

An importance sampling method based on the density transformation of Lévy processes

REIICHIRO KAWAI

Abstract

In this paper, we develop an importance sampling method with the help of flexible control on the Lévy measure in the density transformation. The method has significant efficacy even on evaluating random variables with complex path-dependent structures. Numerical examples are presented to illustrate convergence acceleration through variance reduction with a view towards financial derivatives pricing.

Keywords: CGMY process; Esscher transform; Gamma process; Meixner process; Monte Carlo simulations; series representation; subordination; variance reduction

2000 Mathematics Subject Classification: 65C05, 65D30, 60G51, 60E07.

1 Introduction

In practical finance, one often needs to rely on the Monte Carlo simulations, e.g., for pricing financial derivatives or for computing their Greeks as soon as their analytical formula cannot be derived. In particular, when asset price dynamics are modeled by general Lévy processes, we hardly arrive at desired analytical solutions. In spite of the intense need for the Monte Carlo method due to its wide applicability, its square root convergence is not always satisfactory and thus it has been a major effort to develop various variance reduction techniques for Monte Carlo simulations, including the importance sampling method.

The aim of this paper is to develop an importance sampling method based on the density transformation of Lévy processes. In the Brownian setting, on one hand, there is the literature discussing importance sampling and their methods are based only on transformation of the drift parameter, but not of the variance. Indeed, the Girsanov transformation tells us that in the Brownian case,

Published in *Monte Carlo Methods and Applications* (2006) **12**(2) 171-186.

Email address: reiichiro_kawai@ybb.ne.jp. Postal address : Financial Engineering, Fixed Income Department, Daiwa Securities SMBC Co.Ltd., 1-14-5, Eitai, Koto-ku, Tokyo, 135-0034, Japan. The views expressed in this paper are those of the author and do not reflect those of Daiwa Securities SMBC Co.Ltd. The author would like to thank Prof. Shigeyoshi Ogawa and Prof. Arturo Kohatsu-Higa for important discussions.

the transformation of the drift is the best one could do. On the other hand, although the density transformation of general Lévy processes is of much complexity, its relatively unrestricted control on the Lévy measure may change the shape of the marginal, which is the key of our construction.

The rest of the paper is organized as follows. Section 2 recalls the density transformation and the series representation of Lévy processes, which are building blocks for our entire construction. Section 3 introduces our new importance sampling method and, for illustration purpose, considers a simplified Esscher-transform framework so as to look closely at how our method serves as a variance reduction method. Section 4 presents numerical examples with discussion on some issues arising out of actual numerical procedures. Finally, Section 5 concludes this work.

2 Preliminaries

Let us begin with some notations which will be used throughout the text. \mathbb{R}^d is the d -dimensional Euclidean space with the norm $\|\cdot\|$, $\mathbb{R}_0^d := \mathbb{R}^d \setminus \{0\}$ and $\mathcal{B}(\mathbb{R}_0^d)$ is the Borel σ -field of \mathbb{R}_0^d . $\stackrel{\mathcal{L}}{=}$ and $\stackrel{\mathcal{L}}{\rightarrow}$ denote, respectively, equality and convergence in distribution, or of the finite dimensional distributions when random processes are considered. $\mathbb{D}([0, \infty), \mathbb{R}^d)$ is the space of càdlàg functions from $[0, \infty)$ into \mathbb{R}^d . $\mathbb{P}|_{\mathcal{F}_t}$ is the restriction of a probability measure \mathbb{P} to the σ -field \mathcal{F}_t . ΔX_t denotes the jump of X at time t , that is, $\Delta X_t := X_t - X_{t-}$, while the continuous part of X is denoted by $\tilde{X}_t := X_t - \sum_{s \in (0, t]} \Delta X_s$. As usual, for $a \in \mathbb{R}$, $(a)^+ := \max(a, 0)$. Finally, we say that a Lévy process $\{X_t : t \geq 0\}$ in \mathbb{R}^d is generated by the triplet (γ, A, ν) if its characteristic function is given by

$$\mathbb{E}[e^{i\langle y, X_t \rangle}] = \exp \left[t \left(i\langle y, \gamma \rangle - \frac{1}{2} \langle y, Ay \rangle + \int_{\mathbb{R}_0^d} (e^{i\langle y, z \rangle} - 1 - i\langle y, z \rangle \mathbb{1}_{(0,1]}(\|z\|)) \nu(dz) \right) \right],$$

where $\gamma \in \mathbb{R}^d$, A is a symmetric nonnegative-definite $d \times d$ matrix, and ν is a Lévy measure on \mathbb{R}_0^d , i.e.,

$$\int_{\mathbb{R}_0^d} (\|z\|^2 \wedge 1) \nu(dz) < +\infty. \quad (2.1)$$

2.1 Density transformation of Lévy processes

We here recall the density transformation of (multivariate) Lévy processes in the most general form. For more details, we refer the reader to Sato [7].

