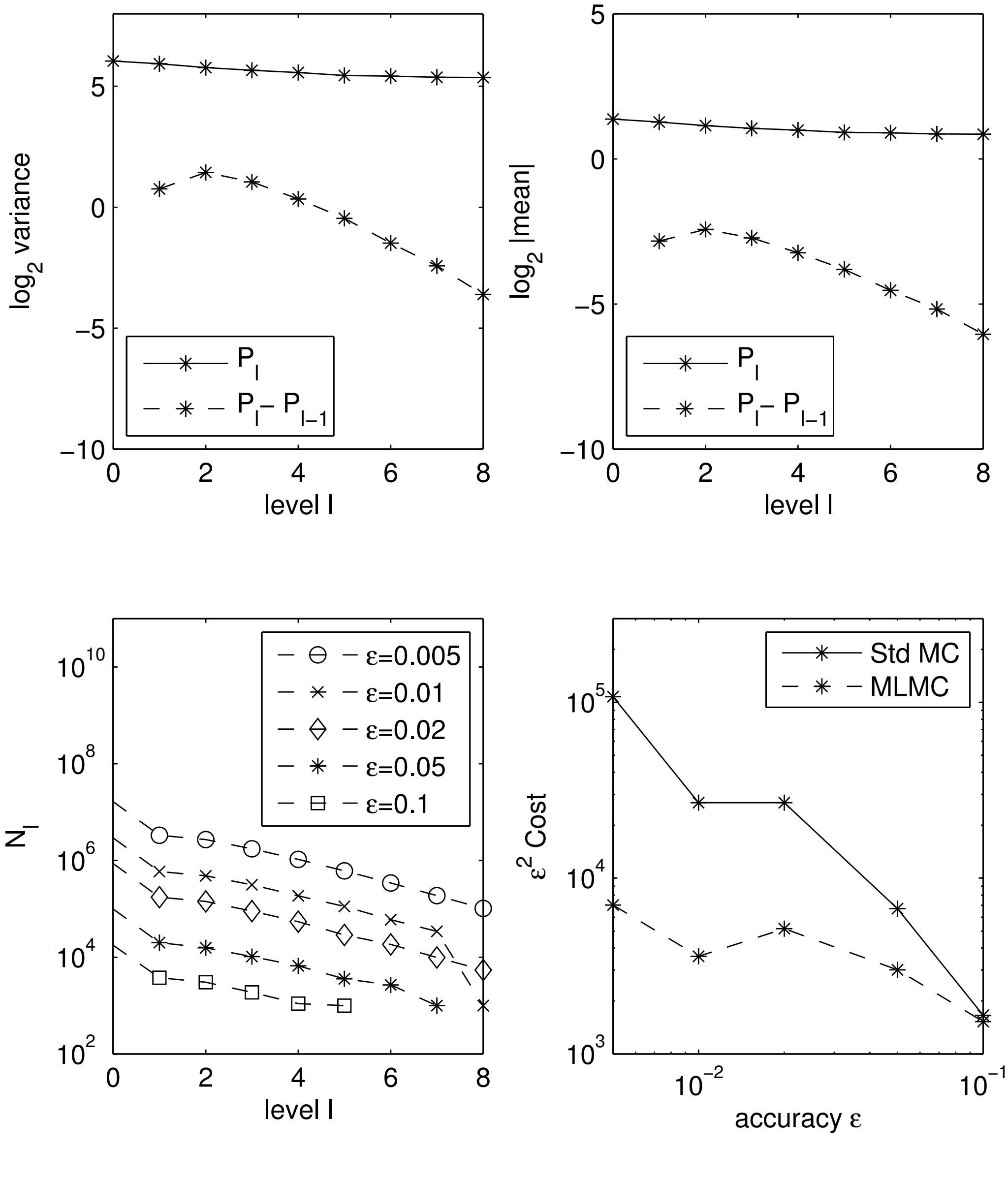

Figure 5.9: Barrier option in Normal Inverse Gaussian model

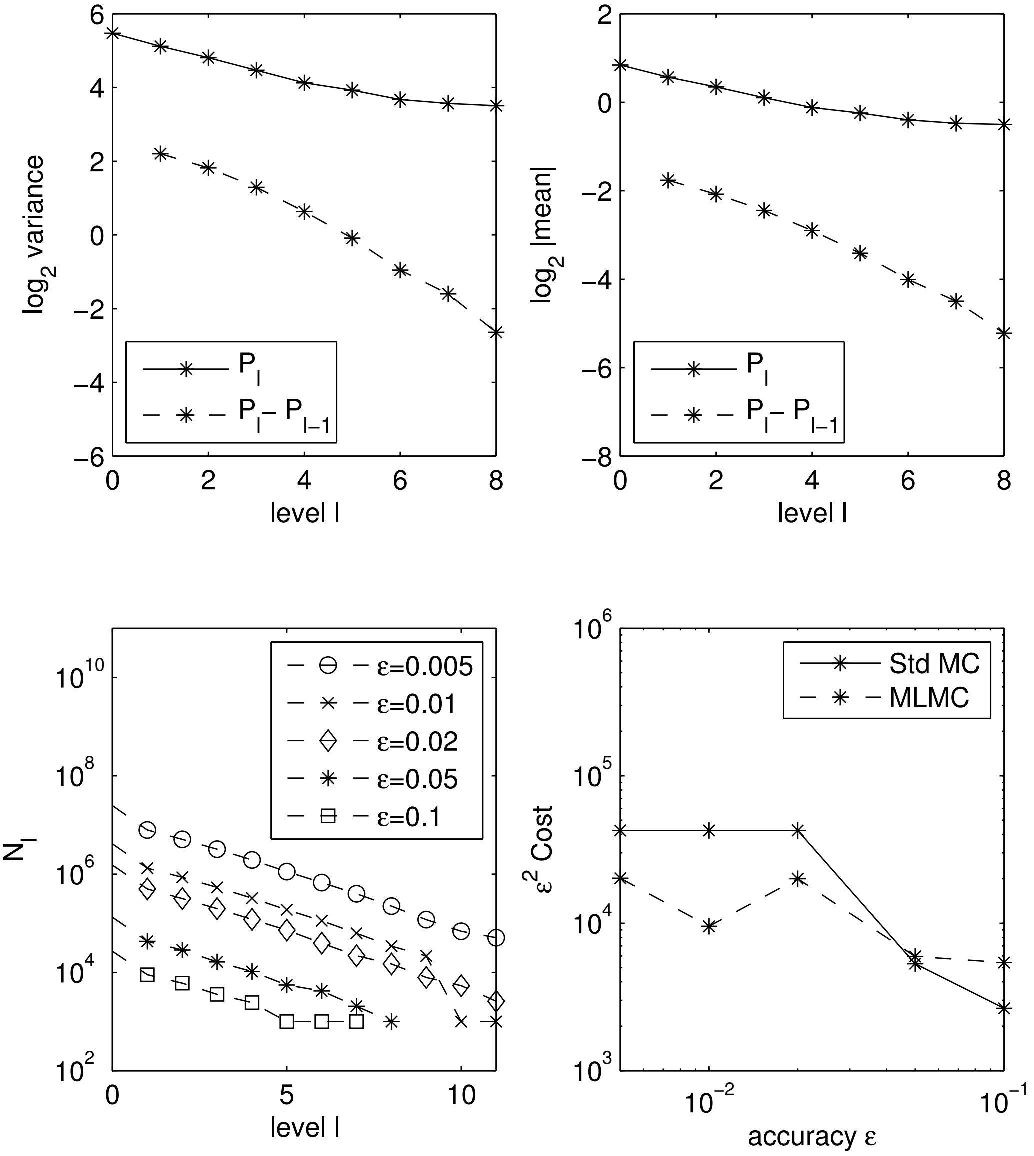

Figure 5.10: Barrier option in spectrally negative α-stable model

	VG			
	numerical		analysis	
option	weak	var	weak	var
Asian	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$
lookback	$\mathcal{O}(h)$	$\mathcal{O}(h^{1.5})$	$\mathcal{O}(h\lvert\log h\rvert)$	$\mathcal{O}(h)$
barrier	$\mathcal{O}(h^{0.7})$	$\mathcal{O}(h)$	$o\left(h^{1-\delta}\right)$	$o\left(h^{1-\delta}\right)$

	NIG			
	numerical		analysis	
option	weak	var	weak	var
Asian	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$
lookback	$\mathcal{O}(h^{0.7})$	$\mathcal{O}(h^{1.5})$	$o\left(h^{1-\delta}\right)$	$\mathcal{O}(h\lvert\log h\rvert)$
barrier	$\mathcal{O}(h^{0.7})$	$\mathcal{O}(h^{0.9})$	$o\left(h^{0.5-\delta}\right)$	$o\left(h^{0.5-\delta}\right)$

	spectrally negative α-stable with $\alpha > 1$			
	numerical for $\alpha = 1.5597$		analysis	
option	weak	var	weak	var
Asian	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$
lookback	$\mathcal{O}(h^{0.5})$	$\mathcal{O}(h^{1.5})$	$o\left(h^{1/\alpha-\delta}\right)$	$o\left(h^{2/\alpha-\delta}\right)$
barrier	$\mathcal{O}(h^{0.5})$	$\mathcal{O}(h^{0.7})$	$o\left(h^{1/\alpha-\delta}\right)$	$o\left(h^{1/\alpha-\delta}\right)$

Table 5.1: Convergence rates of V_l for VG, NIG and α-stable processes; δ can be any small positive constant. $h \equiv h_l$.

Numerical results and analysis of convergence rates of $V_l = \mathbb{V}[\widehat{P}_l - \widehat{P}_{l-1}]$ by Propositions 5.3.15, 5.3.16, 5.3.20 in the following sections are summerised in Table 5.1.

Note that we have a better variance convergence rate in exponential Lévy models driven by VG or NIG processes than in the Geometric Brownian Motion model.

5.3.2 Complexity of MLMC

For the lookback put payoff, numerically to achieve a root-mean-square (RMS) error of $\mathcal{O}(\epsilon)$, it needs approximate $\mathcal{O}(\epsilon^{-2})$ computational cost for VG, NIG and spectrally negative α-stable processes. For the analytical justification, in Section 5.3.4.1, weak convergence for VG, NIG and α-stable processes is obtained. Combined with the upper bound on the multilevel variance V_l, Theorem 1.3.1 indicates that: a computational complexity of $\mathcal{O}(\epsilon^{-2}(\log\epsilon)^2)$ is required for VG; a computational complexity of $o(\epsilon^{-2+\delta})$ where δ is any small positive number, is guaranteed for NIG; a computational complexity of $\mathcal{O}(\epsilon^{-2})$ for spectrally negative α-stable processes with $1 < \alpha < 2$ is also guaranteed.

For the Asian payoff, numerically to achieve a root-mean-square (RMS) error of $\mathcal{O}(\epsilon)$, it needs $\mathcal{O}(\epsilon^{-2})$ computational cost for VG, NIG and spectrally negative α-stable processes. In the analysis, first order weak convergence for all payoffs using VG, NIG and α-stable processes is implied by second order strong convergence. Combined with the upper bounds on the multilevel variances, Theorem 1.3.1 indicates that the optimal computational complexity of $\mathcal{O}(\epsilon^{-2})$ is guaranteed for Asian payoffs in the VG, NIG and spectrally negative α-stable model.

Those two payoffs are examples where multilevel can achieve the optimal computational complexity up to a $o(\epsilon^{\delta})$ factor for the payoff Lipschitz continuous with respect to the supremum norm, regardless of the Blumenthal-Getoor index of the process.

For the up-and-out barrier payoff, numerically to achieve a root-mean-square (RMS) error of $\mathcal{O}(\epsilon)$, it needs at least $\mathcal{O}(\epsilon^{-2})$ computational cost for VG, NIG and spectrally negative α-stable processes. In the analysis, combining weak convergence results with the upper bound on the multilevel variance convergence rate, it indicates the computational complexity of $o(\epsilon^{-2+\delta})$ for VG processes, where δ is any small positive number. For NIG processes, the weak convergence order proved is approximately one half. A closer look at the proof of Theorem 1.3.1 in [32], the RHS of the inequality (7) becomes $o(\epsilon^{-2+\delta})$ which leads to an upper bound on the computational complexity of $o(\epsilon^{-3-\delta})$, where δ is any small positive number. Currently there is a gap between numerical results and the upper bound by the analysis. The reason will be explained later in the analysis section. For spectrally negative α-stable processes, the computational complexity is $o(\epsilon^{-2+\delta})$, where δ is any small positive number.

The barrier payoff in a model driven by the spectrally negative α-stable process is an example where multilevel can achieve the optimal computational complexity up to a $o(\epsilon^{\delta})$ factor for the non-Lipschitz path-dependent payoff regardless of the Blumenthal-Getoor index of the process.

5.3.3 Discretely monitored error

For a Lévy process X_t, we denote the discrepancy between discretely and continuous monitored supremums by

$$V_n = \sup_{0 \le t \le 1} X_t - \max_{i=0,1,\dots,n} X_{\frac{i}{n}}.$$

For the weak convergence of lookback-type payoffs, we are concerned with the expectation of the discretely monitored error $\mathbb{E}[V_n]$, which is extensively studied in the literature. For example, [18], [10] and [11] derive asymptotic expansions for jump-diffusion, VG, NIG processes and estimates for general Lévy processes, by using Spitzer's identity:

If $\int_{x>1} |x| \ \nu(\mathrm{d}x) < \infty$, then

$$\mathbb{E}\left[\max_{i=0,1,\dots,n} X_{\frac{i}{n}}\right] = \sum_{i=1}^{n} \frac{1}{i} \mathbb{E}\left[X^+_{\frac{i}{n}}\right];$$

$$Y := \sup_{0 \le t \le 1} X_t = \int_0^1 \frac{1}{t} X_t^+ \mathrm{d}t.$$

We quote the result in [10]:

Theorem 5.3.1. *Suppose X_t is a scalar Lévy process with triple (m, σ, ν), with finite first moment, i.e.*

$$\int_{|x|>1} |x| \ \nu(\mathrm{d}x) < \infty.$$

Then $V_n = \sup_{0 \le t \le 1} X_t - \max_{i=0,1,\dots,n} X_{\frac{i}{n}}$

satisfies

1. *If $\sigma > 0$*

$$\mathbb{E}[V_n] = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right);$$

2. *If $\sigma = 0$ and X_t is of finite variation, i.e. $\int_{|x|<1} |x| \, \nu(\mathrm{d}x) < \infty$*

$$\mathbb{E}[V_n] = \mathcal{O}\left(\frac{\log n}{n}\right);$$

3. *If $\sigma = 0$ and X_t is of infinite variation, then*

$$\mathbb{E}[V_n] = \mathcal{O}\left(\frac{1}{n^r}\right),$$

where $\frac{1}{2} < r = \frac{1}{\beta} - \delta < \frac{1}{\beta} \leq 1$ *for any small* $\delta > 0$ *and*

$$\beta = \inf \left\{ \alpha > 0 : \int_{|x|<1} |x|^{\alpha} \nu(\mathrm{d}x) < \infty \right\}$$

is the Blumenthal-Getoor index of X_t.

The VG process has finite variation with Blumenthal-Getoor index 0; the NIG process has infinite variation with Blumenthal-Getoor index 1. They correspond to the second and third cases of Theorem 5.3.1 respectively.

For the multilevel variance convergence rate of the payoff with respect to the running supremum of X_t, we are interested in the second moment of U_n, which is the discrepancy between discretely monitored supremums of a Lévy process on two scales:

$$\begin{aligned} U_n &= \max_{i=0,1,\dots,2n} X_{\frac{i}{2n}} - \max_{i=0,1,\dots,n} X_{\frac{i}{n}} \\ &= V_n - V_{2n}. \end{aligned}$$

Since $0 < U_n \leq V_n$, we have

$$\mathbb{E}\left[U_n^2\right] \leq \mathbb{E}\left[V_n^2\right].$$

In the following we will analyse V_n instead of U_n. First we review existing approaches on bounds of the second moment of V_n in the case of diffusion or general Lévy processes.

For the second moment of V_n, the method of expectation expansion by Spitzer's identity faces difficulties in obtaining a positive lower bound on the covariance of the supremum and the state value appearing in $\mathbb{E}\left[V_n^2\right]$:

$$\mathbb{E}\left[\max_{i=0,1,\dots,n} X_{\frac{i}{n}} \sup_{0\leq t\leq 1} X_t\right] = \sum_{i=1}^{n} \frac{1}{i} \mathbb{E}\left[X^+_{\frac{i}{n}} \sup_{0\leq t\leq 1} X_t\right].$$

From another perspective, in the Brownian case, Asmussen et al. ([3]) obtain the asymptotic distribution of V_n. This implies the asymptotic behavior of $\mathbb{E}\left[V_n^2\right]$. [18] extends the result to finite activity jump processes with non-zero diffusion. The obstacle to following this approach comes from the fact that here we are looking at infinite activity jump processes.

On the other hand, Extreme Value theory [21] offers precise estimates of any moments of V_n in the diffusion case. A naive estimate gives

$$\begin{aligned}
&\sup_{0\le t\le 1} X_t - \max_{i=0,1,\dots,n} X_{\frac{i}{n}} \\
= &\max_{i=1,\dots,n}\left(\sup_{[\frac{i-1}{n},\frac{i}{n}]} X_t - X_{\frac{i}{n}} + X_{\frac{i}{n}}\right)^+ - \left(\max_{i=1,\dots,n} X_{\frac{i}{n}}\right)^+ \\
\le &\max_{i=1,\dots,n}\left(\sup_{[\frac{i-1}{n},\frac{i}{n}]} X_t - X_{\frac{i}{n}}\right)^+ . \qquad (5.6)
\end{aligned}$$

In law, the final bound has the same distribution as $\max_{i=1,\dots,n}(\sup_{[0,\frac{1}{n}]}(-X_t)(i))^+$, where $\sup_{[0,\frac{1}{n}]}(-X_t)(i)$ denotes i.i.d. samples of $\sup_{[0,\frac{1}{n}]}(-X_t)$.

By Extreme Value theory arguments in [21], using the tail distribution property of the sup process of Brownian Motion,

$$\mathbb{E}\left[\left(\max_{i=1,\dots,n}(\sup_{[0,\frac{1}{n}]}(-X_t)(i)\right)^k\right] = \mathcal{O}(\left(\frac{1}{n}\log n\right)^{k/2}).$$

When it comes to the VG process, it is difficult to obtain the tail distribution property of the sup process, and the exponentially decaying tail of the VG process is unable to provide a useful estimate.

In the following, we decompose the process into a finite-activity part, a drift part and a remainder part consisting of small jumps and refine the inequality (5.6) to analyse the finite-activity part.

5.3.3.1 Approximation of Lévy processes

For the preparation of the numerical analysis, we introduce the approximation of an infinite-activity pure jump Lévy process by a compound Poisson process.