Let $(\{X_t : t \geq 0\}, \mathbb{P})$ and $(\{X_t : t \geq 0\}, \mathbb{Q})$ be Lévy processes in \mathbb{R}^d generated, respectively, by $(\gamma_{\mathbb{P}}, A_{\mathbb{P}}, \nu_{\mathbb{P}})$ and $(\gamma_{\mathbb{Q}}, A_{\mathbb{Q}}, \nu_{\mathbb{Q}})$, and let $(\mathcal{F}_t)_{t \geq 0}$ be the natural filtration of $\{X_t : t \geq 0\}$. If the following four conditions hold;

- (i) $A_{\mathbb{P}} = A_{\mathbb{Q}}$,
- (ii) the Lévy measures $\nu_{\mathbb{P}}$ and $\nu_{\mathbb{Q}}$ are equivalent,

(iii) the function $\varphi : \mathbb{R}_0^d \mapsto \mathbb{R}$, defined via

$$e^{\varphi(z)} = \frac{d\nu_{\mathbb{P}}}{d\nu_{\mathbb{Q}}}(z), \quad (2.2)$$

satisfies

$$\int_{\mathbb{R}_0^d} (e^{\varphi(z)/2} - 1)^2 \nu_{\mathbb{Q}}(dz) < +\infty, \quad (2.3)$$

(iv) the constants $\gamma_{\mathbb{P}}$ and $\gamma_{\mathbb{Q}}$ satisfy

$$\gamma_{\mathbb{Q}} - \gamma_{\mathbb{P}} - \int_{\|z\| \leq 1} z(\nu_{\mathbb{Q}} - \nu_{\mathbb{P}})(dz) = A_{\mathbb{P}}\eta, \quad (2.4)$$

for some $\eta \in \mathbb{R}^d$,

then the probability measures \mathbb{P} and \mathbb{Q} are equivalent with the Radon-Nikodym derivative,

$$\frac{d\mathbb{P}}{d\mathbb{Q}}|_{\mathcal{F}_t} = e^{U_t}, \quad \mathbb{Q}\text{-a.s.},$$

where the stochastic process $\{U_t : t \geq 0\}$ in \mathbb{R} is defined by, $\mathbb{Q}\text{-a.s.}$,

$$\begin{aligned} U_t := & \langle \eta, \tilde{X}_t \rangle - \frac{t}{2} \langle \eta, A_{\mathbb{Q}} \eta \rangle - t \langle \gamma_{\mathbb{Q}}, \eta \rangle \\ & + \lim_{\varepsilon \downarrow 0} \left[\sum_{(s, \Delta X_s) \in (0, t] \times \{\|z\| > \varepsilon\}} \varphi(\Delta X_s) - t \int_{\|z\| > \varepsilon} (e^{\varphi(z)} - 1) \nu_{\mathbb{Q}}(dz) \right]. \end{aligned} \quad (2.5)$$

Here, $\{U_t : t \geq 0\}$ is uniformly convergent in t on any bounded interval $\mathbb{Q}\text{-a.s.}$, and satisfies

$$\mathbb{E}_{\mathbb{Q}}[e^{U_t}] = \mathbb{E}_{\mathbb{P}}[e^{-U_t}] = 1, \quad t \in [0, \infty).$$

2.2 Series representation of Lévy processes

As seen in (2.5), the density transformation requires jumps of sample paths, and so the increment-based sample path is of no use for our purpose. We will use a well known jump-based sample paths generation method, the so-called series representation of Lévy processes. Let us here briefly summarize its essence. (See Rosiński [5] for details.) Assume that a Lévy measure ν can be decomposed in the following form

$$\nu(B) = \int_0^\infty \mathbb{P}(H(r, V) \in B) dr, \quad B \in \mathcal{B}(\mathbb{R}_0^d), \quad (2.6)$$

where V is a random variable taking values in a suitable space S with a distribution F , and where $H : (0, \infty) \times S \mapsto \mathbb{R}_0^d$ is such that for each $v \in S$, $r \mapsto \|H(r, v)\|$ is non-increasing. Then, a Lévy

process $\{X_t : t \in [0, T]\}$ generated by $(0, 0, \nu)$ admits the following series representation in the finite dimensional distributions sense,

$$\{X_t : t \in [0, T]\} \stackrel{\mathcal{L}}{=} \left\{ \sum_{i=1}^{\infty} \left[H(\Gamma_i/T, V_i) \mathbb{1}(T_i \leq t) - c_i \frac{t}{T} \right] : t \in [0, T] \right\}, \quad (2.7)$$

where $\{\Gamma_i\}_{i \geq 1}$ are arrival times of a standard Poisson process, $\{V_i\}_{i \geq 1}$ is a sequence of iid random variables in S with the distribution F , $\{T_i\}_{i \geq 1}$ is a sequence of iid uniform random variables on $[0, T]$, $\{c_i\}_{i \geq 1}$ is defined by

$$c_i := \mathbb{E}[H(\Gamma_i/T, V_i) \mathbb{1}(\|H(\Gamma_i/T, V_i)\| \leq 1)],$$

and the sequences $\{\Gamma_i\}_{i \geq 1}$, $\{V_i\}_{i \geq 1}$ and $\{T_i\}_{i \geq 1}$ are mutually independent. Moreover, several Lévy processes can be compared at once in Monte Carlo simulations by putting those random sequences in common, that is,

$$\left\{ \sum_{i=1}^{\infty} \left[H_k(\Gamma_i/T, V_i) \mathbb{1}(T_i \leq t) - c_i^k \frac{t}{T} \right] : t \in [0, T] \right\}, \quad k = 1, 2, \dots,$$

where H_k and $\{c_i^k\}_{i \geq 1}$ are derived suitably for each k . Let us note that the decomposition (2.6) is not necessarily unique.