Let X_t be an $(m, 0, \nu)$-Lévy process. Hence

$$X_t = mt + \int_0^t \int_{|z|\ge 1} z\mu(\mathrm{d}z,\mathrm{d}s) + \int_0^t \int_{|z|<1} z\left(\mu(\mathrm{d}z,\mathrm{d}s) - \nu(\mathrm{d}z)\mathrm{d}s\right).$$

Let

$$X_t^\varepsilon = \int_0^t \int_{\varepsilon<|z|} z\mu(\mathrm{d}z,\mathrm{d}s) = \sum_{i=1}^{N_t} Y_i \qquad (5.7)$$

be the compound Poisson process truncating the jumps of X_t smaller than ε. The intensity of N_t is

$$\lambda_\varepsilon := \int_{\varepsilon<|z|} \nu(\mathrm{d}z). \tag{5.8}$$

The c.d.f. of Y_i is

$$\mathbb{P}\left(Y_i < y\right) = \lambda_\varepsilon^{-1} \int_{z<y} 1_{\{\varepsilon<|z|\}} \nu(\mathrm{d}z).$$

Let

$$\mu_\varepsilon = m - \int_{\varepsilon<|z|<1} z\ \nu(\mathrm{d}z).$$

be a drift rate. As ε goes to zero, μ_ε has a finite limit if the process has finite variation,

but is not necessarily finite in other cases. It is worth noting that although NIG processes have infinite variation, by examining the near-zero asymptotic expansion (5.1), we conclude that NIG satisfies $\mu_\varepsilon < \infty$.

$$X_t^\varepsilon + \mu_\varepsilon t$$

is an approximation of X_t by truncating the jumps smaller than ε. The residual term of the jump part is a martingale:

$$R_t^\varepsilon := X_t - X_t^\varepsilon - \mu_\varepsilon t = \int_0^t \int_{|z|\le\varepsilon} z\left(\mu(\mathrm{d}z,\mathrm{d}s) - \nu(\mathrm{d}z)\mathrm{d}s\right).$$

One measure of the error in the approximation is the variance of R_t^ε :

$$\mathbb{V}\left[R_t^\varepsilon\right] = \int_{|z|\le\varepsilon} z^2 \nu(\mathrm{d}z)\ t \equiv \sigma_\varepsilon^2 t. \tag{5.9}$$

This is by differentiating the characteristic function of R_t^ε

$$\mathbb{E}\left[\exp\left(iuR_t^\varepsilon\right)\right] = \exp\left(t\int_{|z|\le\varepsilon} (e^{iuz} - 1 - iuz)\nu(\mathrm{d}z)\right)$$

with respect to u twice and then let $u = 0$.

μ_ε, λ_ε and σ_ε will play a major role in the subsequent numerical analysis.

Let X_t be an $(m, 0, \nu)$-Lévy process. We have the decomposition of $X_t = X_t^\varepsilon + R_t^\varepsilon + \mu_\varepsilon t$. We bound V_n by the difference between continuous maxima and 2-point maxima over all timesteps:

$$\begin{aligned} V_n &= \sup_{0\le t\le 1} X_t - \max_{i=0,1,\ldots,n} X_{\frac{i}{n}} \\ &\le \max_{i=1,\ldots,n}\left(\sup_{[\frac{i-1}{n},\frac{i}{n}]} X_t - \max\left(X_{\frac{i}{n}}, X_{\frac{i-1}{n}}\right)\right) \end{aligned} \tag{5.10}$$

where $\sup_{[\frac{i-1}{n},\frac{i}{n}]} X_t - \max\left(X_{\frac{i}{n}}, X_{\frac{i-1}{n}}\right)$ is equivalent in law to

$$\begin{aligned} &\sup_{[0,\frac{1}{n}]} X_t - \left(X_{\frac{1}{n}}\right)^+ \\ =\ &\sup_{[0,\frac{1}{n}]} \left(X_t^\varepsilon + R_t^\varepsilon + \mu_\varepsilon t\right) - \left(X_{\frac{1}{n}}^\varepsilon + R_{\frac{1}{n}}^\varepsilon + \mu_\varepsilon \frac{1}{n}\right)^+ \\ \le\ &\sup_{[0,\frac{1}{n}]} \left(X_t^\varepsilon + R_t^\varepsilon\right) - \left(X_{\frac{1}{n}}^\varepsilon + R_{\frac{1}{n}}^\varepsilon\right)^+ + \frac{|\mu_\varepsilon|}{n} \\ \le\ &\sup_{[0,\frac{1}{n}]} X_t^\varepsilon - \left(X_{\frac{1}{n}}^\varepsilon\right)^+ + \frac{|\mu_\varepsilon|}{n} + \sup_{[0,\frac{1}{n}]} R_t^\varepsilon + \left(-R_{\frac{1}{n}}^\varepsilon\right)^+ \\ \le\ &\sup_{[0,\frac{1}{n}]} X_t^\varepsilon - \left(X_{\frac{1}{n}}^\varepsilon\right)^+ + \frac{|\mu_\varepsilon|}{n} + 2\sup_{[0,\frac{1}{n}]} \left|R_t^\varepsilon\right| \end{aligned} \tag{5.11}$$

where we use $(a+b)^+ \le a^+ + b^+$ with $a = X_{\frac{1}{n}}^\varepsilon + R_{\frac{1}{n}}^\varepsilon + \mu_\varepsilon \frac{1}{n}$ and $b = -\mu_\varepsilon \frac{1}{n}$ in the first inequality; $a = X_{\frac{1}{n}}^\varepsilon + R_{\frac{1}{n}}^\varepsilon$ and $b = -R_{\frac{1}{n}}^\varepsilon$ in the second inequality.

The following lemma is used in bounding $\mathbb{E}\left[V_n^p\right]$ if $p < 1$.

Lemma 5.3.2. *For nonnegative numbers x, y, if $0 < p < 1$ then we have*

$$(x+y)^p \le x^p + y^p.$$

In (5.11), let

$$S_n := 2\sup_{[0,\frac{1}{n}]} \left|R_t^\varepsilon\right|. \tag{5.12}$$

For $p > 0$, we have

$$\begin{aligned} \mathbb{E}\left[V_n^p\right] &\le \mathbb{E}\left[\max_{i=1,\ldots,n}\left(\sup_{[0,\frac{1}{n}]} X_t^\varepsilon(i) - \left(X_{\frac{1}{n}}^\varepsilon(i)\right)^+ + \frac{|\mu_\varepsilon|}{n} + S_n^{(i)}\right)^p\right] \\ &\le \mathbb{E}\left[C_p \max_{i=1,\ldots,n}\left(\sup_{[0,\frac{1}{n}]} X_t^\varepsilon(i) - \left(X_{\frac{1}{n}}^\varepsilon(i)\right)^+\right)^p + C_p\left(\frac{|\mu_\varepsilon|}{n}\right)^p + C_p\left(\max_{i=1,\ldots,n} S_n^{(i)}\right)^p\right] \\ &\le C_p n \mathbb{E}\left[\left(\sup_{[0,\frac{1}{n}]} X_t^\varepsilon - \left(X_{\frac{1}{n}}^\varepsilon\right)^+\right)^p\right] + C_p\left(\frac{|\mu_\varepsilon|}{n}\right)^p \\ &\quad + C_p \mathbb{E}\left[\left(\max_{i=1,\ldots,n} S_n^{(i)}\right)^p\right]. \end{aligned} \tag{5.13}$$

where in the second step we use $\max_{i=1,\dots,n}(a_i+b_i) \le \max_{i=1,\dots,n} a_i + \max_{i=1,\dots,n} b_i$, Jensen's inequality if $p \ge 1$ and Lemma 5.3.2 if $0 < p < 1$. $C_p = \max(3^{p-1}, 1)$.

We shall give an upper bound on the moments of $\sup_{[0,\frac{1}{n}]} |R_t^\varepsilon|$.

Proposition 5.3.3. *Let X_t be a scalar Lévy process with a triple $(m, 0, \nu)$. Suppose the Lévy measure satisfies*

$$\int_{|z|\le\varepsilon} z^2\nu(\mathrm{d}z)\ t \equiv \sigma_\varepsilon^2 < +\infty.$$

Let

$$R_t^\varepsilon = \int_0^t \int_{|z|\le\varepsilon} z\left(\mu(\mathrm{d}z,\mathrm{d}s) - \nu(\mathrm{d}z)\mathrm{d}s\right).$$

be the jump process with jumps smaller than ε. Then R_t^ε satisfies

$$\mathbb{E}\left[\sup_{0\le t\le T} |R_t^\varepsilon|^p\right] \le \begin{cases} K_p\left(T^{p/2}\sigma_\varepsilon^p + T\int_{|z|\le\varepsilon}|z|^p\,\nu(\mathrm{d}x)\right), \ p>2; \\ K_pT^{p/2}\sigma_\varepsilon^p, \ 0<p\le 2, \end{cases}$$

where K_p is a constant depending on p.

Proof. For any $0 < p \le 2$, by Jensen's inequality and Doob's inequality (A.2.10),

$$\begin{aligned} \mathbb{E}\left[\sup_{0\le t\le T} |R_t^\varepsilon|^p\right] &\le \mathbb{E}\left[\sup_{0\le t\le T} |R_t^\varepsilon|^2\right]^{p/2} \\ &\le 2^p\mathbb{E}\left[|R_T^\varepsilon|^2\right]^{p/2} \\ &= 2^pT^{p/2}\sigma_\varepsilon^p. \end{aligned}$$

In particular when $p = 2$,

$$\mathbb{E}\left[\sup_{0\le t\le T} |R_t^\varepsilon|^2\right] \le 4T\sigma_\varepsilon^2.$$

For any $p > 2$ we can use the method on page 347 of [52]. In the following the constants C_p and K_p vary from line to line.

By the Burkholder–Davis–Gundy inequality (A.2.9), we have

$$\begin{aligned} \mathbb{E}\left[\sup_{0\le t\le 1} |R_t^\varepsilon|^p\right] &\le \mathbb{E}\left[[R^\varepsilon]_1^{p/2}\right] \\ &= \mathbb{E}\left[\left(\sum_{s\le 1}(\Delta R_s^\varepsilon)^2\right)^{p/2}\right] \end{aligned}$$

where $[R^\varepsilon]_t = \int_0^t \int_{|z|\le\varepsilon} z^2 \mu(\mathrm{d}z, \mathrm{d}s)$ is the quadratic variation of R^ε_t and ΔR^ε_s are the jumps of R^ε.

We have

$$\mathbb{E}\left[\left(\sum_{s\le 1}(\Delta R^\varepsilon_s)^2\right)^{p/2}\right] \le C_p\mathbb{E}\left[\left(\sum_{s\le 1}(\Delta R^\varepsilon_s)^2 - \int_{|z|\le\varepsilon}|z|^2\,\nu(\mathrm{d}z)\right)^{p/2} + \left(\int_{|z|\le\varepsilon}|z|^2\,\nu(\mathrm{d}z)\right)^{p/2}\right].$$

Apply the Burkholder–Davis–Gundy inequality to the first term on the RHS above:

$$\mathbb{E}\left[\left(\sum_{s\le 1}(\Delta R^\varepsilon_s)^2\right)^{p/2}\right] \le \mathbb{E}\left[K_p\left(\sum_{s\le 1}(\Delta R^\varepsilon_s)^4\right)^{p/4} + C_p\left(\int_{|z|\le\varepsilon}|z|^2\,\nu(\mathrm{d}z)\right)^{p/2}\right].$$

Continue recursively until $1 \le p2^{-k} < 2$:

$$\mathbb{E}\left[\left(\sum_{s\le 1}(\Delta R^\varepsilon_s)^2\right)^{p/2}\right] \le \mathbb{E}\left[K_p\left(\sum_{s\le 1}(\Delta R^\varepsilon_s)^{2^{k+1}}\right)^{p2^{-(k+1)}} + C_p\sum_{i=1}^{k}\left(\int_{|z|\le\varepsilon}|z|^{2^i}\,\nu(\mathrm{d}z)\right)^{p2^{-i}}\right].$$

We know if $0 < q_1 \le q_2 < \infty$ then for an infinite sequence x we have $\|x\|_{q_2} \le \|x\|_{q_1}$. Let $q_1 = p2^{-k}$ and $q_2 = 2$. As $1 \le p2^{-k} < 2$,

$$\begin{aligned}\left(\sum_{s\le 1}(\Delta R^\varepsilon_s)^{2^{k+1}}\right)^{p2^{-(k+1)}} &= \left(\sum_{s\le 1}\left[(\Delta R^\varepsilon_s)^{2^k}\right]^2\right)^{\frac{1}{2}p2^{-k}} \\ &\le \sum_{s\le 1}|\Delta R^\varepsilon_s|^p.\end{aligned}$$

We have

$$\mathbb{E}\left[\sum_{s\le 1}|\Delta R^\varepsilon_s|^p\right] = \int_{|z|\le\varepsilon}|z|^p\,\nu(\mathrm{d}z).$$

On the other hand, on page 348 of [52] we have

$$\left(\int_{|z|\le\varepsilon}|z|^{2^i}\,\nu(\mathrm{d}z)\right)^{p2^{-i}} \le \left(\int_{|z|\le\varepsilon}|z|^2\,\nu(\mathrm{d}z)\right)^{p/2} + \int_{|z|\le\varepsilon}|z|^p\,\nu(\mathrm{d}z).$$

Therefore

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{s\le 1}(\Delta R^\varepsilon_s)^2\right)^{p/2}\right] &\le K_p\left[\left(\int_{|z|\le\varepsilon}z^2\nu(\mathrm{d}z)\right)^{p/2} + \int_{|z|\le\varepsilon}|z|^p\,\nu(\mathrm{d}z)\right] \\ &= K_p\left(\sigma^p_\varepsilon + \int_{|z|\le\varepsilon}|z|^p\,\nu(\mathrm{d}z)\right).\end{aligned}$$

Finally

$$\mathbb{E}\left[\sup_{0\le t\le 1}|R_t^\varepsilon|^p\right]\le K_p\left[\left(\int_{|z|\le\varepsilon}z^2\nu(\mathrm{d}z)\right)^{p/2}+\int_{|z|\le\varepsilon}|z|^p\,\nu(\mathrm{d}z)\right].$$

We consider a change of time coordinate, $t'=t/T$. Define $R_{t'}^{\varepsilon\prime}=R_t^\varepsilon$. Then

$$\begin{aligned}
R_T^\varepsilon &= \int_0^T\int_{|z|\le\varepsilon}z\left(\mu(\mathrm{d}z,\mathrm{d}t)-\nu(\mathrm{d}z)\mathrm{d}t\right)\\
&= \int_0^1\int_{|z|\le\varepsilon}z\left(\mu(\mathrm{d}z,\mathrm{d}t'T)-\nu(\mathrm{d}z)\mathrm{d}t'T\right)\\
&\equiv \int_0^1\int_{|z|\le\varepsilon}z\left(\mu'(\mathrm{d}z,\mathrm{d}t')-\nu'(\mathrm{d}z)\mathrm{d}t'\right)\\
&= R_1^{\varepsilon\prime}.
\end{aligned}$$

where $\nu'(\mathrm{d}z)=T\ \nu(\mathrm{d}z)$. By the estimate for $\mathbb{E}\left[\sup_{0\le t\le 1}|R_t^{\varepsilon\prime}|^p\right]$ we have

$$\mathbb{E}\left[\sup_{0\le t\le 1}|R_t^\varepsilon|^p\right]\le K_p\left[T^{p/2}\left(\int_{|z|\le\varepsilon}z^2\nu(\mathrm{d}z)\right)^{p/2}+T\int_{|z|\le\varepsilon}|z|^p\,\nu(\mathrm{d}z)\right].$$

where K_p is a constant depending on p. □

Corollary 5.3.4. *Let X_t be a scalar Lévy process with a triple $(m,0,\nu)$. In the (5.13), for $p>2$ we have*

$$\mathbb{E}\left[\left(\max_{i=1,\dots,n}S_n^{(i)}\right)^p\right]\le K_p\left(n^{1-p/2}\sigma_\varepsilon^p+\int_{|z|\le\varepsilon}|z|^p\,\nu(\mathrm{d}x)\right), \tag{5.14}$$

where K_p is a constant depending on p.

Proof. We have

$$\begin{aligned}
\mathbb{E}\left[\left(\max_{i=1,\dots,n}S_n^{(i)}\right)^p\right] &\le n\mathbb{E}\left[S_n^p\right]\\
&= n2^p\left(\mathbb{E}\left[\sup_{0\le t\le 1}|R_t^\varepsilon|^p\right]\right)\\
&\le K_p\left(n^{1-p/2}\sigma_\varepsilon^p+\int_{|z|\le\varepsilon}|z|^p\,\nu(\mathrm{d}x)\right).
\end{aligned}$$

□

We now come to an important theorem which gives an upper bound on $n\mathbb{E}\left[\left(\sup_{[0,\frac{1}{n}]}X_t^\varepsilon-\left(X_{\frac{1}{n}}^\varepsilon\right)^+\right)^p\right]$ in (5.13).

Theorem 5.3.5. *Let X_t be a scalar Lévy process with a triple $(m, 0, \nu)$, and*

$$X_t^\varepsilon := \int_0^t \int_{\varepsilon<|z|} z\mu(\mathrm{d}z, \mathrm{d}s)$$

be a compound Poisson process with jumps from X_t.

For any power index $p > 0$, assuming that there exists a constant K such that $\lambda_\varepsilon \leq Kn$, we have

$$n\mathbb{E}\left[\left(\sup_{[0,\frac{1}{n}]} X_t^\varepsilon - \left(X_{\frac{1}{n}}^\varepsilon\right)^+\right)^p\right] \leq K_p\left(\varepsilon^p + \frac{L_\varepsilon(p)}{\lambda_\varepsilon^2}\right)\frac{\lambda_\varepsilon^2}{n}, \tag{5.15}$$

where μ_ε, λ_ε and σ_ε are as defined in Section 5.3.3.1, K_p is a constant, and $L_\varepsilon(p) = p\int_{x>\varepsilon} x^{p-1}\lambda_x^2 \mathrm{d}x$ is a function depending on the Lévy measure $\nu(x)$.

Proof. Let

$$Z = \sup_{[0,\frac{1}{n}]} X_t^\varepsilon - \left(X_{\frac{1}{n}}^\varepsilon\right)^+.$$

We will give an upper bound on $\mathbb{E}[Z^p]$ by analysing the jump behavior of the finite-activity processes in a single interval $[0, \frac{1}{n}]$.

Let N_i be the number of jumps for replication i. Note that conditional on $N_i \leq 1$, $Z = 0$. Conditional on $N_i = 2$, $Z \leq \min(|Y_1|, |Y_2|)$. This can be seen from the behavior of Z in different scenarios of two jumps in figure 5.11. Based on this fact, we have

$$\begin{aligned}
\mathbb{E}[Z^p \mid N_i = 2] &= p\int x^{p-1}\mathbb{P}(Z > x \mid N_i = 2)\,\mathrm{d}x \\
&\leq p\int x^{p-1}\mathbb{P}(\min(|Y_1|, |Y_2|) > x)\,\mathrm{d}x \\
&= p\int x^{p-1}\mathbb{P}(|Y_1| > x)^2\,\mathrm{d}x \\
&= \frac{p}{\lambda_\varepsilon^2}\int x^{p-1}\left(\int_{|z|>x} 1_{\{\varepsilon<|z|\}}\nu(\mathrm{d}z)\right)^2\mathrm{d}x \\
&= \frac{p}{\lambda_\varepsilon^2}\left(\int_{0\leq x\leq\varepsilon} x^{p-1}\lambda_\varepsilon^2\mathrm{d}x + \int_{x>\varepsilon} x^{p-1}\lambda_x^2\mathrm{d}x\right) \\
&= \varepsilon^p + \frac{p}{\lambda_\varepsilon^2}\int_{x>\varepsilon} x^{p-1}\lambda_x^2\mathrm{d}x \\
&\equiv \varepsilon^p + \frac{L_\varepsilon(p)}{\lambda_\varepsilon^2}.
\end{aligned} \tag{5.16}$$

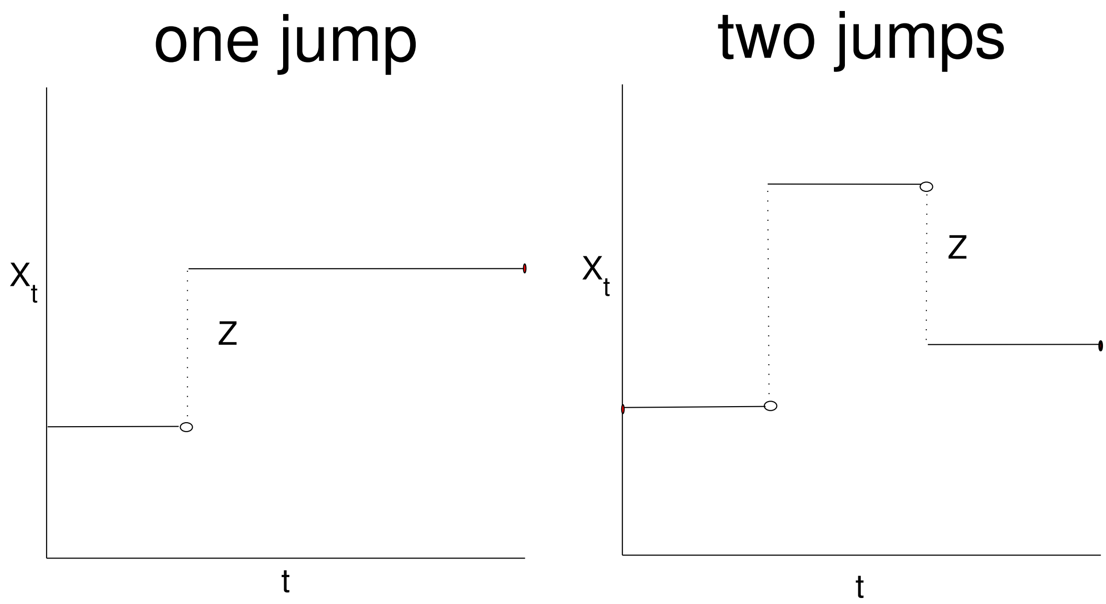

Figure 5.11: Illustration of the behavior of $Z(i)$ in the case of two jumps

More generally, conditional on $N_i = k,\ k \geq 2$

$$\begin{aligned} Z > x \quad &\Rightarrow \quad \exists 1 \leq j \leq k-1 \text{ s.t. } \left|\sum_{l=1}^{j} Y_l\right| > x,\ \left|\sum_{l=j+1}^{k} Y_l\right| > x \\ &\Rightarrow \quad \exists j_1, j_2 \text{ s.t. } |Y_{j_1}| > \frac{x}{k-1},\ |Y_{j_2}| > \frac{x}{k-1}. \end{aligned}$$

$$\begin{aligned} \mathbb{P}\left(\exists j_1, j_2 \text{ s.t. } |Y_{j_1}| > \frac{x}{k-1},\ |Y_{j_2}| > \frac{x}{k-1}\right) \quad &\leq \quad \sum_{(j_1,j_2)} \mathbb{P}(|Y_{j_1}| > \frac{x}{k-1},\ |Y_{j_2}| > \frac{x}{k-1}) \\ &= \quad \frac{k\,(k-1)}{2}\mathbb{P}(|Y_1| > \frac{x}{k-1})^2. \end{aligned}$$

Thus

$$\begin{aligned}
\mathbb{E}\left[Z^{p} \mid N_{i}=k\right] &= p \int x^{p-1} \mathbb{P}\left(Z>x \mid N_{i}=k\right) \mathrm{d} x \\
&\leq \frac{k(k-1)}{2} p \int x^{p-1} \mathbb{P}(|Y_{1}|>\frac{x}{k-1})^{2} \mathrm{d} x \\
&= \frac{k(k-1)}{2} \frac{p}{\lambda_{\varepsilon}^{2}} \int x^{p-1}\left(\int_{|z|>x/(k-1)} 1_{\{\varepsilon<|z|\}} \nu(\mathrm{d} z)\right)^{2} \mathrm{d} x \\
&= \frac{k(k-1)^{p+1}}{2} \frac{p}{\lambda_{\varepsilon}^{2}} \int\left(\frac{x}{k-1}\right)^{p-1}\left(\int_{|z|>x/(k-1)} 1_{\{\varepsilon<|z|\}} \nu(\mathrm{d} z)\right)^{2} \mathrm{d} \frac{x}{k-1} \\
&= \frac{k(k-1)^{p+1}}{2} \frac{p}{\lambda_{\varepsilon}^{2}} \int x^{p-1}\left(\int_{|z|>x} 1_{\{\varepsilon<|z|\}} \nu(\mathrm{d} z)\right)^{2} \mathrm{d} x \\
&\equiv d_{k,p}\left(\varepsilon^{p}+\frac{L_{\varepsilon}(p)}{\lambda_{\varepsilon}^{2}}\right), \qquad (5.17)
\end{aligned}$$

where

$$d_{k,p}=\frac{k(k-1)^{p+1}}{2}.$$

We then have

$$\begin{aligned}
\mathbb{E}\left[Z^{p}\right] &= \sum_{k=2}^{\infty} \mathbb{E}\left[Z^{p} \mid N=k\right]\ \mathbb{P}(N=k) \\
&\leq \left(\varepsilon^{p}+\frac{L_{\varepsilon}(p)}{\lambda_{\varepsilon}^{2}}\right)\ \exp\left(-\frac{\lambda_{\varepsilon}}{n}\right)\ \sum_{k=2}^{\infty} d_{k,p}\left(\frac{\lambda_{\varepsilon}}{n}\right)^{k} \frac{1}{k!}
\end{aligned}$$

For $k_p = \lceil p\rceil + 2$ there exists C_p such that for any $k \geq k_p$, $d_{k,p} \leq C_p \frac{k!}{(k-k_p)!}$, so

$$\begin{aligned}
\sum_{k=2}^{\infty} d_{k,p}\left(\frac{\lambda_{\varepsilon}}{n}\right)^{k} \frac{1}{k!} &\leq \sum_{k=2}^{k_p-1} d_{k,p}\left(\frac{\lambda_{\varepsilon}}{n}\right)^{k} \frac{1}{k!}+C_{p} \sum_{k=k_p}^{\infty}\left(\frac{\lambda_{\varepsilon}}{n}\right)^{k} \frac{1}{(k-k_{p})!} \\
&\leq \sum_{k=2}^{k_p-1} d_{k,p}\left(\frac{\lambda_{\varepsilon}}{n}\right)^{k} \frac{1}{k!}+C_{p}\left(\frac{\lambda_{\varepsilon}}{n}\right)^{k_p} \exp\left(\frac{\lambda_{\varepsilon}}{n}\right) \\
&\leq K_{p}\left(\frac{\lambda_{\varepsilon}}{n}\right)^{2}
\end{aligned}$$

for some constant K_p, where the last step is by assuming that there exists constant K such that $\lambda_\varepsilon \leq Kn$.

Therefore

$$\mathbb{E}\left[Z^{p}\right] \leq K_{p}\left(\varepsilon^{p}+\frac{L_{\varepsilon}(p)}{\lambda_{\varepsilon}^{2}}\right)\left(\frac{\lambda_{\varepsilon}}{n}\right)^{2}$$

and

$$n \mathbb{E}\left[Z^{p}\right] \leq K_{p}\left(\varepsilon^{p}+\frac{L_{\varepsilon}(p)}{\lambda_{\varepsilon}^{2}}\right) \frac{\lambda_{\varepsilon}^{2}}{n}.$$

□

By (5.13) and (5.15) we have: assuming that there exists a constant K such that $\lambda_\varepsilon \leq Kn$, then

$$\mathbb{E}\left[V_n^p\right] \prec \mathbb{E}\left[\left(\max_{i=1,\dots,n} S_n^{(i)}\right)^p\right] + \varepsilon^p \frac{\lambda_\varepsilon^2}{n} + \frac{L_\varepsilon(p)}{n} + \left(\frac{|\mu_\varepsilon|}{n}\right)^p. \tag{5.18}$$

where the notation $u \prec v$ means there exists constant $c > 0$ independent of n such that $u < cv$. This is the foundation of the following numerical analysis.

5.3.3.2 Analysis of individual processes

We will focus on analysing the convergence rate of second and higher moments of V_n.

Under the assumption $\lambda_\varepsilon \prec n$ of Theorem 5.3.5, if we can choose a proper scaling of λ_ε so that the RHS of (5.18) is convergent, the convergence rate of $\mathbb{E}\left[V_n^p\right]$ can be estimated. In the following, we propose the analysis for a broad class of Lévy processes where the Lévy measure admits almost power-law behavior near the origin and the tail of the Lévy measure decays exponentially. The particular cases of VG, NIG and tempered α-stable processes will follow as a corollary.

Proposition 5.3.6. *Let X_t be a scalar pure jump Lévy process. Suppose the Lévy measure $\nu(x)$ satisfies*

$$C_2 |x|^{-1-\alpha} \leq \nu(x) \leq C_1 |x|^{-1-\alpha}, \text{ as } |x| \leq 1;$$

where $C_1, C_2 > 0,\ 0 \leq \alpha < 2$ are constants. For $p > 0$, we have for arbitrary $\delta > 0$,

$$\mathbb{E}\left[\left(\max_{i=1,\dots,n} S_n^{(i)}\right)^p\right] \prec \varepsilon^{p-\delta}.$$

Proof. For $p > 2$, by corollary 5.3.4

$$\mathbb{E}\left[\left(\max_{i=1,\dots,n} S_n^{(i)}\right)^p\right] \leq K_p \left(n^{1-p/2}\sigma_\varepsilon^p + \int_{|z|\leq\varepsilon} |z|^p \,\nu(\mathrm{d}x)\right).$$

Recall

$$\sigma_\varepsilon = \left(\int_{|z|\leq\varepsilon} z^2 \,\nu(\mathrm{d}z)\right)^{1/2}.$$

Since

$$\int_{|z|\leq\varepsilon} z^2 \frac{1}{|z|^{\alpha+1}} \mathrm{d}z = \frac{2}{2-\alpha}\varepsilon^{2-\alpha},$$

we have

$$\frac{2C_2}{2-\alpha}\varepsilon^{2-\alpha} \le \sigma_\varepsilon^2 \le \frac{2C_1}{2-\alpha}\varepsilon^{2-\alpha}$$

and

$$\sigma_\varepsilon^p \sim \varepsilon^{p-p\alpha/2},$$

where the notation $u \sim v$ means there exists constant $c_1, c_2 > 0$ independent of n such that $c_1 v \le u \le c_2 v$.

For $p > 2 > \alpha$, since

$$\int_{|z|\le\varepsilon} |z|^p \frac{1}{|z|^{\alpha+1}} \mathrm{d}z = \frac{2}{p-\alpha}\varepsilon^{p-\alpha},$$

we have

$$\frac{2C_2}{p-\alpha}\varepsilon^{p-\alpha} \le \int_{|z|\le\varepsilon} |z|^p \, \nu(\mathrm{d}x) \le \frac{2C_1}{p-\alpha}\varepsilon^{p-\alpha}$$

i.e.

$$\int_{|z|\le\varepsilon} |z|^p \, \nu(\mathrm{d}x) \sim \varepsilon^{p-\alpha}. \tag{5.19}$$

Assuming $0 < \varepsilon < 1$, we work out a lower bound λ_ε:

$$\begin{aligned} \lambda_\varepsilon &\ge C_2 \int_{\varepsilon<|z|<1} \frac{1}{|z|^{\alpha+1}} \mathrm{d}z \\ &= \begin{cases} 2C_2 \log\frac{1}{\varepsilon}, & \alpha = 0; \\ \frac{2C_2}{\alpha}\varepsilon^{-\alpha}, & 0 < \alpha < 2. \end{cases} \end{aligned} \tag{5.20}$$

Hence the assumption of Theorem 5.3.5

$$\lambda_\varepsilon \le n$$

implies

$$\begin{cases} 2C_2 \log\frac{1}{\varepsilon} \le n, & \alpha = 0; \\ \frac{2C_2}{\alpha}\varepsilon^{-\alpha} \le n, & 0 < \alpha < 2. \end{cases} \tag{5.21}$$

We have

$$\frac{n^{1-p/2}\sigma_\varepsilon^p}{\int_{|z|\le\varepsilon} |z|^p \, \nu(\mathrm{d}x)} \le \frac{C_1}{C_2}\frac{p-\alpha}{2-\alpha}\left(\frac{\varepsilon^{-\alpha}}{n}\right)^{\frac{p}{2}-1} \le \frac{C_1}{C_2}\frac{p-\alpha}{2-\alpha}\left(\frac{\alpha}{2C_2}\right)^{\frac{p}{2}-1}. \tag{5.22}$$

where the last step is due to (5.21).

Furthermore by (5.14), given $p > 0$, for any $q > \max(2, p)$, by Jensen's inequality we have

$$\begin{aligned}
\mathbb{E}\left[\left(\max_{i=1,\ldots,n} S_n^{(i)}\right)^p\right] &\leq \mathbb{E}\left[\left(\max_{i=1,\ldots,n} S_n^{(i)}\right)^q\right]^{p/q} \\
&\leq \left(1 + \frac{C_1}{C_2}\frac{q-\alpha}{2-\alpha}\left(\frac{\alpha}{2C_2}\right)^{\frac{q}{2}-1}\right)^{p/q} K_q^{p/q}\left(\int_{|z|\leq\varepsilon} |z|^q \,\nu(\mathrm{d}x)\right)^{p/q} \\
&\prec \left(\varepsilon^{q-\alpha}\right)^{p/q} \\
&= \varepsilon^{p-\alpha p/q}.
\end{aligned}$$

For any $p > 0$, we can take q as large as we like to get a bound

$$\mathbb{E}\left[\left(\max_{i=1,\ldots,n} S_n^{(i)}\right)^p\right] \prec \varepsilon^{p-\delta}$$

for any $\delta > 0$. □

Proposition 5.3.7. *Let X_t be a scalar pure jump Lévy process. Suppose the Lévy measure $\nu(x)$ satisfies*

$$\begin{array}{c} C_2\,|x|^{-1-\alpha} \leq \nu(x) \leq C_1\,|x|^{-1-\alpha}, \text{ as } |x| \leq 1; \\ \nu(x) \leq \exp\left(-C_3\,|x|\right), \text{ as } |x| > 1, \end{array} \tag{5.23}$$

where $C_1, C_2, C_3 > 0$, $0 \leq \alpha < 2$ are constants. Then $V_n = \sup_{0\leq t\leq 1} X_t - \max_{i=0,1,\ldots,n} X_{\frac{i}{n}}$ satisfies

a) $p \geq 2\alpha$, $p \geq 1$,

$$\mathbb{E}\left[V_n^p\right] = \begin{cases} \mathcal{O}\left(\frac{1}{n}\right), & p > 2\alpha; \\ \mathcal{O}\left(\frac{\log n}{n}\right), & p = 2\alpha. \end{cases}$$

b) $p < 2\alpha$ *and* $\alpha \geq \frac{1}{2}$,

$$\mathbb{E}\left[V_n^p\right] = \mathcal{O}\left(\left(\frac{\log n}{n}\right)^{\frac{p}{2\alpha}}\right).$$

c) $p < 1$ *and* $\alpha < \frac{1}{2}$,

$$\mathbb{E}\left[V_n^p\right] = \mathcal{O}\left(n^{-p}\right).$$

These three (p, α) regions are illustrated in figure 5.12.

In particular for $p = 2$,

1. *if $0 \leq \alpha < 1$, then*

$$\mathbb{E}\left[V_n^2\right] = \mathcal{O}\left(\frac{1}{n}\right);$$

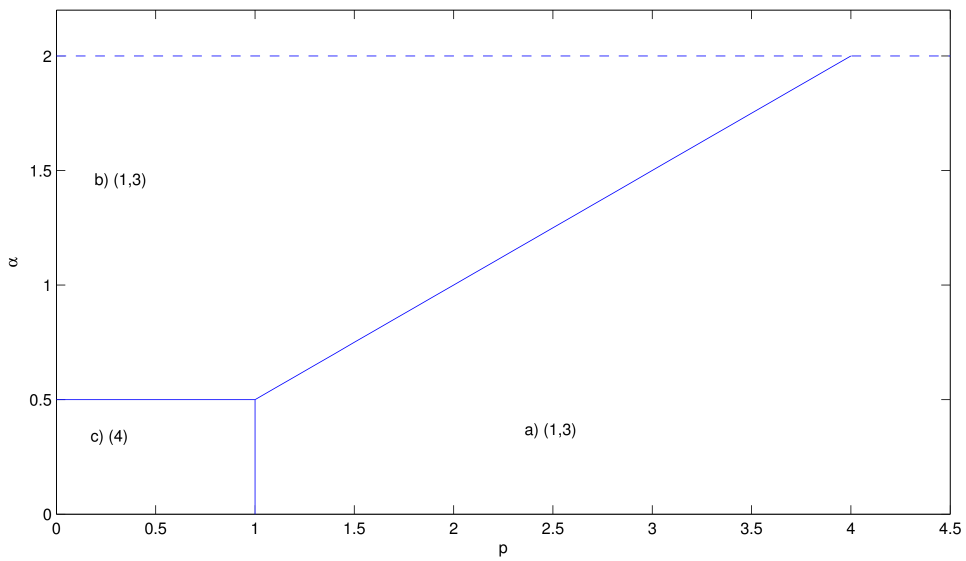

Figure 5.12: The pair of numbers after the label of regime indicates the leading order terms among four terms in (5.18).

2. if $\alpha = 1$*, then*

$$\mathbb{E}\left[V_n^2\right] = \mathcal{O}\left(\frac{\log n}{n}\right);$$

3. if $1 < \alpha < 2$*, then*

$$\mathbb{E}\left[V_n^2\right] = \mathcal{O}\left(\left(\frac{\log n}{n}\right)^{\frac{1}{\alpha}}\right).$$

In addition, if X_t *is an* α*-stable process with* $\alpha > 1$*, then for* $p < 2\alpha$*,*

$$\mathbb{E}\left[V_n^p\right] = \mathcal{O}\left(n^{-(p-\delta)(2\alpha-\delta)^{-1}}\right),$$

where δ *can be any small positive constant.*

Proof. According to (5.18), we work out the upper bounds of μ_ε, λ_ε and $L_\varepsilon(p)$ when p, α are in different sets of value and then balance their asymptotic orders.