Example 2.1. (CGMY process) The CGMY process (C, G, M, Y) $\{X_t : t \geq 0\}$ is a Lévy process in \mathbb{R} defined via the Lévy measure of the form

$$\nu(dz) = C|z|^{-1-Y} \left[e^{-G|z|} \mathbb{1}(z < 0) + e^{-M|z|} \mathbb{1}(z > 0) \right] dz,$$

where $C > 0$, $G \geq 0$, $M \geq 0$, and $Y < 2$. (See Carr et al[2] for details.) If $G, M \in (0, \infty)$ and $Y \in (0, 2)$, then it is a special class of the tempered stable process of Rosiński [6] and a series representation can be derived as follows,

$$\{X_t : t \in [0, T]\} \stackrel{\mathcal{L}}{=} \left\{ \sum_{i=1}^{\infty} \left[\left(\frac{Y\Gamma_i}{2CT} \right)^{-1/Y} \wedge E_i U_i^{1/Y} |V_i| \right] \frac{V_i}{|V_i|} \mathbb{1}(T_i \leq t) + \gamma_1 t : t \in [0, T] \right\}, \quad (2.8)$$

where $\{\Gamma_i\}_{i \geq 1}$ are arrival times of a standard Poisson process, $\{E_i\}_{i \geq 1}$ is a sequence of iid standard exponential random variables, $\{U_i\}_{i \geq 1}$ is a sequence of iid uniform random variables on $[0, 1]$, $\{V_i\}_{i \geq 1}$ is a sequence of iid random variables taking values $-G^{-1}$ and M^{-1} with equal probability, $\{T_i\}_{i \geq 1}$ is a sequence of iid uniforms on $[0, T]$, and γ_1 is a suitable constant. Here, the sequences $\{\Gamma_i\}_{i \geq 1}$, $\{E_i\}_{i \geq 1}$, $\{U_i\}_{i \geq 1}$, $\{V_i\}_{i \geq 1}$ and $\{T_i\}_{i \geq 1}$ are mutually independent.

Example 2.2. (Gamma process) The Gamma process $\{X_t : t \geq 0\}$ is an increasing Lévy process in $[0, \infty)$ whose characteristic function is given by

$$\mathbb{E}[e^{iyX_t}] = \exp \left[t \int_{(0, \infty)} (e^{iyz} - 1) \nu(dz) \right],$$

where

$$v(dz) = a \frac{e^{-z/b}}{z} dz, \quad z \in (0, \infty),$$

with $a, b \in (0, \infty)$. Its sample paths can be generated via the series representation

$$\{X_t : t \in [0, T]\} \stackrel{\mathcal{L}}{=} \left\{ \sum_{i=1}^{\infty} b e^{-\frac{\Gamma_i}{aT}} V_i \mathbb{1}(T_i \leq t) : t \in [0, T] \right\},$$

where $\{\Gamma_i\}_{i \geq 1}$, $\{V_i\}_{i \geq 1}$ and $\{T_i\}_{i \geq 1}$ are same as in the last example.

3 Main framework

We are now in a position to construct our method. Consider a Lévy process $\{X_t : t \geq 0\}$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$. Let F be a functional from $\mathbb{D}([0, T], \mathbb{R}^d)$ to \mathbb{R} and suppose that we want to evaluate

$$\mathbb{E}_{\mathbb{P}}[F(X)] (= I)$$

by Monte Carlo simulation, provided that the expectation is well defined. Following Monte Carlo procedures, on one hand, we first generate a sequence $\{X_t^{\mathbb{P}, i} : t \in [0, T]\}_{i \geq 1}$ of iid Lévy process sample paths, each of which replicates $\{X_t : t \in [0, T]\}$ under the probability measure \mathbb{P} , and then compute $n^{-1} \sum_{i=1}^n F(X^{\mathbb{P}, i})$. By the strong law of large numbers, $n^{-1} \sum_{i=1}^n F(X^{\mathbb{P}, i}) \rightarrow I$ a.s. as $n \rightarrow \infty$. The speed of convergence is usually measured in terms of the variance of the estimator, i.e.,

$$\text{Var} \left(\frac{1}{n} \sum_{i=1}^n F(X^{\mathbb{P}, i}) \right) = \frac{1}{n} \text{Var}_{\mathbb{P}}(F(X)).$$

Clearly, the convergence is likely to be faster with smaller $\text{Var}_{\mathbb{P}}(F(X))$. On the other hand, in view of the equality

$$\mathbb{E}_{\mathbb{P}}[F(X)] = \mathbb{E}_{\mathbb{Q}} \left[\frac{d\mathbb{P}}{d\mathbb{Q}} \Big|_{\mathcal{F}_T} F(X) \right] = \mathbb{E}_{\mathbb{Q}} [e^{U_T} F(X)],$$

we also get, \mathbb{Q} -a.s.,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n e^{U_T^i} F(X^{\mathbb{Q}, i}) = I, \quad (3.1)$$

where $\{X_t^{\mathbb{Q}, i} : t \in [0, T]\}_{i \geq 1}$ is a sequence of iid Lévy processes replicating $\{X_t : t \in [0, T]\}$ under \mathbb{Q} , and where $\{U_t^i : t \in [0, T]\}_{i \geq 1}$ is a corresponding sequence sampled in accordance with (2.5).

The Monte Carlo computation (3.1) after the density transformation will perform more efficiently once we find a probability measure \mathbb{Q} which makes the new variance $\text{Var}_{\mathbb{Q}}(e^{U_T} F(X))$ smaller than the original $\text{Var}_{\mathbb{P}}(F(X))$. A widely used criterion in choosing such a probability measure \mathbb{Q} is the importance sampling of the random variable $e^{U_T} F(X)$. The fundamental idea of the importance sampling is to choose \mathbb{Q} under which we may more likely sample Monte Carlo summands of large values, which contributes more to the computation.