For $0 < \varepsilon < 1$,

$$\begin{aligned} \lambda_\varepsilon &\le C_1 \int_{\varepsilon<|z|<1} \frac{1}{|z|^{\alpha+1}} \mathrm{d}z + \int_{1<|z|} \exp\left(-C_3 |z|\right) \mathrm{d}z \\ &\le \begin{cases} 2C_1 \log \frac{1}{\varepsilon} + l_1, \ \alpha = 0; \\ l_2 \varepsilon^{-\alpha}, \ 0 < \alpha < 2. \end{cases} \end{aligned} \tag{5.24}$$

For a positive x,

$$\lambda_x \le \begin{cases} 1_{\{x\le 1\}}\left(2C_1 \log\frac{1}{x} + l_1\right) + 1_{\{x>1\}}\frac{1}{C_3}\exp\left(-C_3\left|x\right|\right), & \alpha = 0; \\ l_2\left(1_{\{x\le 1\}}x^{-\alpha} + 1_{\{x>1\}}\exp\left(-C_3\left|x\right|\right)\right), & 0<\alpha<2. \end{cases}$$

If $\alpha = 0$, then

$$\begin{aligned} L_\varepsilon(p) &= p\int_{x>\varepsilon} x^{p-1}\lambda_x^2 \mathrm{d}x \\ &\le l_1^2 p\int_{x>\varepsilon} x^{p-1}\left(1_{\{x\le 1\}}\left(C_1\log\frac{1}{x}+l_1\right) + 1_{\{x>1\}}\frac{1}{C_3}\exp\left(-C_3\left|x\right|\right)\right)^2 \mathrm{d}x \\ &\le l_3; \end{aligned}$$

If $\alpha > 0$, then

$$\begin{aligned} L_\varepsilon(p) &= p\int_{x>\varepsilon} x^{p-1}\lambda_x^2 \mathrm{d}x \\ &\le l_2^2 p\int_{x>\varepsilon} x^{p-1}\left(1_{\{x<1\}}x^{-\alpha} + 1_{\{x>1\}}\exp\left(-C_3\left|x\right|\right)\right)^2 \mathrm{d}x \\ &\le \begin{cases} l_4, & p>2\alpha; \\ l_4\log\frac{1}{\varepsilon} + l_5, & p=2\alpha; \\ l_4\varepsilon^{-2\alpha+p} + l_5, & p<2\alpha. \end{cases} \end{aligned}$$

where l_i, $i = 1, \ldots, 5$ are constants. This upper bound on $L_\varepsilon(p)$ also includes the $\alpha = 0$ case.

Remark 5.3.8. *The upper bound on $L_\varepsilon(p)$ requires that the right tail of the Lévy measure decays fast enough. For example, for an α-stable process we have*

$$\lambda_x \le l_2 1_{\{x\le 1\}} x^{-\alpha},\ 0<\alpha<2$$

and

$$\begin{aligned} L_\varepsilon(p) &= p\int_{x>\varepsilon} x^{p-1}\lambda_x^2 \mathrm{d}x \\ &\le l_2^2 p\int_{x>\varepsilon} x^{p-1}\left(1_{\{x<1\}}x^{-\alpha}\right)^2 \mathrm{d}x \\ &\le l_4\varepsilon^{-2\alpha+p},\ p<2\alpha. \end{aligned} \tag{5.25}$$

Hence for $0 < \alpha \le 1$ we cannot obtain an upper bound on $L_\varepsilon(p)$ for $p \ge 2$. Fortunately, in the numerical experiment we are interested in the application where spectrally negative processes don't have a right tail and $\alpha > 1$.

Since

$$\begin{aligned}
\int_{\varepsilon<|z|<1} z\ \nu(\mathrm{d}z) &\leq C_1 \int_{\varepsilon<z<1} z\ \frac{1}{z^{\alpha+1}}\mathrm{d}z + C_2 \int_{-1<z<-\varepsilon} z\ \frac{1}{|z|^{\alpha+1}}\mathrm{d}z; \\
-\int_{\varepsilon<|z|<1} z\ \nu(\mathrm{d}z) &\leq -C_2 \int_{\varepsilon<z<1} z\ \frac{1}{z^{\alpha+1}}\mathrm{d}z - C_1 \int_{-1<z<-\varepsilon} z\ \frac{1}{|z|^{\alpha+1}}\mathrm{d}z \\
&= C_1 \int_{\varepsilon<z<1} z\ \frac{1}{z^{\alpha+1}}\mathrm{d}z + C_2 \int_{-1<z<-\varepsilon} z\ \frac{1}{|z|^{\alpha+1}}\mathrm{d}z,
\end{aligned}$$

assuming $0<\varepsilon<1$ we have

$$\begin{aligned}
|\mu_\varepsilon| &= \left| m - \int_{\varepsilon<|z|<1} z\ \nu(\mathrm{d}z)\right| \\
&\leq |m| + \left| C_2 \int_{\varepsilon<z<1} z\ \frac{1}{z^{\alpha+1}}\mathrm{d}z + C_1 \int_{-1<z<-\varepsilon} z\ \frac{1}{|z|^{\alpha+1}}\mathrm{d}z \right| \\
&= \begin{cases} |m| + \left|\frac{C_1-C_2}{\alpha-1}\left[\varepsilon^{1-\alpha}-1\right]\right|, & \alpha \neq 1; \\ |m| + |C_1-C_2|\log\frac{1}{\varepsilon}, & \alpha = 1. \end{cases}
\end{aligned} \tag{5.26}$$

By (5.18) and (5.22),

$$\mathbb{E}\left[V_n^p\right] \prec \underbrace{\mathbb{E}\left[\left(\max_{i=1,\dots,n} S_n^{(i)}\right)^p\right]}_{1)} + \underbrace{\varepsilon^p \frac{\lambda_\varepsilon^2}{n}}_{2)} + \underbrace{\frac{L_\varepsilon(p)}{n}}_{3)} + \underbrace{\left(\frac{|\mu_\varepsilon|}{n}\right)^p}_{4)}.$$

For the ease of balancing asymptotic orders, we assume ε is sufficiently small such that

$$\begin{array}{l} \log\frac{1}{\varepsilon} > \max\left(l_5, l_1\right), \ \text{if } p = 2\alpha \text{ or } \alpha = 0; \\ \varepsilon^{-2\alpha+p} > l_5, \ \text{if } p < 2\alpha. \end{array} \tag{5.27}$$

Under this assumption, we have

1)

$$\mathbb{E}\left[\left(\max_{i=1,\dots,n} S_n^{(i)}\right)^p\right] \prec \varepsilon^{p-\delta}.$$

2)

$$\varepsilon^p \frac{\lambda_\varepsilon^2}{n} \prec \begin{cases} n^{-1}\varepsilon^p \log\frac{1}{\varepsilon}, & \alpha = 0; \\ n^{-1}\varepsilon^{p-2\alpha}, & 0<\alpha<2. \end{cases}$$

3)

$$\frac{L_\varepsilon(p)}{n} \prec n^{-1} \times \begin{cases} 1, & p > 2\alpha; \\ \log\frac{1}{\varepsilon}, & p = 2\alpha; \\ \varepsilon^{-2\alpha+p}, & p < 2\alpha. \end{cases}$$

4)

$$\left(\frac{|\mu_\varepsilon|}{n}\right)^p \prec n^{-p} \times \begin{cases} 1 + |C_1 - C_2|^p \, \varepsilon^{p(1-\alpha)}, \ \alpha > 1; \\ 1 + \left(|C_1 - C_2| \log \frac{1}{\varepsilon}\right)^p, \ \alpha = 1; \\ 1, \ \alpha < 1. \end{cases}$$

We balance the RHS of four terms as p, α take value in different regions. All possible scenarios are illustrated in Figure 5.12.

In the following we assume $C_1 \neq C_2$.

1. $p \geq 2\alpha$.

   In this case, the order of 2) is greater than or equal to the one of 3).

   If $\alpha = 0$, we can take $\varepsilon = n^{-(p-\delta)^{-1}}$ which gives

   $$\mathbb{E}\left[V_n^p\right] \prec \begin{cases} n^{-1}, \ p \geq 1; \\ n^{-p}, \ p < 1. \end{cases}$$

   If $\alpha > 0$, we can take

   $$\varepsilon = n^{-1/\alpha}.$$

   According to (5.24),

   $$\lambda_\varepsilon \prec \varepsilon^{-\alpha} = n,$$

   satisfying the constraint $\lambda_\varepsilon = \mathcal{O}(n)$.

   $$\begin{aligned} \mathbb{E}\left[\left(\max_{i=1,\ldots,n} S_n^{(i)}\right)^p\right] &\prec \varepsilon^{p-\delta} \\ &= n^{-(p-\delta)/\alpha}. \end{aligned}$$

   is less than n^{-1} if $\delta < \alpha$ is chosen. Under this scaling of ε, the leading order among 1) to 3) is

   $$\begin{cases} n^{-1}, \ p > 2\alpha; \\ n^{-1} \log n, \ p = 2\alpha. \end{cases} \tag{5.28}$$

   If $\alpha > 1$, then

   $$\begin{aligned} \left(\frac{|\mu_\varepsilon|}{n}\right)^p &\prec n^{-p} \varepsilon^p n^p \\ &= n^{-p/\alpha} \end{aligned}$$

   is negligible compared to (5.28).

Hence the order of 4) is greater than or equal to n^{-1}.

If $\alpha = 1$, then

$$\left(\frac{|\mu_\varepsilon|}{n}\right)^p \prec n^{-p}\left(1 + \left(|C_1 - C_2| \log\frac{1}{\varepsilon}\right)^p\right).$$

Hence the order of 4) is greater than or equal to (5.28).

If $\alpha < 1$, then

$$\left(\frac{|\mu_\varepsilon|}{n}\right)^p \prec n^{-p}$$

is less than or equal to n^{-1} when $p \geq 1$. In those cases,

$$\mathbb{E}\left[V_n^p\right] = \begin{cases} \mathcal{O}\left(\frac{1}{n}\right), & p > 2\alpha; \\ \mathcal{O}\left(\frac{\log n}{n}\right), & p = 2\alpha. \end{cases}$$

Remark 5.3.9. *When* $2\alpha \leq p < 1$,

$$\mathbb{E}\left[V_n^p\right] = \mathcal{O}\left(n^{-p}\right).$$

2. $0 < p < 2\alpha$.

In this case, 3) $\prec n^{-1}\varepsilon^{p-2\alpha}$ and 2) $\prec n^{-1}\varepsilon^{p-2\alpha}$ are of the same order. Balancing the RHS of 1) and 3) we get

$$\varepsilon \sim n^{-(2\alpha-\delta)^{-1}}.$$

According to (5.21), the constraint on ε is satisfied $\varepsilon^{-\alpha} \sim n^{\alpha(2\alpha-\delta)^{-1}} \prec n$ if δ is chosen to fulfill $\alpha\left(2\alpha - \delta\right)^{-1} \leq 1$ i.e.

$$\delta \leq \alpha. \tag{5.29}$$

Under this scaling, the leading order among 1) to 3) is

$$n^{-(p-\delta)(2\alpha-\delta)^{-1}}.$$

Now we examine the order of 4) :

As $\alpha > 1$,

$$\begin{aligned} \left(\frac{|\mu_\varepsilon|}{n}\right)^p &\prec n^{-p}\left(1 + |C_1 - C_2|^p \varepsilon^{p(1-\alpha)}\right) \\ &\sim n^{-p} + n^{-p+p(\alpha-1)(2\alpha-\delta)^{-1}}. \end{aligned}$$

Note that

$$-p + p(\alpha - 1)(2\alpha - \delta)^{-1} \le -(p - \delta)(2\alpha - \delta)^{-1}$$

$\Leftrightarrow$

$$\delta \le \alpha.$$

Hence by choosing sufficiently small δ the contribution from 4) is not the leading order.

As $\alpha = 1$,

$$\left(\frac{|\mu_\varepsilon|}{n}\right)^p \prec n^{-p}\left(1 + \left(|C_1 - C_2| \log \frac{1}{\varepsilon}\right)^p\right),$$

is a higher order term.

As $\alpha < 1$,

$$\left(\frac{|\mu_\varepsilon|}{n}\right)^p \prec n^{-p}.$$

Note that

$$-p \le -(p - \delta)(2\alpha - \delta)^{-1}$$

$\Leftrightarrow$

$$(p - 1)\delta \le (2\alpha - 1)p \text{ or } \delta \ge 2\alpha.$$

If $1 > \alpha > \frac{1}{2}$, then by choosing sufficiently small δ, the contribution from 4) is not the leading order. Thus

$$\mathbb{E}[V_n^p] = \mathcal{O}\left(n^{-(p-\delta)(2\alpha-\delta)^{-1}}\right).$$

If $2\alpha \le 1$, then $p < 2\alpha \le 1$. Due to constraint (5.29), by solving

$$\alpha(p - 1) \le (p - 1)\delta \le (2\alpha - 1)p$$

we have

$$p \le \frac{\alpha}{1 - \alpha}.$$

We conclude that for $p \le \frac{\alpha}{1-\alpha}$, we still have

$$\mathbb{E}[V_n^p] = \mathcal{O}\left(n^{-(p-\delta)(2\alpha-\delta)^{-1}}\right).$$

For $p > \frac{\alpha}{1-\alpha}$, n^{-p} is the leading order term so

$$\mathbb{E}[V_n^p] = \mathcal{O}\left(n^{-p}\right).$$

On the other hand, note that when $2\alpha \geq 1$, for $q = 2\alpha > p$,

$$\begin{aligned}\mathbb{E}\left[V_n^p\right] &\leq \mathbb{E}\left[V_n^q\right]^{p/q} \\ &= \mathcal{O}\left(\left(\frac{\log n}{n}\right)^{\frac{p}{2\alpha}}\right).\end{aligned}$$

When $2\alpha < 1$, for $q = 1 > p$,

$$\begin{aligned}\mathbb{E}\left[V_n^p\right] &\leq \mathbb{E}\left[V_n\right]^p \\ &= \mathcal{O}\left(n^{-p}\right).\end{aligned}$$

This shows that by simply using Hölder's inequality on the results obtained for the $p = 2\alpha$ case, we can get better bounds compared to what we get from balancing terms, whenever $2\alpha \geq 1$ or $2\alpha < 1$.

For a general α-stable process with $\alpha > 1$, as shown by (5.25) in the remark, the only difference is that we can only use the estimate of $L_\varepsilon(p)$ when $p < 2\alpha$. Thus it is not possible to use Hölder's inequality on the results obtained for the $p = 2\alpha$ case to get a better bound. In this case, the result would be

$$\mathbb{E}\left[V_n^p\right] = \mathcal{O}\left(n^{-(p-\delta)(2\alpha-\delta)^{-1}}\right).$$

□

We use a lemma which can be proved by Hölder's inequality to interpolate the result of Theorem 5.3.1 and Proposition 5.3.7.

Lemma 5.3.10. *Suppose $r_1 < r < r_2$, and $\mathbb{E}\left[X^{r_2}\right]$ exists, then we have*

$$\mathbb{E}\left[X^r\right] \leq \mathbb{E}\left[X^{r_1}\right]^{\frac{r_2-r}{r_2-r_1}} \mathbb{E}\left[X^r\right]^{\frac{r-r_1}{r_2-r_1}}.$$

Taking advantage of Theorem 5.3.1 and this lemma, we obtain better results:

Proposition 5.3.11. *Let X_t be a scalar pure jump Lévy process. Suppose the Lévy measure $\nu(x)$ satisfies*

$$\begin{aligned}C_2\left|x\right|^{-1-\alpha} \leq \nu(x) \leq C_1\left|x\right|^{-1-\alpha}, \ \textit{as} \ |x| \leq 1; \\ \nu(x) \leq \exp\left(-C_3\left|x\right|\right), \ \textit{as} \ |x| > 1,\end{aligned}$$

where $C_1, C_2, C_3 > 0, \ 0 \leq \alpha < 2$ are constants.

Then $V_n = \sup_{0\leq t\leq 1} X_t - \max_{i=0,1,\ldots,n} X_{\frac{i}{n}}$ satisfies *a)* $p \geq 2\alpha$, $p \geq 1$,

$$\mathbb{E}\left[V_n^p\right] = \begin{cases} \mathcal{O}\left(\frac{1}{n}\right), & p > 2\alpha; \\ \mathcal{O}\left(\frac{\log n}{n}\right), & p = 2\alpha. \end{cases}$$

b) $1 \leq p < 2\alpha$ *and* $\frac{1}{2} \leq \alpha < 1$,

$$\mathbb{E}\left[V_n^p\right] = \mathcal{O}\left(\left(\frac{\log n}{n}\right)^{\frac{p}{2\alpha}}\right).$$

c) $1 \leq p < 2\alpha$ *and* $1 \leq \alpha < 2$,

$$\mathbb{E}\left[V_n^p\right] = \mathcal{O}\left(\left(\frac{1}{n}\right)^{r\frac{2\alpha-p}{2\alpha-1}}\left(\frac{\log n}{n}\right)^{\frac{p-1}{2\alpha-1}}\right).$$

d) $p < 1$ *and* $0 \leq \alpha < 1$,

$$\mathbb{E}\left[V_n^p\right] = \mathcal{O}\left(\left(\frac{\log n}{n}\right)^{p}\right).$$

e) $p < 1$ *and* $1 \leq \alpha < 2$,

$$\mathbb{E}\left[V_n^p\right] = \mathcal{O}\left(n^{-rp}\right).$$

where $\frac{1}{2} < r = \frac{1}{\alpha} - \theta < \frac{1}{\alpha} \leq 1$ *for any small* $\theta > 0$.

In particular for $p = 2$,

1. *if* $0 \leq \alpha < 1$, *then*

$$\mathbb{E}\left[V_n^2\right] = \mathcal{O}\left(\frac{1}{n}\right);$$

2. *if* $\alpha = 1$, *then*

$$\mathbb{E}\left[V_n^2\right] = \mathcal{O}\left(\frac{\log n}{n}\right);$$

3. *if* $1 < \alpha < 2$, *then*

$$\mathbb{E}\left[V_n^2\right] = \mathcal{O}\left(\left(\frac{1}{n}\right)^{r\frac{2\alpha-2}{2\alpha-1}}\left(\frac{\log n}{n}\right)^{\frac{1}{2\alpha-1}}\right).$$

For a variance gamma process with parameters A, B and κ, $\nu(z) = \frac{1}{\kappa|x|}e^{A-B|x|}$, let $\alpha = 0$. There exists C_1, C_2, C_3 such that (5.23) is satisfied. This corresponds to the first case of Proposition 5.3.7.