Now, among the density transformations of Section 2.1, there is a simple class ending up with a path-independent structure and thus does not need jumps of sample paths. It is well known as the Esscher transform, where the function φ in (2.2) is linear, i.e., $\varphi(z) = \lambda z$ with some $\lambda \in \mathbb{R}$, so that $\sum \varphi(\Delta X)$ only depends on the terminal value of X . This simple structure enables us to investigate closely how our method may contribute to variance reduction in Monte Carlo simulations.

We first recall the Esscher transform in brief. Let $\{X_t : t \geq 0\}$ be a Lévy process in \mathbb{R} generated by (γ, σ^2, ν) and let $(\mathcal{F}_t)_{t \geq 0}$ be the natural filtration of $\{X_t : t \geq 0\}$. Assume that there exists $\lambda \in \mathbb{R}$ such that $\mathbb{E}_{\mathbb{P}}[e^{\lambda X_1}] < +\infty$. Under the probability measure \mathbb{Q}_λ , which is defined via the Radon-Nikodym derivative, \mathbb{P} -a.s.,

$$\frac{d\mathbb{Q}_\lambda}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = \frac{e^{\lambda X_t}}{\mathbb{E}_{\mathbb{P}}[e^{\lambda X_t}]},$$

the stochastic process $\{X_t : t \geq 0\}$ is again a Lévy process generated by $(\gamma_\lambda, \sigma^2, \nu_\lambda)$ where

$$\gamma_\lambda = \gamma + \int_{\|z\| \leq 1} z(\nu_\lambda - \nu)(dz)$$

and

$$\nu_\lambda(dz) = e^{\lambda z} \nu(dz).$$

The probability measure \mathbb{Q}_λ is then equivalent to \mathbb{P} . Therefore, we also get $\mathbb{E}_{\mathbb{Q}_\lambda}[e^{-\lambda X_1}] < +\infty$, and \mathbb{Q}_λ -a.s.,

$$\frac{d\mathbb{P}}{d\mathbb{Q}_\lambda} \Big|_{\mathcal{F}_t} = \frac{e^{-\lambda X_t}}{\mathbb{E}_{\mathbb{Q}_\lambda}[e^{-\lambda X_t}]}.$$

Moreover, letting $T > 0$ and let p be the probability density function of the random variable X_T under \mathbb{P} , the density function p_λ of X_T under \mathbb{Q}_λ is then given by

$$p_\lambda(x) = \frac{e^{\lambda x}}{\mathbb{E}_{\mathbb{P}}[e^{\lambda X_T}]} p(x), \quad x \in \mathbb{R},$$

and, conversely,

$$p(x) = \frac{e^{-\lambda x}}{\mathbb{E}_{\mathbb{Q}_\lambda}[e^{-\lambda X_T}]} p_\lambda(x), \quad x \in \mathbb{R}.$$

In the simplified Esscher-transform framework, suppose that we want to evaluate

$$\mathbb{E}_{\mathbb{P}}[f(X_T)] (= I)$$

by Monte Carlo simulation, where $f : \mathbb{R} \mapsto \mathbb{R}$ satisfies $\mathbb{E}_{\mathbb{P}}[f(X_T)] < +\infty$ and where the density p of X_T under \mathbb{P} is known. We generate a sequence of iid random variables $\{x_i^{\mathbb{P}}\}_{i \geq 1}$ with common density p and computes $n^{-1} \sum_{i=1}^n f(x_i^{\mathbb{P}})$. Just as before, in view of the equality

$$\mathbb{E}_{\mathbb{P}}[f(X_T)] = \mathbb{E}_{\mathbb{Q}_\lambda} \left[\frac{d\mathbb{P}}{d\mathbb{Q}_\lambda} \Big|_{\mathcal{F}_T} f(X_T) \right] = \mathbb{E}_{\mathbb{Q}_\lambda} \left[\frac{e^{-\lambda X_T}}{\mathbb{E}_{\mathbb{Q}_\lambda}[e^{-\lambda X_T}]} f(X_T) \right],$$

it also holds that, \mathbb{Q}_λ -a.s.,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{e^{-\lambda x_i^{\mathbb{Q}_\lambda}}}{\mathbb{E}_{\mathbb{Q}_\lambda}[e^{-\lambda X_T}]} f(x_i^{\mathbb{Q}_\lambda}) = I,$$

where $\{x_i^{\mathbb{Q}_\lambda}\}_{i \geq 1}$ is a sequence of iid random variables with common density p_λ , which is the density of X_T under \mathbb{Q}_λ . The new variance is then given by

$$\begin{aligned} \text{Var}_{\mathbb{Q}_\lambda} \left(\frac{e^{-\lambda x_1^{\mathbb{Q}_\lambda}}}{\mathbb{E}_{\mathbb{Q}_\lambda}[e^{-\lambda X_T}]} f(x_1^{\mathbb{Q}_\lambda}) \right) &= \int_{\mathbb{R}} \left(\frac{e^{-\lambda x}}{\mathbb{E}_{\mathbb{Q}_\lambda}[e^{-\lambda X_T}]} f(x) \right)^2 p_\lambda(x) dx - I^2 \\ &= \int_{\mathbb{R}^2} e^{\lambda(y-x)} f(x)^2 p(x) p(y) dx dy - I^2 =: J(\lambda). \end{aligned}$$