For a NIG process, $\nu(x) = \frac{C}{|x|}e^{Ax}\ K_1\left(B\left|x\right|\right)$, where $K_1\left(\cdot\right)$ is the modified Bessel function of the second kind. Recall that $K_1\left(\cdot\right)$ has near-zero and tail asymptotic expansion (5.1), (5.2). Let $\alpha = 1$. There exists C_1, C_2, C_3 such that (5.23) is satisfied. This corresponds to the second case of Proposition 5.3.7.

For a tempered α-stable process, $\nu(x) = \frac{A}{x^{\alpha+1}} e^{-\lambda_+ x} 1_{\{x>0\}} + \frac{B}{|x|^{\alpha+1}} e^{\lambda_- x} 1_{\{x<0\}}$. There exists C_1, C_2, C_3 such that (5.23) is satisfied. This corresponds to the third case of Proposition 5.3.11.

For a spectrally negative α-stable process ($A = 0$) which we are interested in the numerical results, we have a strong result:

Proposition 5.3.12. *Let X_t be a spectrally negative α-stable process, i.e. its Lévy measure $\nu(x)$ is*

$$\nu(x) = \frac{B}{|x|^{\alpha+1}} 1_{\{x<0\}}$$

where $B > 0,\ 0 < \alpha < 2$ are constants. Then $V_n = \sup_{0 \le t \le 1} X_t - \max_{i=0,1,\dots,n} X_{\frac{i}{n}}$ satisfies

1. *if $0 < \alpha < 1$, then*
$$\mathbb{E}\left[V_n^p\right] = \mathcal{O}\left(n^{-p}\right);$$
2. *if $\alpha = 1$, then*
$$\mathbb{E}\left[V_n^p\right] = \mathcal{O}\left(\left(\frac{\log n}{n}\right)^p\right);$$
3. *if $1 < \alpha < 2$, then*
$$\mathbb{E}\left[V_n^p\right] = o\left(n^{-rp}\right)$$
where $\frac{1}{2} < r = \frac{1}{\alpha} - \theta < \frac{1}{\alpha} \le 1$ for any small $\theta > 0$.

Proof. Since X_t doesn't have positive jumps, $\sup_{[0,\frac{1}{n}]} X_t^\varepsilon - \left(X_{\frac{1}{n}}^\varepsilon\right)^+$ in (5.13) is zero. We have

$$\mathbb{E}\left[V_n^p\right] \le \mathbb{E}\left[\left(\max_{i=1,\dots,n} S_n^{(i)}\right)^p\right] + \left(\frac{|\mu_\varepsilon|}{n}\right)^p.$$

According to Proposition 5.3.4, and the proof of Proposition 5.3.7, we have for any $\delta > 0$ and $0 < \varepsilon < 1$,

$$\mathbb{E}\left[\left(\max_{i=1,\dots,n} S_n^{(i)}\right)^p\right] \prec \varepsilon^{p-\delta}.$$

$$\left(\frac{|\mu_\varepsilon|}{n}\right)^p \prec n^{-p} \times \begin{cases} \varepsilon^{p(1-\alpha)},\ \alpha > 1; \\ \left(\log \frac{1}{\varepsilon}\right)^p,\ \alpha = 1; \\ 1,\ \alpha < 1. \end{cases}$$

We can make

$$\varepsilon = n^{-\left(\alpha-\delta p^{-1}\right)^{-1}}$$

under which

$$\mathbb{E}\left[V_n^p\right] \prec \begin{cases} n^{-\frac{p-\delta}{\alpha-\delta p^{-1}}}, \ \alpha > 1; \\ \left(\frac{\log n}{n}\right)^p, \ \alpha = 1; \\ n^{-p}, \ \alpha < 1. \end{cases}$$

By presetting a sufficiently small δ we get the result. □

The result of Proposition 5.3.12 is near optimal in a sense that it matches the numerical result. The numerical results show that for a spectrally negative α-stable process with α tending to 2, $\mathbb{E}\left[U_n^2\right]$ behaves like $\mathcal{O}\left(n^{-1}\right)$.The intuition for this is that when α is close to 2 the process has so many small jumps that it behaves like a Brownian motion. However, for a (tempered) α-stable process the result of Proposition 5.3.11 is not optimal. The limitation of current analysis might be due to the estimate (5.10) and since the bound of Theorem 5.3.5 on the contribution from the large jumps is not tight enough.

5.3.4 Lookback options

The lookback option payoff in the numerical experiments is a Lipschitz function of $\max_{0\le t\le T} S(t)$, i.e.

$$P = P(\max_{0\le t\le T} S(t)) \tag{5.30}$$

such that $|P(S_1) - P(S_2)| \le L_K \left|S_1 - S_2\right|$.

To avoid moment explosion of unbounded payoffs, we also consider the lookback put option, where the payoff is a Lipschitz function of the exponential exponent $\max_{0\le t\le T} X(t)$, i.e.

$$P = P(\max_{0\le t\le T} X(t)) \tag{5.31}$$

such that $|P(X_1) - P(X_2)| \le K\left|X_1 - X_2\right|$. This is since for $P\left(x\right) = \left(K - S_0 \exp\left(x\right)\right)^+$ we have $\left|P'\left(x\right)\right| \le K$.

Without loss of generality, we assume $T = 1$ in the following.

5.3.4.1 Weak convergence

For the put option case, by the Lipschitz property and boundness of the payoff we have

$$\mathbb{E}\left[\widehat{P} - P\right] \leq K\ \mathbb{E}\left[\left|\max_{i=0,1,\ldots,n} X_{\frac{i}{n}} - \sup_{0\leq t\leq 1} X_t\right|\right].$$

Therefore we obtain weak convergence for the process discussed by Theorem 5.3.1, and the convergence rate is exactly the same. Hence we have almost first order weak convergence for the estimator of Lipschitz lookback put payoff in the VG and NIG model.

For the call option case, firstly by $|e^x - 1| \leq |x|$ for $x \leq 0$, we have

$$\begin{aligned}\mathbb{E}\left[\widehat{P} - P\right] &\leq L_K\ \mathbb{E}\left[\left|\exp\left(\max_{i=0,1,\ldots,n} X_{\frac{i}{n}}\right) - \exp\left(\sup_{0\leq t\leq 1} X_t\right)\right|\right] \\ &\leq L_K\ \mathbb{E}\left[\exp\left(\sup_{0\leq t\leq 1} X_t\right) V_n\right].\end{aligned}$$

Then we state Lemma 6.3 of [18] which uses Hölder's inequality to reduce the analysis to $\mathbb{E}\left[V_n\right]$:

Lemma 5.3.13. *If* $\exists q > 1$ *such that*

$$\mathbb{E}\left[\exp\left(q \sup_{0\leq t\leq 1} X_t\right)\right] < \infty, \tag{5.32}$$

then for any $\delta > 0$, *there exists constant* C *such that*

$$\mathbb{E}\left[\widehat{P} - P\right] \leq C\ \mathbb{E}\left[V_n\right]^{\frac{q-1}{q} - \delta}.$$

We will see in the next section that the weak convergence holds for the processes using parameters in the numerical simulation.

5.3.4.2 Moment explosion in exponential Lévy models

In the case of diffusion processes, the moment generating function

$$\mathbb{E}\left[\exp\left(q \sup_{0\leq t\leq 1} X_t\right)\right]$$

is always finite for any q. However, the moment generating function of a Lévy process can be infinite for some q. We discuss the conditions under which the Lévy process admits a finite exponential moment.

The solution of an SDE driven by a Lévy process admits a moment estimate; see e.g. Theorem 67 on page 349 of [52]. We state the estimate addressing the condition on the Lévy measure.

Theorem 5.3.14. *Let σ be continuous differentiable with a bounded derivative. Suppose X_t is a Lévy process with a Lévy measure ν such that $\int_{|z|>1} z^k \nu(\mathrm{d}z) < \infty$ for an integer $k > 0$. Then the solution of the SDE*

$$\mathrm{d}S_t = \sigma\left(S_{t-}\right)\mathrm{d}X_t$$

admits

$$\mathbb{E}\left[\sup_{0\le t\le T} |S_t|^p\right] < C\left(1 + |S_0|^p\right).$$

where $p \le k$ is an integer, C is a positive constant.

It is reasonable that if the Lévy measure has a finite exponential moment then the running maximum of the Lévy process has a finite exponential moment. In fact, for a given $q > 1$, by Doob's inequality and Jensen's inequality,

$$\mathbb{E}\left[\exp\left(q \sup_{0\le t\le 1} X_t\right)\right] < \infty,$$

holds if and only if

$$\mathbb{E}\left[\exp\left(qX_t\right)\right] < \infty.$$

By Theorem 3.6 in [47], the above is equivalent to

$$\int_{|z|>1} e^{qz}\nu(\mathrm{d}z) < \infty.$$

Hence (5.32) is equivalent to $\exists q > 1$ such that

$$\int_{|z|>1} e^{qz}\nu(\mathrm{d}z) < \infty. \tag{5.33}$$

For a Variance Gamma process with parameters (σ, θ, κ), $\nu(x) = \frac{1}{\kappa|x|}e^{Ax-B|x|}$, where $A = \frac{\theta}{\sigma^2}$ and $B = \frac{\sqrt{\theta^2+2\sigma^2/\kappa}}{\sigma^2}$, in order that $\exists q > 1$ such that

$$\int_{|x|>1} e^{qx}\frac{1}{\kappa\,|x|}e^{Ax-B|x|}\mathrm{d}x < \infty,$$

it is necessary that $B - A > 1$, which is equivalent to $\sqrt{\theta^2 + 2\sigma^2/\kappa} + \theta < 2$. Recall the parameters σ, κ, θ represent the standard deviation of BM and Gamma subordinator, drift of BM respectively. In fact, in our numerical test, according to option prices data calibration in [48], parameters are estimated as $\kappa = 0.1686$; $\sigma = 0.1213$; $\theta = -0.1436$. A calculation shows that $A = -9.7596, B = 30.0244$ so (5.33) is satisfied.

For a NIG process with a parameter set (σ, θ, κ), $\nu(x) = \frac{C}{|x|} e^{Ax}\ K_1(B|x|)$ with $A = \frac{\theta}{\sigma^2}$, $B = \frac{\sqrt{\theta^2 + \sigma^2/\kappa}}{\sigma^2}$, and $C = \frac{\sqrt{\theta^2 + 2\sigma^2/\kappa}}{2\pi\sigma\sqrt{\kappa}}$ where $K_1(\cdot)$ is the modified Bessel function of the second kind. Recall that $K_1(\cdot)$ has tail asymptotic expansion (5.2). In order to guarantee the boundness of

$$\begin{aligned} & \int_1^\infty e^{qx} \frac{1}{\kappa |x|} e^{Ax} K_1(B|x|) \, \mathrm{d}x \\ \sim \; & \int_1^\infty e^{qx} \frac{1}{\kappa |x|} e^{Ax - B|x|} \sqrt{\frac{\pi}{2|x|}} \left(1 + \mathcal{O}\left(\frac{1}{|x|}\right)\right) \mathrm{d}x, \end{aligned}$$

for a $q > 1$, it is necessary that $B - A > 1$. In our numerical test $A = -3.8941$, $B = 6.1882$, and $\sigma = 0.1622$ which fulfills the condition.

For an α-stable process, only in the case when $A = 0$, we have all exponential moments finite according to Proposition 1 in [8].

In summary, for the processes used in our numerical results, the condition for weak convergence is guaranteed.

The reason why the exponential moment of the Lévy process can blow up is due to the heavy tail property. In the exponential Lévy model, the expected value of a discounted payoff is not always finite under a given risk-neutral measure, e.g. the mean correcting measure discussed above. A way to resolve this issue is to choose another risk-neutral measure for calibration, e.g. the pricing measure defined by the Esscher transform which does an exponential tilting on the tail to make the moment finite [14]. Alternatively, the Lévy-driven SDE model only requires the boundness of up to kth moments of Lévy measure to provide finite up to kth order moments for the solution [17, 16].

It is worth noting that similar phenomena are studied in the stochastic volatility model literature [24, 44].

5.3.4.3 Analysis of Variance

Recall that the performance of the multilevel method relies on the variance convergence rate of the multilevel correction term:

$$\mathbb{V}\left[\widehat{P}_l - \widehat{P}_{l-1}\right].$$

We are interested in its upper bound $\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^2\right]$. The results of VG, NIG and α-stable processes for lookback options are summarised in the following:

Proposition 5.3.15. *Let X_t be a scalar Lévy process by which the exponential Lévy model is driven. For the Lipschitz lookback put payoff (5.31), we have multilevel variance convergence rate results:*

- *If X_t is a Variance Gamma (VG) process, then*

$$\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^2\right] = \mathcal{O}\left(h_l\right);$$

- *If X_t is a Normal Inverse Gaussian (NIG) process, then*

$$\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^2\right] = \mathcal{O}\left(h_l \left|\log h_l\right|\right);$$

- *If X_t is a spectrally negative α-stable process with $\alpha > 1$, then*

$$\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^2\right] = o\left(h_l^{2/\alpha-\delta}\right)$$

  for any small $\delta > 0$.

  If $\exists q > 1$ such that

$$\int_{|z|>1} e^{2qz}\nu(\mathrm{d}z) < \infty \tag{5.34}$$

  then for the lookback call payoff (5.30), we have multilevel variance convergence results:

- *If X_t is a Variance Gamma (VG) process, then*

$$\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^2\right] = \mathcal{O}\left(h_l^{1-\frac{1}{q}}\right);$$

- *If X_t is a Normal Inverse Gaussian (NIG) process, then*

$$\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^2\right] = \mathcal{O}\left(h_l^{1-\frac{1}{q}}\right);$$

- *If X_t is a spectrally negative α-stable process with $\alpha > 1$, then for any small $\delta > 0$*

$$\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^2\right] = o\left(h_l^{2/\alpha-\delta}\right),$$

where q is the constant depending on the parameters of the process in (5.34).

The best choice of q is $(B-A)/2$ for both VG and NIG processes, where A, B are defined in sections 5.1.2.1 and 5.1.2.2.

Proof. Denote $n = 2^{l-1}$. For the put option case, by the Lipschitz property and boundness of the payoff we have

$$\begin{aligned}\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^2\right] &\leq L_K^2\, \mathbb{E}\left[U_n^2\right] \\ &\leq L_K^2 \mathbb{E}\left[V_n^2\right].\end{aligned}$$

Thus the results follow from Proposition 5.3.7.

For the call option case,

Since $|e^x - 1| \leq |x|$ for $x \leq 0$,

$$\begin{aligned}\left|\widehat{P}_l - \widehat{P}_{l-1}\right| &\leq L_K \left|\exp\left(\max_{i=0,1,\dots,2n} X_{\frac{i}{2n}}\right) - \exp\left(\max_{i=0,1,\dots,n} X_{\frac{i}{n}}\right)\right| \\ &\leq L_K \exp\left(\max_{i=0,1,\dots,2n} X_{\frac{i}{2n}}\right)\left|\max_{i=0,1,\dots,2n} X_{\frac{i}{2n}} - \max_{i=0,1,\dots,n} X_{\frac{i}{n}}\right|.\end{aligned}$$

By Hölder's inequality,

$$\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^2\right] \leq L_K\, \mathbb{E}\left[\exp\left(2q \max_{i=0,1,\dots,n} X_{\frac{i}{n}}\right)\right]^{\frac{1}{q}} \mathbb{E}\left[|U_n|^{\frac{2q}{q-1}}\right]^{\frac{q-1}{q}}.$$

For $q > 1$, by Doob's inequality and Theorem 3.6 in [47],

$$\mathbb{E}\left[\exp\left(2q \max_{i=0,1,\dots,2n} X_{\frac{i}{2n}}\right)\right] < \infty, \tag{5.35}$$

is equivalent to

$$\int_{|z|>1} e^{2qz}\nu(\mathrm{d}z) < \infty.$$

When $0 \leq \alpha \leq 1$, since $\frac{2q}{q-1} > 2 \geq 2\alpha$, case a) of Proposition 5.3.7 implies the results of VG ($\alpha = 0$) and NIG ($\alpha = 1$).

For a spectrally negative α-stable process ($1 < \alpha < 2$), we have all exponential moments finite according to Proposition 1 in [8]. The result follows from Proposition 5.3.12. □

It is easy to check whether (5.34) is satisfied for a process with particular parameters. In fact, following the calculations of the last section, the condition is fulfilled for all processes in the numerical results.

5.3.5 Asian options

We consider the analysis of Lipschitz arithmetic Asian payoff, which means the payoff of the Asian option is

$$P = P\left(\overline{S}\right) \tag{5.36}$$

where P is Lipschitz such that $|P(S_1) - P(S_2)| \le L_K\, |S_1 - S_2|$ and

$$\overline{S} = S_0\, T^{-1} \int_0^T \exp\left(X_t\right) \mathrm{d}t\ .$$

We approximate the payoff using a trapezoidal approximation:

$$\overline{\widehat{S}} := S_0 \sum_{j=0}^{n-1} \tfrac{1}{2}\, h\ \left(\exp\left(X_{jh}\right) + \exp\left(X_{(j+1)h}\right)\right). \tag{5.37}$$

Proposition 5.3.16. *Let X_t be a scalar Lévy process with triple (m, Σ, ν) as in (5.1.2).*

Suppose the asset price follows an exponential Lévy model, i.e. $S_t = S_0 \exp\left(X_t\right)$.

For Arithmetic Asian option $P = P\left(\overline{S}\right)$ (5.36), suppose we use (5.37) as an n-step trapezoidal approximation of $\overline{S}$, and the subsequent approximated payoff is $\widehat{P} = f\left(\overline{\widehat{S}}\right)$.