We want to find λ minimizing $J(\lambda)$, or such that $J(\lambda) < \text{Var}_{\mathbb{P}}(f(X_T))$. Here, the sign of the first derivative of J ,

$$J'(\lambda) = - \int_{\mathbb{R}^2} (x-y) e^{-\lambda(x-y)} p(x) p(y) f(x)^2 dx dy,$$

depends on p and f , while J is convex since

$$J''(\lambda) = \int_{\mathbb{R}^2} (x-y)^2 e^{-\lambda(x-y)} p(x) p(y) f(x)^2 dx dy \geq 0,$$

provided that for $n = 1, 2$,

$$\int_{\mathbb{R}^2} |x-y|^n e^{-\lambda(x-y)} p(x) p(y) f^2(x) dx dy < \infty.$$

Example 3.1. (European call) Consider the European call option,

$$f(X) = (\exp(X) - K)^+,$$

where K is a prescribed strike price. We set X to be a random variable with Meixner (a, b, d, m) distribution of Schoutens and Teugels [8] under the probability measure \mathbb{P} . Recall that the Meixner (a, b, d, m) distribution is infinitely divisible on \mathbb{R} defined via the Lévy measure of the form

$$\nu(dz) = d \frac{\exp(bz/a)}{z \sinh(\pi z/a)} dz, \quad z \in \mathbb{R}_0,$$

and the probability density function p of X is given in closed form by

$$p(x) = \frac{(2 \cos(b/2))^{2d}}{2a\pi\Gamma(2d)} e^{b(x-m)/a} \left| \Gamma \left(d + \frac{i(x-m)}{a} \right) \right|^2,$$

where $a > 0$, $b \in (-\pi, \pi)$, $d > 0$, and $m \in \mathbb{R}$. Note that the Meixner distribution is closed under the Esscher transform with $\lambda \in ((-\pi - b)/a, (\pi - b)/a)$, since

$$\nu_\lambda(dz) = e^{\lambda z} \nu(dz) = d \frac{\exp\left(\frac{a\lambda + b}{a} z\right)}{z \sinh(\pi z/a)} dz.$$

We here fix $(a, b, d) = (0.1, -1.0, 0.1)$, put $m = -2d \ln(\cos(b/2)/\cos((a+b)/2))$ to guarantee an arbitrage-free state, i.e., $\mathbb{E}_{\mathbb{P}}[\exp(X)] = 1$, and restrict to $\lambda \in (-21.416, 41.416)$. Following the fundamental idea of the importance sampling, we wish to evaluate intensively a domain of x with larger values of $(\exp(x) - K)^+ p(x)$. Clearly, the interval $x \in (-\infty, \ln K)$ makes no contribution to the computation since then $(\exp(x) - K)^+ = 0$. In this case, we intuitively want X to realize greater values, and thus wish to tilt the density p to the right, i.e. we set λ to be positive. For exact comparison purpose in the simulations, we will sample the sequences $\{x_i^{\mathbb{P}}\}_{i \geq 1}$ and $\{x_i^{\mathbb{Q}_\lambda}\}_{i \geq 1}$ by the transformation method with a common sequence of iid uniform random variables, i.e., for each i , $x_i^{\mathbb{P}} = G^{-1}(U_i)$ and $x_i^{\mathbb{Q}_\lambda} = G_\lambda^{-1}(U_i)$, where $G(x) = \int_{-\infty}^x p(y)dy$, $G_\lambda(x) = \int_{-\infty}^x p_\lambda(y)dy$, and $\{U_i\}_{i \geq 1}$ is a sequence of iid uniform random variables on $[0, 1]$.

The left of Figure 1 shows the variance ratios $J(\lambda)/J(0)$ for each fixed moneyness $K/S_0 (= K$, in this case). We find the variance reduced uniformly with $\lambda > 0$, and our method performs more effectively for the far-out-of-the-money cases. The right figure shows typical convergences with $K = 1.02$ under the original \mathbb{P} and under \mathbb{Q}_λ with $\lambda = 30$. The convergence is clearly accelerated, and 5,775 non-zero realizations out of 20,000 under \mathbb{Q}_λ , while only 1,994 under \mathbb{P} .

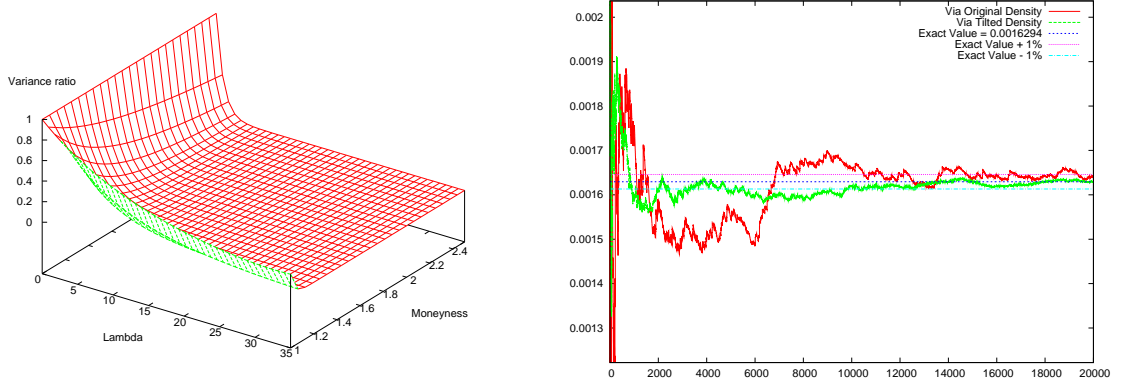


Figure 1: Variance ratios (left) and a comparison of convergence in Monte Carlo simulations with $\lambda = 30$ and $K = 1.02$ (right)

4 Numerical examples and considerations

We have observed that our method effectively accelerates convergence of Monte Carlo simulations in the simplified Esscher-transform framework. In this section, we put our method to full use with illustrating numerical examples. Before proceeding to numerical experiments, however, let us discuss two new possible numerical problems.