If $\int_{|z|>1} e^{2z}\ \nu(\mathrm{d}z) < \infty$, then we have

$$\mathbb{E}\left[\left(P - \widehat{P}\right)^2\right] = \mathcal{O}(h^2).$$

Proof.

$$\begin{aligned}
\left|\overline{S}-\widehat{\overline{S}}\right|/S_0 &= T^{-1}\left|\sum_{j=0}^{n-1}\tfrac{1}{2}\,h\ \left(\exp\left(X_{jh}\right)+\exp\left(X_{(j+1)h}\right)\right)-\int_0^T \exp\left(X_t\right)\mathrm{d}t\right| \\
&= T^{-1}\left|\sum_{j=0}^{n-1}\tfrac{1}{2}\,h\ \left(\exp\left(X_{jh}\right)+\exp\left(X_{(j+1)h}\right)-\int_{jh}^{(j+1)h}\exp\left(X_t\right)\mathrm{d}t\right)\right| \\
&= T^{-1}\Bigg|\sum_{j=0}^{n-1}\exp\left(X_{jh}\right)\int_{jh}^{(j+1)h}\left(\exp\left(X_t-X_{jh}\right)-1\right)\mathrm{d}t \\
&\quad -\frac{1}{2}h\exp\left(X_T\right)+\frac{1}{2}h\Bigg|.
\end{aligned}$$

Let us denote

$$\begin{aligned}
b_j &= \exp\left(X_{jh}\right); \\
I_j &= \int_{jh}^{(j+1)h}\left(\exp\left(X_t-X_{jh}\right)-1\right)\mathrm{d}t; \\
R_A &= -\frac{1}{2}h\exp\left(X_T\right)+\frac{1}{2}h.
\end{aligned}$$

$$\mathbb{E}\left[R_A^2\right]=\mathcal{O}\left(h^2\right).$$

We are concerned with the second moment:

$$\begin{aligned}
\mathbb{E}\left[\left(\widehat{\overline{S}}-\overline{S}\right)^2\right] &= T^{-2}S_0^2\,\mathbb{E}\left[\left|\sum_{j=0}^{n-1}b_jI_j+R_A\right|^2\right] \\
&\leq 2T^{-2}S_0^2\left(\mathbb{E}\left[\left|\sum_{j=0}^{n-1}b_jI_j\right|^2\right]+\mathbb{E}\left[R_A^2\right]\right).
\end{aligned}$$

$$\mathbb{E}\left[\left|\sum_{j=0}^{n-1}b_jI_j\right|^2\right]=\mathbb{E}\left[\sum_{j=0}^{n-1}b_j^2I_j^2+2\sum_{m=0}^{n-1}\sum_{j=0}^{m-1}b_mI_mb_jI_j\right] \tag{5.38}$$

By independence of b_j and I_j we have

$$\mathbb{E}\left[\sum_{j=0}^{n-1}b_j^2I_j^2\right]=\sum_{j=0}^{n-1}\mathbb{E}\left[b_j^2\right]\mathbb{E}\left[I_j^2\right].$$

The first part $\mathbb{E}\left[b_j^2\right]=\mathbb{E}\left[\exp\left(2X_{jh}\right)\right]=e^{Ajh}$, where $A=2m+\int\left(e^{2z}-2z1_{|z|<1}\right)\nu\left(\mathrm{d}z\right)$. Here we assume $\int_{|z|>1}e^{2z}\ \nu(\mathrm{d}z)<\infty$.

For the second part, by Cauchy-Schwartz inequality,

$$
\begin{aligned}
\mathbb{E}\left[I_j^2\right] &= \mathbb{E}\left[\left(\int_{jh}^{(j+1)h} \left(\exp\left(X_t - X_{jh}\right) - 1\right) \mathrm{d}t\right)^2\right] \\
&\leq \mathbb{E}\left[\int_{jh}^{(j+1)h} \left(\exp\left(X_t - X_{jh}\right) - 1\right)^2 \mathrm{d}t\right] h \\
&= \mathbb{E}\left[\int_0^h \left(\exp\left(X_t\right) - 1\right)^2 \mathrm{d}t\right] h \\
&= \left[\frac{1}{A}\left(e^{Ah} - 1\right) - 2\frac{1}{r}\left(e^{rh} - 1\right) + h\right] h \\
&\leq C(A, r)\, h^3, \text{ for small } h
\end{aligned}
$$

where the penultimate step is by $\mathbb{E}[\exp(X_t)] = \exp(rt)$ under the risk-neutral measure. The last step is due to Taylor expansion, and $C(A, r)$ is a constant depending on A and r. Thus

$$
\begin{aligned}
\sum_{j=0}^{n-1} \mathbb{E}\left[b_j^2\right] \mathbb{E}\left[I_j^2\right] &\leq C(A, r)\, h^3 \sum_{j=0}^{n-1} e^{Ajh} \\
&= C(A, r)\, h^3 \frac{e^{AT} - 1}{e^{Ah} - 1} \\
&\leq C(A, r)\, T \frac{e^{AT} - 1}{A} h^2.
\end{aligned}
$$

Now we calculate the second term in (5.38). Note that $m > j$, I_m is independent of $b_m b_j I_j$, and b_m / b_{j+1} is independent of $b_{j+1} b_j I_j$, so

$$
\mathbb{E}\left[\sum_{m=0}^{n-1} \sum_{j=0}^{m-1} b_m I_m b_j I_j\right] = \sum_{m=0}^{n-1} \mathbb{E}\left[I_m\right] \sum_{j=0}^{m-1} \mathbb{E}\left[b_m / b_{j+1}\right] \mathbb{E}\left[b_{j+1} b_j I_j\right].
$$

Firstly we have

$$
\begin{aligned}
\mathbb{E}\left[I_m\right] &= \mathbb{E}\left[\int_{mh}^{(m+1)h} \left(\exp\left(X_t - X_{mh}\right) - 1\right) \mathrm{d}t\right] \\
&= \int_{mh}^{(m+1)h} \left(\mathbb{E}[\exp\left(X_t - X_{mh}\right)] - 1\right) \mathrm{d}t \\
&= \int_0^h (\exp(rt) - 1) \mathrm{d}t \\
&= r^{-1}\left(e^{rh} - 1\right) - h.
\end{aligned}
$$

Moreover,

$$
\mathbb{E}\left[b_m / b_{j+1}\right] = e^{r(m-j-1)h}.
$$

On the other hand, we have

$$
\begin{aligned}
\mathbb{E}\left[b_{j+1}b_jI_j\right] &= \mathbb{E}\left[\exp\left(2X_{jh}\right)\exp\left(X_{(j+1)h}-X_{jh}\right)\int_{jh}^{(j+1)h}\left(\exp\left(X_t-X_{jh}\right)-1\right)\mathrm{d}t\right]\\
&= \mathbb{E}\left[\exp\left(2X_{jh}\right)\right]\mathbb{E}\left[\exp\left(X_{(j+1)h}-X_{jh}\right)\int_{jh}^{(j+1)h}\left(\exp\left(X_t-X_{jh}\right)-1\right)\mathrm{d}t\right]\\
&= e^{Ajh}\ \mathbb{E}\left[\exp\left(X_h\right)\int_0^h\left(\exp\left(X_t\right)-1\right)\mathrm{d}t\right]\\
&= e^{Ajh}\int_0^h\left(\mathbb{E}\left[\exp\left(X_h-X_t+2X_t\right)\right]-\mathbb{E}\left[\exp\left(X_h\right)\right]\right)\mathrm{d}t\\
&= e^{Ajh}\int_0^h\left(\mathbb{E}\left[\exp\left(X_h-X_t\right)\exp\left(2X_t\right)\right]-e^{rh}\right)\mathrm{d}t\\
&= e^{Ajh}\int_0^h\left(e^{At}e^{r(h-t)}\ \mathrm{d}t-he^{rh}\right)\mathrm{d}t\\
&= e^{Ajh}e^{rh}\left[(A-r)^{-1}\left(e^{(A-r)h}-1\right)-h\right].
\end{aligned}
$$

Thus

$$
\sum_{m=0}^{n-1}\sum_{j=0}^{m-1}\mathbb{E}\left[b_m/b_{j+1}\right]\mathbb{E}\left[b_{j+1}b_jI_j\right]=e^{rh}\left[(A-r)^{-1}\left(e^{(A-r)h}-1\right)-h\right]\sum_{m=0}^{n-1}\sum_{j=0}^{m-1}e^{r(m-j-1)h}e^{Ajh}.
$$

$$
\begin{aligned}
\sum_{m=0}^{n-1}\sum_{j=0}^{m-1}e^{r(m-j-1)h}e^{Ajh} &= \sum_{m=0}^{n-1}e^{r(m-1)h}\frac{e^{(A-r)mh}-1}{e^{(A-r)h}-1}\\
&= \frac{e^{-rh}}{e^{(A-r)h}-1}\left(\frac{e^{AT}-1}{e^{Ah}-1}-\frac{e^{rT}-1}{e^{rh}-1}\right)\\
&\leq \frac{e^{-rh}}{e^{(A-r)h}-1}\frac{e^{AT}-1}{e^{Ah}-1}\\
&\leq e^{-rh}\frac{e^{AT}-1}{(A-r)\,A}T^{-2}n^2,
\end{aligned}
$$

where in the last step we use $e^x\geq 1+x$ for $x>0$.

Note that for small h,

$$
e^{rh}\left[(A-r)^{-1}\left(e^{(A-r)h}-1\right)-h\right]=\mathcal{O}\left(h^2\right).
$$

Hence

$$
\begin{aligned}
\sum_{m=0}^{n-1}\sum_{j=0}^{m-1}\mathbb{E}\left[b_m/b_{j+1}\right]\mathbb{E}\left[b_{j+1}b_jI_j\right] &\leq e^{-rh}\frac{e^{AT}-1}{(A-r)\,A}T^{-2}n^2e^{rh}\left[(A-r)^{-1}\left(e^{(A-r)h}-1\right)-h\right]\\
&= \mathcal{O}(1).
\end{aligned}
$$

Therefore

$$\begin{aligned}\mathbb{E}\left[\sum_{m=0}^{n-1}\sum_{j=0}^{m-1} b_m I_m b_j I_j\right] &= \sum_{m=0}^{n-1}\mathbb{E}\left[I_m\right]\sum_{j=0}^{m-1}\mathbb{E}\left[b_m/b_{j+1}\right]\mathbb{E}\left[b_{j+1}b_j I_j\right] \\ &= \left(r^{-1}\left(e^{rh}-1\right)-h\right)\sum_{m=0}^{n-1}\sum_{j=0}^{m-1}\mathbb{E}\left[b_m/b_{j+1}\right]\mathbb{E}\left[b_{j+1}b_j I_j\right] \\ &= \mathcal{O}(h^2).\end{aligned}$$

Finally we conclude

$$\mathbb{E}\left[\left(\overline{\widehat{S}}-\overline{S}\right)^2\right]=\mathcal{O}(h^2).$$

□

For ML variance, let $n = 2^{l-1}$ in the definition of $\overline{\widehat{S}}$. Note that the payoff is Lipschitz, so the variance convergence follows from

$$\begin{aligned}&\mathbb{E}\left[\left(\widehat{P}_l-\widehat{P}_{l-1}\right)^2\right] \\ \le\ & 2\,\mathbb{E}\left[\left(\widehat{P}_l-P\right)^2\right]+2\,\mathbb{E}\left[\left(\widehat{P}_{l-1}-P\right)^2\right] \\ \le\ & 2L_K^2\mathbb{E}\left[\left(\overline{\widehat{S}}_l-\overline{S}\right)^2\right]+2L_K^2\mathbb{E}\left[\left(\overline{\widehat{S}}_{l-1}-\overline{S}\right)^2\right].\end{aligned}$$

5.3.6 Barrier options

Let us first present the convergence result of the approximation of the crossing probability:

Proposition 5.3.17. *Let X_t be a scalar Lévy process with triple $(m,0,\nu)$ as in (5.1.2). Suppose the Lévy measure satisfies*

$$\begin{gathered}C_2\left|x\right|^{-1-\alpha}\le\nu(x)\le C_1\left|x\right|^{-1-\alpha},\ \text{as } |x|\le 1;\\ \begin{cases}\text{If } \alpha>0,\ \nu(x)\le C_3\left|x\right|^{-1-\alpha},\ \text{as } |x|\ge 1;\\ \text{If } \alpha=0,\ \nu(x)\le\exp\left(-C_3\left|x\right|\right),\ \text{as } |x|\ge 1\end{cases}\end{gathered}$$

where $C_1, C_2, C_3 > 0, 0 \le \alpha < 2$ are constants.

Suppose also we use $\widehat{M}_n = \max_{i=0,1,\dots,n}\exp\left(X_{\frac{i}{n}T}\right)$ as a discretely monitored approximation of $M_T = \max_{0\le t\le T} S(t)$. Provided there is a bounded density for M_T in the neighbourhood of B, we have

$$\mathbb{E}\left[\mathbf{1}_{\{\widehat{M}_n<B\}}-\mathbf{1}_{\{M_T<B\}}\right]=o\left(n^{-r}\right),$$

where $r\in\left(0,(1+2\alpha)^{-1}\right)$ is an arbitrary positive number.

Proof. Note that

$$\mathbf{1}_{\{\widehat{M}_n<B\}} - \mathbf{1}_{\{M_T<B\}} = 1 \text{ or } 0,$$

and "$= 1$ " holds only if either M_T is close to the barrier or the difference between the discretely and the continuously monitored maximum $V_n = M_T - \widehat{M}_n$ is large. More precisely,

$$\left\{\mathbf{1}_{\{\widehat{M}_n<B\}} - \mathbf{1}_{\{M_T<B\}} = 1\right\} \subset F \cup G,$$

where

$$\begin{aligned} F &:= \left\{B \leq M_T \leq B + n^{-r}\right\}, \text{ for an } r > 0 \text{ to be determined;} \\ G &:= \left\{V_n > n^{-r}\right\}. \end{aligned}$$

Hence the expected discrepancy between crossing probabilities admits the estimate:

$$\mathbb{E}\left[\mathbf{1}_{\{\widehat{M}_n<B\}} - \mathbf{1}_{\{M_T<B\}}\right] \leq \mathbb{P}(F) + \mathbb{P}(G).$$

We first investigate the contribution from F. Due to bounded density for M_T, we have

$$\mathbb{P}(F) = \mathcal{O}\left(n^{-r}\right).$$

According to (5.11) and (5.12),

$$V_n \leq \max_{i=1,\ldots,n} \left(\sup_{[0,\frac{1}{n}]} X_t^{\varepsilon}(i) - \left(X_{\frac{1}{n}}^{\varepsilon}(i)\right)^+ + S_n^{(i)} \right) + \frac{|\mu_\varepsilon|}{n}.$$

Denote

$$Z_i = \sup_{[0,\frac{1}{n}]} X_t^{\varepsilon}(i) - \left(X_{\frac{1}{n}}^{\varepsilon}(i)\right)^+.$$

We first eliminate the drift of V_n. Assuming for sufficiently large n,

$$|\mu_\varepsilon| \leq \frac{1}{2} n^{1-r}, \tag{5.39}$$

we have

$$\begin{aligned} \mathbb{P}\left(V_n > n^{-r}\right) &= \mathbb{P}\left(\max_{i=1,\ldots,n} \left(Z_i + S_n^{(i)}\right) > n^{-r} - \frac{|\mu_\varepsilon|}{n}\right) \\ &\leq \mathbb{P}\left(\max_{i=1,\ldots,n} \left(Z_i + S_n^{(i)}\right) > \frac{1}{2} n^{-r}\right). \end{aligned}$$

We can decompose it into two components:

$$\mathbb{P}\left(V_n > n^{-r}\right) \leq \mathbb{P}\left(\max_{i=1,\ldots,n} Z_i > \frac{1}{4} n^{-r}\right) + \mathbb{P}\left(\max_{i=1,\ldots,n} S_n^{(i)} > \frac{1}{4} n^{-r}\right).$$

Since

$$\max_{i=1,\ldots,n} Z(i) > \frac{1}{4} n^{-r}$$

implies that there are at least two jumps in some $Z(i)$, we have

$$\mathbb{P}\left(\max_{i=1,\ldots,n} Z(i) > \frac{1}{4} n^{-r}\right) \leq 1 - \exp(-\lambda_\varepsilon)\left(1 + \frac{\lambda_\varepsilon}{n}\right)^n$$

where $\exp(-\lambda_\varepsilon)\left(1 + \frac{\lambda_\varepsilon}{n}\right)^n$ is the probability that each $Z(i)$ has at most one jump.

Provided $\lambda_\varepsilon < n$,

$$\begin{aligned} & \exp(-\lambda_\varepsilon)\left(1 + \frac{\lambda_\varepsilon}{n}\right)^n \\ = & \exp\left(n \log\left(1 + \frac{\lambda_\varepsilon}{n}\right) - \lambda_\varepsilon\right) \\ = & \exp\left(-\frac{\lambda_\varepsilon^2}{2n} + o\left(\frac{\lambda_\varepsilon^3}{n^2}\right)\right). \end{aligned}$$

To have a meaningful bound such that $1 - \exp(-\lambda_\varepsilon)\left(1 + \frac{\lambda_\varepsilon}{n}\right)^n$ converges to zero as $n \to \infty$, we have to assume $\lambda_\varepsilon = o(\sqrt{n})$, under which we have

$$1 - \exp(-\lambda_\varepsilon)\left(1 + \frac{\lambda_\varepsilon}{n}\right)^n = \mathcal{O}\left(\frac{\lambda_\varepsilon^2}{n}\right).$$

Hence

$$\mathbb{P}\left(\max_{i=1,\ldots,n} Z(i) > \frac{1}{4} n^{-r}\right) \prec \frac{\lambda_\varepsilon^2}{n}.$$

Then we use Markov's inequality on the remainder terms: for $p > 0$,

$$\mathbb{P}\left(\max_{i=1,\ldots,n} S_n^{(i)} > \frac{1}{4} n^{-r}\right) \leq \frac{\mathbb{E}\left[\max_{i=1,\ldots,n}\left(S_n^{(i)}\right)^p\right]}{(n^{-r}/4)^p}.$$

According to Proposition 5.3.3, estimates (5.19) and (5.22), we have

$$\mathbb{E}\left[\max_{i=1,\ldots,n}\left(S_n^{(i)}\right)^p\right] \prec \varepsilon^{p-\delta}$$

and thus

$$\mathbb{P}\left(\max_{i=1,\ldots,n} S_n^{(i)} > \frac{1}{4} n^{-r}\right) \prec \varepsilon^{p-\delta} n^{rp}.$$

In summary, assuming $|\mu_\varepsilon| \leq \frac{1}{2} n^{1-r}$ and $\lambda_\varepsilon = o(\sqrt{n})$, we have

$$\mathbb{E}\left[\mathbf{1}_{\{\widehat{M}_n < B\}} - \mathbf{1}_{\{M_T < B\}}\right] \prec n^{-r} + \varepsilon^{p-\delta} n^{rp} + \frac{\lambda_\varepsilon^2}{n}.$$

Balancing the first two terms, we have

$$\varepsilon \sim n^{-r(1+p)(p-\delta)^{-1}}.$$

For $0 < \varepsilon < 1$,

$$\begin{aligned} \lambda_\varepsilon &= \int_{\varepsilon<|z|} \nu(\mathrm{d}z) \\ &\leq \begin{cases} 2C_1 \log \frac{1}{\varepsilon} + l_1, \ \alpha = 0; \\ l_2 \varepsilon^{-\alpha}, \ 0 < \alpha < 2. \end{cases} \end{aligned}$$

For $\alpha = 0$,

$$\begin{aligned} \lambda_\varepsilon &\prec \log \frac{1}{\varepsilon} \\ &\sim \log n. \end{aligned}$$

$\lambda_\varepsilon = o\left(\sqrt{n}\right)$ is satisfied. Then we have

$$\frac{\lambda_\varepsilon^2}{n} \prec \frac{(\log n)^2}{n}.$$

We can choose a $p > \delta$ and any $r < 1$, under which

$$\mathbb{E}\left[\mathbf{1}_{\{\widehat{M}_n < B\}} - \mathbf{1}_{\{M_T < B\}}\right] \prec n^{-r}.$$

For $0 < \alpha < 2$,

$$\begin{aligned} \lambda_\varepsilon &\prec \varepsilon^{-\alpha} \\ &\sim n^{r\frac{1+p}{p-\delta}\alpha}. \end{aligned}$$

In order that $\lambda_\varepsilon = o\left(\sqrt{n}\right)$, it is required that

$$r\frac{1+p}{p-\delta}\alpha < \frac{1}{2}.$$

We can choose a large p such that the constraint becomes

$$r < \frac{1}{2\alpha}.$$

Then

$$\begin{aligned} \frac{\lambda_\varepsilon^2}{n} &\prec \varepsilon^{-2\alpha} n^{-1} \\ &\sim n^{-1+r\frac{1+p}{p-\delta}2\alpha}. \end{aligned}$$

Balancing n^{-r} and $n^{-1+r\frac{1+p}{p-\delta}2\alpha}$, we have

$$r = \left(1 + \frac{1+p}{p-\delta} 2\alpha\right)^{-1}.$$

Since we can choose an arbitrary $p > \delta$, we have

$$r \in \left(0, (1+2\alpha)^{-1}\right).$$

Hence

$$\mathbb{E}\left[\mathbf{1}_{\{\widehat{M}_n < B\}} - \mathbf{1}_{\{M_T < B\}}\right] \prec n^{-r}$$

where $r \in \left(0, (1+2\alpha)^{-1}\right)$.