One issue is the discretization error, just as in most numerical procedures simulating stochastic process sample paths. In our case, the discretization originates from the series representation since

we cannot simulate infinity many jumps and thus need to discard small jumps. This is typically equivalent to truncating to a finite number of terms of the series due to the non-increasingness of the function $r \mapsto \|H(r, \cdot)\|$ in (2.6). To deal with this problem, we may use the Gaussian approximation of Asmussen and Rosiński [1], which approximates small jumps of Lévy processes by a Gaussian random variable. Let us briefly state its multivariate version (Cohen and Rosiński [3]). Let $\{X_t^\varepsilon : t \geq 0\}$ be a compensated Lévy process in \mathbb{R}^d without Gaussian component and consisting only of bounded jumps, i.e., its characteristic function is given in the form of

$$\mathbb{E}[e^{i\langle y, X_1^\varepsilon \rangle}] = \exp \left[\int_{\mathbb{R}_0^d} (e^{i\langle y, z \rangle} - 1 - i\langle y, z \rangle) \nu^\varepsilon(dz) \right],$$

where ν^ε is a restriction of a Lévy measure ν to $\{z \in \mathbb{R}_0^d : \|z\| \leq \varepsilon\}$ with some $\varepsilon > 0$. Moreover, let Σ_ε be the covariance matrix of X_1^ε , i.e., $\Sigma_\varepsilon = \int_{\mathbb{R}_0^d} zz^T \nu_\varepsilon(dz)$, and let W_d be a standard normal random vector in \mathbb{R}^d . Then, as $\varepsilon \downarrow 0$,

$$\Sigma_\varepsilon^{-1/2} X_1^\varepsilon \xrightarrow{\mathcal{L}} W_d,$$

if and only if for every $\kappa > 0$,

$$\lim_{\varepsilon \downarrow 0} \int_{\langle \Sigma_\varepsilon^{-1} z, z \rangle > \kappa} \langle \Sigma_\varepsilon^{-1} z, z \rangle \nu^\varepsilon(dz) = 0.$$

This result implies that the process consisting of discarded small jumps may also be approximated by a Brownian motion, i.e.,

$$\{\Sigma_\varepsilon^{-1/2} X_t^\varepsilon : t \geq 0\} \xrightarrow{\mathcal{L}} \{W_t : t \geq 0\},$$

where $\{W_t : t \geq 0\}$ is a d -dimensional standard Brownian motion, since for Lévy processes, the weak convergence of the marginal at some fixed time is necessary and sufficient for the weak convergence in $\mathbb{D}([0, \infty), \mathbb{R}^d)$ equipped with the Skorohod topology.

The second problem comes from the explosion of the compensation term in the definition (2.5) of $\{U_t : t \geq 0\}$, i.e.,

$$\int_{\mathbb{R}_0^d} |e^{\varphi(z)} - 1| \nu_{\mathbb{Q}}(dz) = +\infty. \quad (4.1)$$

Although the compensation term after the Gaussian approximation with any positive threshold ε is necessarily finite, i.e., $\int_{\|z\| > \varepsilon} |e^{\varphi(z)} - 1| \nu_{\mathbb{Q}}(dz) < +\infty$, it is still very sensitive to the threshold, since its choice directly relates to the singularity of the Lévy measure around the origin. Indeed, by our numerical experiences, the equation

$$\mathbb{E}_{\mathbb{Q}}[e^{U_t}] = 1, \quad t \in [0, \infty) \quad (4.2)$$

was far from being satisfied. To avoid this, we will replace the sequence $\{e^{U_t^i}\}_{i \geq 1}$ of Radon-Nikodym derivatives by its self-normalized version $\{e^{U_t^i} / \sum_{j=1}^i e^{U_t^j}\}_{i \geq 1}$. Note that the latter

achieves (4.2), while it does no longer consist of iid random variables. Let us mention that this was not a problem in the Esscher-transform framework since then the Esscher transform requires us to choose λ so that $\mathbb{E}_{\mathbb{P}}[e^{\lambda X_T}] < +\infty$, or equivalently, $\mathbb{E}_{\mathbb{Q}_\lambda}[e^{-\lambda X_T}] < +\infty$.

Let us give a precise summary of our method after the suggested modifications. Fix a sufficiently small threshold $\varepsilon \in (0, 1)$ and replace a Lévy process generated by $(\gamma_{\mathbb{P}}, 0, \nu_{\mathbb{P}})$ with one generated by $(\gamma_{\mathbb{P}}, A_\varepsilon, \nu_{\mathbb{P}} - \nu_{\mathbb{P}}^\varepsilon)$. Next, transform $(\gamma_{\mathbb{P}}, A_\varepsilon, \nu_{\mathbb{P}} - \nu_{\mathbb{P}}^\varepsilon)$ into $(\gamma_{\mathbb{Q}}, A_\varepsilon, \nu_{\mathbb{Q}} - \nu_{\mathbb{Q}}^\varepsilon)$, so that the new probability measure \mathbb{Q} is likely to provide a faster convergence in interested Monte Carlo simulations. Here, the transformation is defined via the Radon-Nikodym derivative, \mathbb{Q} -a.s.,

$$\frac{d\mathbb{P}}{d\mathbb{Q}}|_{\mathcal{F}_t} = e^{U_t}, \quad (4.3)$$

where

$$U_t = \sum_{(s, \Delta X_s) \in (0, t] \times \{\|z\| > \varepsilon\}} \varphi(\Delta X_s) - t \int_{\|z\| > \varepsilon} (e^{\varphi(z)} - 1) \nu_{\mathbb{Q}}(dz).$$

Notice that the density transformation via (4.3) has no effect on the approximating Gaussian component. Moreover, the Lévy measure $\nu_{\mathbb{Q}}$ and the constant $\gamma_{\mathbb{Q}}$ are required to satisfy, respectively,

$$\int_{\|z\| > \varepsilon} (e^{\varphi(z)/2} - 1)^2 \nu_{\mathbb{Q}}(dz) < +\infty,$$

where

$$e^{\varphi(z)} = \frac{d\nu_{\mathbb{P}}}{d\nu_{\mathbb{Q}}}(z) \mathbb{1}(\|z\| > \varepsilon),$$

and with $\eta = 0$ in (2.4),

$$\gamma_{\mathbb{Q}} = \gamma_{\mathbb{P}} + \int_{\|z\| \in (\varepsilon, 1]} z(\nu_{\mathbb{Q}} - \nu_{\mathbb{P}})(dz). \quad (4.4)$$

We then compute

$$\frac{1}{n} \sum_{i=1}^n \frac{e^{U_T^{\mathbb{Q}, i}}}{\frac{1}{i} \sum_{j=1}^i e^{U_T^{\mathbb{Q}, j}}} F(X^{\mathbb{Q}, i}),$$

where $\{X_t^{\mathbb{Q}, i} : t \in [0, T]\}_{i \geq 1}$ is a sequence of iid Lévy processes, each of which replicates $\{X_t : t \in [0, T]\}$ under \mathbb{Q} and where $\{U_t^{\mathbb{Q}, i} : t \in [0, T]\}_{i \geq 1}$ is a sequence of iid random processes replicating $\{U_t : t \in [0, T]\}$ under \mathbb{Q} , defined by, \mathbb{Q} -a.s.,

$$U_t = \sum_{s \in (0, t]} \varphi(\Delta X_s) - t \int_{\|z\| > \varepsilon} (e^{\varphi(z)} - 1) \nu_{\mathbb{Q}}(dz).$$

Example 4.1. (A lookback type) Suppose that we are interested in the expectation

$$\mathbb{E}_{\mathbb{P}} \left[\left(\sup_{t \in [0, 1]} \exp(X_t) - \inf_{t \in [0, 1]} \exp(X_t) - K \right)^+ \right] \quad (=:\mathbb{E}_{\mathbb{P}}[F(X)]),$$

where $\{X_t : t \geq 0\}$ is a CGMY process generated by $(\gamma_{\mathbb{P}}, 0, \nu_{\mathbb{P}})$ and $\gamma_{\mathbb{P}}$ is chosen so that $\mathbb{E}_{\mathbb{P}}[\exp(X_1)] = 1$. Its analytical solution is unknown in the Lévy setting. From the path-dependent structure above, we infer that as the “volatility”^{*} of X is greater, the random variable inside the expectation is more likely to generate non-zero realizations. For clear illustration, we will compare three CGMY processes;

- (i) original one; $(C_1, G_1, M_1, Y_1) = (0.03, 4.0, 4.0, 1.5)$,
- (ii) well-transformed one; $(C_2, G_2, M_2, Y_2) = (C_1, 1.0, 1.0, Y_1)$,
- (iii) badly-transformed one; $(C_3, G_3, M_3, Y_3) = (C_1, 9.0, 9.0, Y_1)$.

Clearly, we intend by (ii) greater jumps via fatter tails of the Lévy measure, while smaller ones by (iii) via the heavily tempered Lévy measure.

We use the series representation of (2.8) to generate jump-based sample paths and remove small jumps $|\Delta X| \leq \varepsilon$. In this example, we put $\varepsilon = 1.0 \times 10^{-5}$, where the required number of jumps is then around 5×10^3 for (i), 2×10^4 for (ii) and 3×10^3 for (iii). Denoting by $(\gamma_{\mathbb{P}}, \sigma_{\varepsilon}^2, \nu_{\mathbb{P}} - \nu_{\mathbb{P}}^{\varepsilon})$ the triplet after the Gaussian approximation, where

$$\sigma_{\varepsilon}^2 = \int_{|z| \leq \varepsilon} z^2 \nu_{\mathbb{P}}(dz) = C_1(M_1^{Y_1-2} \Gamma(2 - Y_1, \varepsilon M_1) + G_1^{Y_1-2} \Gamma(2 - Y_1, \varepsilon G_1)), \quad (4.5)$$

and $\Gamma : \mathbb{R}^+ \times \mathbb{R}^+ \mapsto \mathbb{R}^+$ is the lower incomplete gamma function, we perform a transformation into $(\gamma_{\mathbb{Q}}, \sigma_{\varepsilon}^2, \nu_{\mathbb{Q}} - \nu_{\mathbb{Q}}^{\varepsilon})$, where $\nu_{\mathbb{Q}}$ is the well-transformed Lévy measure of (ii) and where $\gamma_{\mathbb{Q}}$ is the constant satisfying (4.4). The sequence $\{e^{U_i^i}\}_{i \geq 1}$ of Radon-Nikodym derivatives is generated in accordance with

$$e^{U_1} = \exp \left[\sum_{s \in (0,1]} \varphi(\Delta X_s) - \gamma_2 \right],$$

where the function φ is defined via

$$e^{\varphi(z)} = e^{-(G_1 - G_2)|z|} \mathbb{1}(z < -\varepsilon) + e^{-(M_1 - M_2)|z|} \mathbb{1}(z > \varepsilon), \quad (4.6)$$

and where the constant γ_2 is given by

$$\gamma_2 = \nu_{\mathbb{P}}(\{z \in \mathbb{R}_0 : |z| > \varepsilon\}) - \nu_{\mathbb{Q}}(\{z \in \mathbb{R}_0 : |z| > \varepsilon\}). \quad (4.7)$$