Now we examine assumption (5.39) for different processes. By (5.26):

$$|\mu_\varepsilon| \le \begin{cases} |m| + \left|\frac{C_1 - C_2}{\alpha - 1}\left[\varepsilon^{1-\alpha} - 1\right]\right|, & \alpha \neq 1; \\ |m| + |C_1 - C_2| \log \frac{1}{\varepsilon}, & \alpha = 1. \end{cases}$$

If $0 < \alpha < 1$, we have $|\mu_\varepsilon| \le |m| + \left|2\frac{C_1 - C_2}{\alpha - 1}\right|$. The assumption (5.39) is satisfied.

If $\alpha = 1$,

$$\begin{aligned} |\mu_\varepsilon| &\prec \log \frac{1}{\varepsilon} \\ &\sim \log n. \end{aligned}$$

Given an r, for sufficiently large n, $|\mu_\varepsilon| < \frac{1}{2} n^{1-r}$.

If $1 < \alpha < 2$, we have

$$\begin{aligned} |\mu_\varepsilon| &\prec \varepsilon^{1-\alpha} \\ &\sim n^{r\frac{1+p}{p-\delta}(\alpha-1)}. \end{aligned}$$

In order that $|\mu_\varepsilon| \le \frac{1}{2} n^{1-r}$ for sufficiently large n,

$$r \frac{1+p}{p-\delta}(\alpha - 1) \le 1 - r.$$

Since we can choose an arbitrary $p > \delta$, we must have

$$r < \frac{1}{\alpha}.$$

Thus in all cases (5.39) can be satisfied by choosing sufficiently large p and sufficiently small δ.

In particular, for VG processes, due to its finite variation, $|\mu_\varepsilon| < \infty$. For NIG processes, according to the expansion of its Lévy measure near the origin

$$K_1(x) \sim \frac{1}{x} + \mathcal{O}(1),$$

we have

$$\begin{aligned}
|\mu_\varepsilon| &\leq |m| + \left|\int_{\varepsilon<|x|<1} x\ \nu(\mathrm{d}x)\right| \\
&\leq |m| + \int_{\varepsilon<x<1} x\frac{C}{|x|}e^{Ax}\left(\frac{1}{B|x|} + K_1\right)\mathrm{d}x + \int_{-1<x<-\varepsilon} x\frac{C}{|x|}e^{Ax}\left(\frac{1}{B|x|} + K_1\right)\mathrm{d}x \\
&\leq |m| + e^A \int_{\varepsilon<|x|<1}\left(\frac{C}{Bx}\right)\mathrm{d}x + e^A K_1(1-\varepsilon) - K_1(1-\varepsilon) \\
&\leq C_2^{NIG},
\end{aligned}$$

where C_2^{NIG} is a constant depending on the NIG parameters. □

We have seen in Proposition 5.3.11 that our estimate is not as fine as Theorem 5.3.1 when $\alpha > \frac{1}{2}$. Based on this fact, we use Theorem 5.3.1 to give a better result when $\alpha > \frac{1}{2}$.

Proposition 5.3.18. *Suppose X_t is a scalar Lévy process with triple $(m, 0, \nu)$, with finite first moment, i.e.*

$$\int_{|x|>1} |x|\ \nu(\mathrm{d}x) < \infty.$$

Suppose also we use $\widehat{M}_n$ (5.41) as a discretely monitored approximation of M_T. Provided there is a bounded density for M_T in the neighbourhood of B, we have

1. *If X_t is of finite variation, i.e. $\int_{|x|<1} |x|\,\nu(\mathrm{d}x) < \infty$, then*

$$\mathbb{E}\left[\mathbf{1}_{\{\widehat{M}_n<B\}} - \mathbf{1}_{\{M_T<B\}}\right] = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right);$$

2. *If $\sigma = 0$ and X_t is of infinite variation, then*

$$\mathbb{E}\left[\mathbf{1}_{\{\widehat{M}_n<B\}} - \mathbf{1}_{\{M_T<B\}}\right] = o\left(\frac{1}{n^r}\right),$$

   where $r \in \left(0, \frac{1}{2\beta}\right)$ is an arbitrary positive number;

$$\beta = \inf\left\{\alpha > 0 : \int_{|x|<1} |x|^\alpha\,\nu(\mathrm{d}x) < \infty\right\}$$

   is the Blumenthal-Getoor index of X_t.

Proof. Repeating the start of the proof for Proposition 5.3.17, we have for an $r > 0$ to be determined

$$\mathbb{E}\left[\mathbf{1}_{\{\widehat{M}_n < B\}} - \mathbf{1}_{\{M_T < B\}}\right] \prec n^{-r} + \mathbb{P}\left(V_n > n^{-r}\right).$$

By Markov's inequality,

$$\mathbb{P}\left(V_n > n^{-r}\right) \leq \mathbb{E}\left[V_n\right] n^r.$$

According to Theorem 5.3.1, balancing n^{-r} and $\mathbb{E}\left[V_n\right] n^r$ we have the results. □

This proposition will be used later in the model driven by the NIG process.

We have a strong result for spectrally negative α-stable processes.

Proposition 5.3.19. *Let X_t be a spectrally negative α-stable process with $\alpha > 1$, i.e. its Lévy measure $\nu(x)$ is*

$$\nu(x) = \frac{B}{|x|^{\alpha+1}} 1_{\{x<0\}}$$

where $B > 0,\ 0 < \alpha < 2$ are constants. Suppose also we use $\widehat{M}_n$ (5.41) as a discretely monitored approximation of M_T. Provided there is a bounded density for M_T in the neighbourhood of B, we have

$$\mathbb{E}\left[\mathbf{1}_{\{\widehat{M}_n < B\}} - \mathbf{1}_{\{M_T < B\}}\right] = o\left(n^{-\frac{1}{\alpha}+\delta}\right)$$

for any small $\delta > 0$.

Proof. Repeating the start of the proof for Proposition 5.3.17, we have for an $r > 0$ to be determined

$$\mathbb{E}\left[\mathbf{1}_{\{\widehat{M}_n < B\}} - \mathbf{1}_{\{M_T < B\}}\right] \prec n^{-r} + \mathbb{P}\left(V_n > n^{-r}\right).$$

By Markov's inequality,

$$\mathbb{P}\left(V_n > n^{-r}\right) \leq \mathbb{E}\left[V_n^p\right] n^{rp}.$$

According to Proposition 5.3.12, balancing n^{-r} and $n^{-\left(\frac{1}{\alpha}-\theta\right)p} n^{rp}$ we have

$$r = \frac{p}{p+1}\left(\frac{1}{\alpha} - \theta\right).$$

Taking sufficiently large p and sufficiently small θ we conclude. □

We now prove the variance convergence and weak convergence of the estimator.

We consider a up-and-out call barrier option for which the discounted payoff is

$$P = \exp(-rT)\,(S_T - K)^+\,\mathbf{1}_{\{M_T<B\}}, \tag{5.40}$$

where $M_T = \max_{0\le t\le T} S(t)$.

The discretely monitored approximation $\widehat{M_n} = \max_{i=0,1,\dots,n} \exp\left(X_{\frac{i}{n}T}\right)$ gives

$$\widehat{P} = \exp(-rT)\,(S_T - K)^+\,\mathbf{1}_{\left\{\widehat{M_n}<B\right\}}. \tag{5.41}$$

Proposition 5.3.20. *Suppose the asset price follows an exponential Lévy model, i.e.* $S_t = S_0 \exp(X_t)$. *Suppose also we use* $\widehat{M_n} = \max_{i=0,1,\dots,n} \exp\left(X_{\frac{i}{n}T}\right)$ *as a discretely monitored approximation of* M_T. *We assume there is a bounded density for* M_T *in the neighbourhood of* B.

Let $n = 2^l$ *and* $h_l = T/2^l$. $\widehat{P}_l$ *is given by* $\widehat{P}$ *where* $n = 2^l$.

For an up-and-out Barrier option (5.40),

- *If* X_t *is a Variance Gamma (VG) process, then*

$$\begin{aligned} \mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^2\right] &= o(h_l^{1-\delta}); \\ \left|\mathbb{E}\left[\widehat{P} - P\right]\right| &= o(h_l^{1-\delta}) \end{aligned}$$

  where δ *is an arbitrary positive number.*

- *If* X_t *is a NIG process, then*

$$\begin{aligned} \mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^2\right] &= o(h_l^{1/2-\delta}); \\ \left|\mathbb{E}\left[\widehat{P} - P\right]\right| &= o(h_l^{1/2-\delta}) \end{aligned}$$

  where δ *is an arbitrary positive number.*

- *If* X_t *is a spectrally negative* α*-stable process with* $\alpha > 1$, *then*

$$\begin{aligned} \mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^2\right] &= o\left(h_l^{\frac{1}{\alpha}-\delta}\right); \\ \left|\mathbb{E}\left[\widehat{P} - P\right]\right| &= o\left(h_l^{\frac{1}{\alpha}-\delta}\right) \end{aligned}$$

where δ is an arbitrary positive number.

For an up-and-in Barrier option

$$P^{(in)} = \exp(-rT)\,(S_T - K)^+\,\mathbf{1}_{\{M_T>B\}},$$

the approximated payoff is

$$\widehat{P}^{(in)} = \exp(-rT)\,(S_T - K)^+\,\mathbf{1}_{\{\widehat{M}_n>B\}}.$$

Let $q>1$ be any constant such that

$$\int_{|z|>1} e^{2qz}\nu(\mathrm{d}z) < \infty.$$

The relationship between the convergence of up-and-in and up-and-out Barrier options is

$$\begin{aligned} \mathbb{E}\left[\left(\widehat{P}^{(in)} - P^{(in)}\right)^2\right] &= \mathbb{E}\left[\left(\widehat{P} - P\right)^2\right]^{1-\frac{1}{q}}; \\ \left|\mathbb{E}\left[\widehat{P}^{(in)} - P^{(in)}\right]\right| &= \left|\mathbb{E}\left[\widehat{P} - P\right]\right|^{1-\frac{1}{2q}}. \end{aligned}$$

Proof. Let $f(S) = (S-K)^+$.

For an up-and-out Barrier option, since the payoff is bounded,

$$\begin{aligned} \mathbb{E}\left[\left(\widehat{P} - P\right)^2\right] &\leq f(B)^2\,\mathbb{E}\left[\mathbf{1}_{\{\widehat{M}_n<B\}} - \mathbf{1}_{\{M_T<B\}}\right]; \\ \left|\mathbb{E}\left[\widehat{P} - P\right]\right| &\leq f(B)\,\mathbb{E}\left[\mathbf{1}_{\{\widehat{M}_n<B\}} - \mathbf{1}_{\{M_T<B\}}\right]. \end{aligned}$$

For an up-and-in Barrier option, first we separate the contribution from the European call payoff. We have

$$\begin{aligned} \mathbb{E}\left[\left(\widehat{P} - P\right)^2\right] &= \mathbb{E}\left[f(S_T)^2\left(\mathbf{1}_{\{\widehat{M}_n<B\}} - \mathbf{1}_{\{M_T<B\}}\right)\right] \\ &\leq \mathbb{E}\left[f(S_T)^{2q}\right]^{\frac{1}{q}}\,\mathbb{E}\left[\mathbf{1}_{\{\widehat{M}_n<B\}} - \mathbf{1}_{\{M_T<B\}}\right]^{1-\frac{1}{q}}. \end{aligned} \tag{5.42}$$

Similarly

$$\left|\mathbb{E}\left[\widehat{P} - P\right]\right| \leq \mathbb{E}\left[f(S_T)^q\right]^{\frac{1}{q}}\,\mathbb{E}\left[\mathbf{1}_{\{\widehat{M}_n<B\}} - \mathbf{1}_{\{M_T<B\}}\right]^{1-\frac{1}{q}}.$$

Note that the European call payoff is Lipschitz,

$$f(S_T) \leq |S_T - S_0| + f(S_0).$$

Thus the $2q$th moment of the European call payoff is finite if and only if the Laplace transform of the Lévy measure is finite at $2q$:

$$\mathbb{E}\left[f\left(S_T\right)^{2q}\right] < \infty \Longleftrightarrow \int e^{2qx}\nu(\mathrm{d}x) < \infty.$$

The variance of the multilevel correction term is bounded by

$$\mathbb{E}\left[\left(\widehat{P}_l - \widehat{P}_{l-1}\right)^2\right] \leq 2\,\mathbb{E}\left[\left(\widehat{P}_l - P\right)^2\right] + 2\,\mathbb{E}\left[\left(\widehat{P}_{l-1} - P\right)^2\right].$$

The results for VG, NIG and spectrally negative $\alpha-$stable follow from $\alpha = 0$ case of Proposition 5.3.17, case 2 of Proposition 5.3.18 and Proposition 5.3.19 respectively.

□

The gap between numerical results and the analysis for the NIG process could be since the bound of Theorem 5.3.5 on the contribution from the large jumps is not tight enough.

Chapter 6

Conclusions

6.1 Summary of results

This work concerns applications of multilevel Monte Carlo(MLMC) to option pricing problems and computing Greeks in scalar jump-diffusion models, and option pricing problems in exponential Lévy models. In the first part of the thesis, for constant rate and state-dependent rate jump-diffusion models, we numerically demonstrate the computational efficiency of MLMC for Lipschitz, Asian, lookback, barrier and digital options. We also prove convergence of the variance of the ML correction term that is a key to the computational complexity of MLMC. Due to Theorem 1.3.1, the computational efficiency is analytically justified. Those numerical and analytic results are summarised in Table 6.1. We then studied computing Greeks in scalar jump-diffusion models. The numerical results of the variance convergence order for Greeks of European call option are summarised in Table 6.2.

	Euler for diffusions		Milstein for diffusions		Jump-adapted Milstein	
option	numerical	analysis	numerical	analysis	numerical	analysis
Lipschitz	$\mathcal{O}(h)$	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h^2)$
Asian	$\mathcal{O}(h)$	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h^2)$
lookback	$\mathcal{O}(h)$	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$	$o(h^{2-\delta})$	$\mathcal{O}(h^2)$	$o(h^{2-\delta})$
barrier	$\mathcal{O}(h^{1/2})$	$o(h^{1/2-\delta})$	$\mathcal{O}(h^{3/2})$	$o(h^{3/2-\delta})$	$\mathcal{O}(h^{3/2})$	$o(h^{1-\delta})$
digital	$\mathcal{O}(h^{1/2})$	$\mathcal{O}(h^{1/2}\log h)$	$\mathcal{O}(h^{3/2})$	$o(h^{3/2-\delta})$	$\mathcal{O}(h^{3/2})$	$o(h^{3/2-\delta})$

Table 6.1: Orders of convergence for V_l as observed numerically and proved analytically for both the Euler and Milstein discretisations for pure diffusions and jump-diffusions; δ can be any strictly positive constant. $h \equiv h_l$.

Greeks	Jump-adapted Milstein weak	Jump-adapted Milstein variance
delta	$\mathcal{O}(h)$	$\mathcal{O}(h)$
vega	$\mathcal{O}(h)$	$\mathcal{O}(h)$
rate via pathwise	$\mathcal{O}(h)$	$\mathcal{O}(h^{\frac{3}{2}})$
rate via change of measure	$\mathcal{O}(h)$	$\mathcal{O}(h^{1.75})$

Table 6.2: Orders of convergence for V_l as observed numerically for Greeks of European call option. $h \equiv h_l$.

In the second part of the thesis, we are concerned with option pricing problems in exponential Lévy models. Numerical results and analysis of convergence rates are summarised in Table 6.3. These results imply the computational complexity of MLMC in each individual case.

In more detail, in Chapter 2 we have extended the multilevel Monte Carlo method to scalar jump-diffusion SDEs using a jump-adapted Milstein discretisation. Second order variance convergence is obtained in the constant rate case for European options with Lipschitz payoffs, and also for lookback options by constructing estimators using a Brownian interpolation which approximates the process within a timestep as a Gaussian process with constant drift and volatility. A variance convergence rate of $\frac{3}{2}$ is obtained for barrier and digital options, which again is the same as the one previously achieved for scalar diffusion SDEs. In the state-dependent rate case, we use a thinning with a change of measure to simulate the process.

In Chapter 3 we verify the the conditions required by the MLMC computational complexity Theorem 1.3.1 for finite-rate jump-diffusion SDEs, using a jump-adapted Milstein discretisation scheme for the constant rate case and a thinning procedure for the state-dependent intensity cases. Therefore we justify the numerical results of MLMC computational complexity presented in Chapter 2.

In Chapter 4 after developing pathwise and likelihood ratio estimators for sensitivity to jump rate, we apply MLMC to calculating Greeks in the constant jump rate case. We also discuss the variable rate case and variance reduction in the presence of jumps. Numerical results for delta, vega and sensitivity to the jump rate show that the estimators we developed are quite efficient.

In Chapter 5 we first demonstrate numerical results of multilevel Monte Carlo in exponential Lévy models, using Variance Gamma, Normal Inverse Gaussian and α-stable processes. Then we give upper bounds on L^p convergence rate of discretely

	VG			
	numerical		analysis	
option	weak	var	weak	var
Asian	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$
lookback	$\mathcal{O}(h)$	$\mathcal{O}(h^{1.5})$	$\mathcal{O}(h\,\lvert\log h\rvert)$	$\mathcal{O}(h)$
barrier	$\mathcal{O}(h^{0.7})$	$\mathcal{O}(h)$	$o\left(h^{1-\delta}\right)$	$o\left(h^{1-\delta}\right)$

	NIG			
	numerical		analysis	
option	weak	var	weak	var
Asian	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$
lookback	$\mathcal{O}(h^{0.7})$	$\mathcal{O}(h^{1.5})$	$o\left(h^{1-\delta}\right)$	$\mathcal{O}(h\,\lvert\log h\rvert)$
barrier	$\mathcal{O}(h^{0.7})$	$\mathcal{O}(h^{0.9})$	$o\left(h^{0.5-\delta}\right)$	$o\left(h^{0.5-\delta}\right)$

	spectrally negative α-stable with $\alpha > 1$			
	numerical for $\alpha = 1.5597$		analysis	
option	weak	var	weak	var
Asian	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$
lookback	$\mathcal{O}(h^{0.5})$	$\mathcal{O}(h^{1.5})$	$o\left(h^{1/\alpha-\delta}\right)$	$o\left(h^{2/\alpha-\delta}\right)$
barrier	$\mathcal{O}(h^{0.5})$	$\mathcal{O}(h^{0.7})$	$o\left(h^{1/\alpha-\delta}\right)$	$o\left(h^{1/\alpha-\delta}\right)$

Table 6.3: Convergence rates of V_l for VG, NIG and α-stable processes; δ can be any small positive constant. $h \equiv h_l$.

monitored error introduced to the running maximum for a broad class of Lévy processes whose Lévy measures admit a power law near the origin and have exponential tails. In particular we give upper bounds on variance convergence rates for MLMC estimators in models using Variance Gamma, Normal Inverse Gaussian and α-stable processes. As an application, the numerical analysis of multilevel Monte Carlo for exponential Lévy models is presented.

6.2 Future work

We name a few potential areas for future work:

It is possible to investigate whether the multilevel quasi-Monte Carlo method [38] will further reduce the cost in all cases.

In Chapter 3, the gap to bound the variance of conditional crossing probabilities for barrier options is still open to fill. Due to the difficulty in simulating Lévy area in the multidimensional case, we only deal with the scalar case. By using the antithetic method in [30], a multidimensional version of the jump-adapted Milstein scheme could be proposed to simulate constant jump rate processes without simulating Lévy area.

In Chapter 4, the issue of pathwise method for non-smooth payoffs can be coped with by smoothing the payoff, including explicit expression of conditional expectation (p399 of [39]) in the 1D case, "splitting" technique [2] in multidimensional case, or combination with Likelihood ratio method, called Vibrato Monte Carlo [33]. It is also interesting to study the effectiveness of MLMC in combination with other variance reduction techniques [39].

In Chapter 5, there is still a gap in the analysis of barrier option since we cannot match the estimated variance convergence rate to the numerical result. Since the weak convergence rate is less than one half in the proof for NIG and α-stable process, the computational complexity is not guaranteed. Furthermore, we only consider the Lévy process of which the increments can be directly simulated. For other Lévy processes, it may be possible to simulate the increments by constructing approximations to the inverse of the cumulative distribution function. Also, simulation technique might be developed through approximating charateristic functions of Lévy processes. In both cases, multilevel estimators can be constructed accordingly.

It is natural to develop numerical schemes for SPDEs driven by finite-rate state-dependent jump-diffusion processes or Lévy processes based on this work. Correspondingly, MLMC method can be applied to this class of problems.

Appendix A

Appendix

A.1 Brownian bridge results

The results in can be easily extended to the case of nonuniform timestep and jump-adapted schemes since the jump effect is separated by the grid point. For constant drift a and volatility b , the SDE (1.1) has solution

$$S(t) = S_0 + a\,t + b\,W(t)$$

We can use this as an approximation for the non-constant drift and volatility SDE within a small timestep $[t_n, t_{n+1}]$ of length h_n

$$S(t) = S_n + v_n(S_{n+1}-S_n) + b\ (W(t) - W_n - v_n\,(W_{n+1}-W_n)) \tag{A.1}$$

where $v_n \equiv (t-t_n)/h_n$. This means that the deviation of $S(t)$ from a piecewise linear interpolation of the values $S_n \equiv S(t_n)$ is proportional to the deviation of $W(t)$ from its piecewise linear interpolation. It can be proved that the distribution of the latter is independent of the Brownian increment $W_{n+1}-W_n$, and furthermore we have the following results (see for example [39]).

Lemma A.1.1. *Conditional on S_n and S_{n+1}, the distribution for the integral of $S(t)$ over the interval $[t_n, t_{n+1}]$ is given by*

$$\int_{t_n}^{t_{n+1}} S(t)\,\mathrm{d}t = \ \tfrac{1}{2}h\,(S_n + S_{n+1}) + b_n\,I_n \tag{A.2}$$

where

$$I_n \equiv \int_{t_n}^{t_{n+1}} (W(t) - W_n - v_n\,(W_{n+1}-W_n))\ \mathrm{d}t$$

is a $N(0,\ \frac{1}{12}h_n^3)$ Normal random variable, independent of $W_{n+1}-W_n$.

Lemma A.1.2. *Conditional on S_n and S_{n+1}, the distributions for the minimum and maximum of $S(t)$ over the interval $[t_n, t_{n+1}]$ are given by*

$$\begin{aligned} S_{n,min} &= \tfrac{1}{2}\left(S_n + S_{n+1} - \sqrt{(S_{n+1}-S_n)^2 - 2\,b^2\,h_n \log U_n}\ \right), \\ S_{n,max} &= \tfrac{1}{2}\left(S_n + S_{n+1} + \sqrt{(S_{n+1}-S_n)^2 - 2\,b^2\,h_n \log V_n}\ \right), \end{aligned}$$

where U_n and V_n are each uniformly distributed on the unit interval $[0,1]$.

Lemma A.1.3. *Conditional on S_n and S_{n+1}, the probability that the minimum (or maximum) of $S(t)$ over the interval $[t_n, t_{n+1}]$ is less than (or greater than) some value B, is*

$$\mathbb{P}\left(\inf_{[t_n,t_{n+1}]} S(t) < B \mid S_n, S_{n+1} \right) = \exp\left(\frac{-2\,(S_n-B)^+(S_{n+1}-B)^+}{b^2\,h_n} \right), \quad \text{(A.3)}$$

$$\mathbb{P}\left(\sup_{[t_n,t_{n+1}]} S(t) > B \mid S_n, S_{n+1} \right) = \exp\left(\frac{-2\,(B-S_n)^+(B-S_{n+1})^+}{b^2\,h_n} \right), \quad \text{(A.4)}$$

where the notation $(x)^+$ means $\max(x,0)$.

Corollary A.1.4. *If $W(t)$ is a Brownian motion with $W(0)=W(1)=0$, then for $x>0$*

$$\mathbb{P}\left(\sup_{[0,1]} W(t) > x \right) = \mathbb{P}\left(\inf_{[0,1]} W(t) < -x \right) = \exp(-2x^2),$$

and hence $\mathbb{E}\left[\sup_{[0,1]} |W(t)|^m\right]$ is finite for all integers $m>0$.

A.2 Assorted inequalities and applications

Here we list a number of classic inequalities with example applications, all of which are used in the main analysis.

Theorem A.2.1 (Jensen's inequality)**.** *If x_i are real variables, X is a scalar random variable and $\phi(x)$ is a scalar convex function, then*

$$\phi\left(n^{-1} \sum_{i=1}^{n} x_i \right) \leq n^{-1} \sum_{i=1}^{n} \phi(x_i)$$

and

$$\phi\left(\mathbb{E}[X]\right) \leq \mathbb{E}\left[\phi(X)\right].$$

Integral form: if μ is a probability measure on $(E, \mathcal{B}(E))$ where $E \subset R^d$, i.e. $\mu(E) = 1$, and f and $\phi(f)$ are integrable then

$$\phi\left(\int f \mathrm{d}\mu\right) \leq \int \phi(f) \mathrm{d}\mu.$$

Example A.2.2. *$\phi(x) = x^m$ is convex for positive integers m, and hence*

$$\left(\sum_{i=1}^{n} x_i\right)^m \leq n^{m-1} \left(\sum_{i=1}^{n} x_i^m\right) \tag{A.5}$$

Furthermore, if X_i are scalar random variables, then

$$\mathbb{E}\left[\,\left|\sum_{i=1}^{n} X_i\right|^m\right] \leq n^{m-1} \left(\sum_{i=1}^{n} \mathbb{E}[\,|X_i|^m]\right) \tag{A.6}$$

Theorem A.2.3 (Minkowski's inequality)**.** *If X, Y are scalar random variables and p is a real constant with $1 \leq p < \infty$, then*

$$\left(\mathbb{E}[\,|X+Y|^p]\right)^{1/p} \leq \left(\mathbb{E}[\,|X|^p]\right)^{1/p} + \left(\mathbb{E}[\,|Y|^p]\right)^{1/p}.$$

Example A.2.4. *Putting $p=2$ and applying the inequality to $X-\mathbb{E}[X]$ and $Y-\mathbb{E}[Y]$ gives*

$$\sqrt{\mathbb{V}[X+Y]} \leq \sqrt{\mathbb{V}[X]} + \sqrt{\mathbb{V}[Y]}. \tag{A.7}$$

Theorem A.2.5 (Hölder's inequality)**.** *If X, Y are scalar random variables, and p, q are real constants with $1 \leq p, q < \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$, then*

$$\mathbb{E}[\,|X\,Y|\,] \;\leq\; \left(\mathbb{E}[\,|X|^p]\right)^{1/p} \; \left(\mathbb{E}[\,|Y|^q]\right)^{1/q}.$$

Example A.2.6. *Taking $p=q=2$ gives Cauchy-Schwarz inequality*

$$\mathbb{E}[\,X\,Y\,] \;\leq\; \sqrt{\mathbb{E}[\,X^2]\;\mathbb{E}[\,Y^2]}\,. \tag{A.8}$$

Theorem A.2.7 (Markov's inequality)**.** *If X is a scalar random variable and $a>0$,*

$$\mathbb{P}\left(|X| \geq a\right) \;\leq\; \frac{\mathbb{E}[\,|X|\,]}{a}.$$

Corollary A.2.8. *For any positive integer m and $a>0$*

$$\mathbb{P}\left(|X| \geq a\right) \;=\; \mathbb{P}\left(|X|^m \geq a^m\right) \;\leq\; \frac{\mathbb{E}[\,|X|^m]}{a^m}.$$

In [52] we have:

Theorem A.2.9 (Burkholder–Davis–Gundy inequality)**.** *Let M be a martingale with cádlág paths. For any real number $m \geq 1$, there exists constant C_m such that*

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |M_t|^m\right] \leq C_m \mathbb{E}\left[\left|[M]_T\right|^{m/2}\right],$$

where $[M]_t$ is the quadratic variation of M_t.

In [19] we have:

Theorem A.2.10 (Doob's inequality)**.** *Let M be a right-continuous martingale. For any real number $m > 1$,*

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |M_t|^m\right]^{1/m} \leq \frac{m}{m-1}\mathbb{E}\left[|M_T|^m\right]^{1/m}.$$

Bibliography

[1] Y. Aït-Sahalia and J. Jacod. Estimating the degree of activity of jumps in high frequency data. *The Annals of Statistics*, 37(5A):2202–2244, 2009.

[2] A. Asmussen and P. Glynn. *Stochastic Simulation.* Springer, New York, 2007.

[3] S. Asmussen, P. Glynn, and J. Pitman. Discretization error in simulation of one-dimensional reflecting Brownian motion. *The Annals of Applied Probability*, 5(4):875–896, 1995.

[4] R. Avikainen. Convergence rates for approximations of functionals of SDEs. *Finance and Stochastics*, 13(3):381–401, 2009.

[5] M. Becker. Comment on 'correcting for simulation bias in monte carlo methods to value exotic options in models driven by lévy processes' by c. ribeiro and n. webber. *Applied Mathematical Finance*, 17(2):133–146, 2010.

[6] F. Black and M. Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3):637–654, 1973.

[7] S. Burgos and M. B. Giles. The computation of greeks with multilevel monte carlo. *Arxiv preprint arXiv:1102.1348*, 2011.

[8] Peter Carr and Liuren Wu. The Finite Moment Log Stable Process and Option Pricing. *The Journal of Finance*, 58(2):753–778, 2003.

[9] John M Chambers, Colin L Mallows, and BW Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–344, 1976.

[10] A. Chen. *Sampling error of the supremum of a Lévy process.* PhD thesis, University of Illinois at Urbana-Champaign, 2011.

[11] A. Chen, L. Feng, and R. Song. On the monitoring error of the supremum of a normal jump diffusion process. *Journal of Applied Probability*, 48(4):1021–1034, 2011.

[12] K. A. Cliffe, M. B. Giles, R. Scheichl, and A. L. Teckentrup. Multilevel monte carlo methods and applications to elliptic pdes with random coefficients. *Computing and Visualization in Science*, 14(1):3–15, 2011.

[13] R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1:223–236, 2001.

[14] R. Cont and P. Tankov. *Financial modelling with jump processes.* Chapman & Hall/CRC, London, 2004.

[15] M.H.A. Davis and M.P. Johansson. Malliavin monte carlo greeks for jump diffusions. *Stochastic Processes and their Applications*, 116(1):101–129, 2006.

[16] S. Dereich. Multilevel Monte Carlo Algorithms for Lévy-driven SDEs with Gaussian correction. *The Annals of Applied Probability*, 21(1):283–311, 2011.

[17] S. Dereich and F. Heidenreich. A multilevel Monte Carlo algorithm for Lévy-driven stochastic differential equations. *Stochastic Processes and their Applications*, 121(7):1565–1587, 2011.

[18] E.H.A. Dia and D. Lamberton. Connecting discrete and continuous lookback or hindsight options in exponential Lévy models. *Advances in Applied Probability*, 43(4):1136–1165, 2011.

[19] J. L. Doob. *Stochastic processes.* Wiley New York, 1953.

[20] Rick Durrett. *Probability: Theory and Examples, 3rd Edtion.* Thomson, 2005.

[21] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events: for Insurance and Finance.* Springer, 2008.

[22] Albert Ferreiro-Castilla, Andreas E Kyprianou, Robert Scheichl, and Gowri Suryanarayana. Multilevel Monte Carlo simulation for Lévy processes based on the Wiener–Hopf factorisation. *Stochastic Processes and their Applications*, 124(2):985–1010, 2014.

[23] J.E. Figueroa-López and P. Tankov. Small-time asymptotics of stopped Lévy bridges and simulation schemes with controlled bias. *Bernoulli,*, Forthcoming, 2013.

[24] Peter Friz and Martin Keller-Ressel. Moment explosions in stochastic volatility models. In Rama Cont, editor, *Encyclopedia of Quantitative Finance*. Wiley Online Library, 2010.

[25] T. Gerstner and S. Heinz. Dimension-and time-adaptive multilevel Monte Carlo methods. *Sparse Grids and Applications*, pages 107–120, 2012.

[26] Kay Giesecke and Dmitry Smelov. Exact sampling of jump diffusions. *Operations Research*, 61(4):894–907, 2013.

[27] M. B. Giles. Multilevel monte carlo for basket options. In *Simulation Conference (WSC), Proceedings of the 2009 Winter*, pages 1283–1290. IEEE, 2009.

[28] M. B. Giles. Multilevel Monte Carlo methods. In *Monte Carlo and Quasi-Monte Carlo Methods 2012*. Springer-Verlag, 2012.

[29] M. B. Giles and Christoph Reisinger. Stochastic finite differences and multi-level monte carlo for a class of spdes in finance. *SIAM Journal on Financial Mathematics*, 3(1):572–592, 2012.

[30] M. B. Giles and L. Szpruch. Antithetic multilevel monte carlo estimation for multi-dimensional sdes without lévy area simulation. *Annals of Applied Probability*, (to appear), 2013.

[31] M.B. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In A. Keller, S. Heinrich, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 343–358. Springer-Verlag, Berlin Heidelberg New York, 2007.

[32] M.B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[33] M.B. Giles. Vibrato Monte Carlo sensitivities. In P. L'Ecuyer and A. Owen, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2008*, pages 369–382. Springer, 2009.

[34] M.B. Giles, K. Debrabant, and A. Rößler. Numerical analysis of multilevel Monte Carlo path simulation using the Milstein discretisation. *ArXiv preprint arXiv:1302.4676*, 2013.

[35] M.B. Giles and P. Glasserman. Smoking adjoints: fast Monte Carlo Greeks. *RISK*, January 2006.

[36] M.B. Giles, D. Higham, and X. Mao. Analysing multilevel Monte Carlo for options with non-globally Lipschitz payoff. *Finance and Stochastics*, 13(3):403–413, 2009.

[37] M.B. Giles and L. Szpruch. Multilevel monte carlo methods for applications in finance. In *Recent Developments in Computational Finance*, chapter 1, pages 3–47. World Scientific, 2013.

[38] M.B. Giles and B.J. Waterhouse. Multilevel quasi-Monte Carlo path simulation. *Advanced Financial Modelling*, page 165, 2009.

[39] P. Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, New York, 2004.

[40] P. Glasserman and N. Merener. Numerical solution of jump-diffusion LIBOR market models. *Finance and Stochastics*, 7(1):1–27, 2003.

[41] P. Glasserman and N. Merener. Convergence of a discretization scheme for jump-diffusion processes with state-dependent intensities. *Proc. Royal Soc. London A*, 460:111–127, 2004.

[42] J. Jacod and A.N. Shiryaev. *Limit theorems for stochastic processes*, volume 288. Springer-Verlag Berlin, 1987.

[43] Jean Jacod, Thomas G Kurtz, Sylvie Méléard, and Philip Protter. The approximate euler method for lévy driven stochastic differential equations. 41(3):523–558, 2005.

[44] M. Keller-Ressel. Moment explosions and long-term behavior of affine stochastic volatility models. *Mathematical Finance*, 21(1):73–98, 2011.

[45] P.E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin, 1992.

[46] A. Kuznetsov, A.E. Kyprianou, J.C. Pardo, and K. Van Schaik. A Wiener–Hopf Monte Carlo simulation technique for Lévy processes. *The Annals of Applied Probability*, 21(6):2171–2190, 2011.

[47] A.E. Kyprianou. *Introductory lectures on fluctuations of Lévy processes with applications*. Springer Verlag, Berlin Heidelberg New York, 2006.

[48] D.B. Madan, P.P. Carr, and E.C. Chang. The variance gamma process and option pricing. *European Finance Review*, 2(1):79–105, 1998.

[49] R.C. Merton. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1-2):125–144, 1976.

[50] E. Platen. A generalized Taylor formula for solutions of stochastic equations. *Sankhyā: The Indian Journal of Statistics, Series A*, 44(2):163–172, 1982.

[51] E. Platen and N. Bruti-Liberati. *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*, volume 64 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, 1st edition, 2010.

[52] P.E. Protter. *Stochastic integration and differential equations*. Springer Verlag, Berlin Heidelberg New York, 2004.

[53] C. Ribeiro and N. Webber. Correcting for simulation bias in monte carlo methods to value exotic options in models driven by lévy processes. *Applied Mathematical Finance*, 13(4):333–352, 2006.

[54] W. Schoutens. *Lévy processes in finance: pricing financial derivatives*. Wiley-Blackwell, New Jersey, 2003.

[55] AL Teckentrup, R Scheichl, M. B. Giles, and E Ullmann. Further analysis of multilevel monte carlo methods for elliptic pdes with random coefficients. *Numerische Mathematik*, pages 1–32, 2012.

[56] Jan Vecer and Mingxin Xu. Pricing asian options in a semimartingale model. *Quantitative Finance*, 4(2):170–175, 2004.