Finally, we use the self-normalized sequence $\{e^{U_i^i} / \frac{1}{i} \sum_{j=1}^i e^{U_j^j}\}_{i \geq 1}$, instead of $\{e^{U_i^i}\}_{i \geq 1}$. The entire procedure for (iii) can be taken just in a similar manner.

Table 4 summarizes the variance ratios, with respect to the original CGMY, estimated from 20,000 Monte Carlo summands for several K 's. We see that the simulation works more efficiently for greater K , just as expected. The right of Figure 2 shows a typical convergence when $K = 0.5$. The number of non-zero realizations is 4,487 out of 20,000 in the “Original”, 6,100 in the “Well-transformed” achieving faster convergence, and 2,974 for the “Badly-transformed” with an undesirable performance.

K	0.0	0.1	0.2	0.3	0.4	0.5
Ratio1	2.1902	1.3055	0.79732	0.51505	0.53705	0.38410
Ratio2	25.475	11.539	15.913	8.9055	6.5652	43.029

Table 1: Empirical variance ratios ; Ratio1= $\frac{(\text{well-transformed})}{(\text{original})}$ and Ratio2= $\frac{(\text{badly-transformed})}{(\text{original})}$

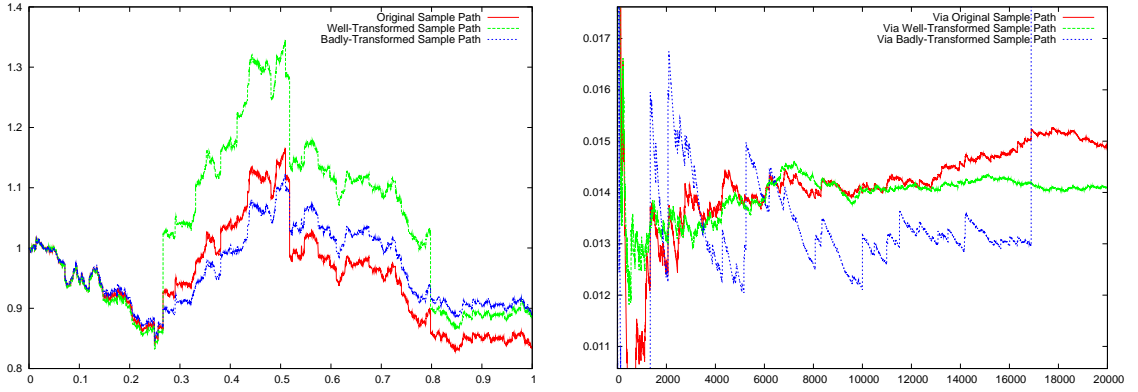


Figure 2: Typical sample path and its transformed sample paths (left) and a typical convergence in Monte Carlo simulations with $K = 0.5$ (right)

One might have wondered whether or not our importance sampling method with the jump-based sample paths outperforms an ordinary Monte Carlo simulation with the increment-based sample paths in terms of the CPU time. We give in Table 2 the elapsed time required for $\pm 0.5\%$ accuracy, together with that by the Monte Carlo simulation with increment-based sample paths. For the increment-based MC, each sample path is approximated by 500 equidistant increments and the Fourier inversion computation time (only once) for the unknown CGMY density is included in the CPU time. Since the true value is unknown, we set the average of 5×10^4 samples from the “Original” simulation to be the basis for the $\pm 0.5\%$ accuracy for “Original” and “Well-transformed”, while the average of 1×10^5 samples from the “increment-based MC” simulation to be its basis. This distinction is made since the rough discretization by 500 increments in “increment-based MC” yields a smaller expected value, due to $\sup_{t \in [0,1]} \exp(X_t) \geq \sup_{t \in \{t_1, t_2, \dots\}} \exp(X_t)$ and $\inf_{t \in [0,1]} \exp(X_t) \leq \inf_{t \in \{t_1, t_2, \dots\}} \exp(X_t)$. Let us also mention that the extra computations of the Radon-Nikodym derivatives and of more jumps for “Well-transformed” did not have much impact on the CPU times. (Remember that we need to generate around 2×10^4 jumps to achieve the threshold $\varepsilon = 1.0 \times 10^{-5}$ for each sample path in “Well-transformed”, while only 5×10^3 jumps in “Original.”) Table 2 indicates that our method is still behind the increment-based sample paths

*We use the word “volatility” for an association with the Black-Scholes setting although it is an abuse of word for a Lévy process without Gaussian component.

framework in terms of the CPU time at least for this example.

Method	#Sample ($\times 10^4$)	CPU time ($\times 10^2$ sec)
Original	4	13
Well-transformed	2	7
Increment-based MC	6	4

Table 2: #Sample and CPU time required for 4 digits accuracy

Example 4.2. (Lookback type revisited) On comparing CPU times in the last example, we need only one CGMY density, which can be used throughout the entire Monte Carlo simulations. *What if each Monte Carlo summand requires us to perform the Fourier transform to compute a density function?* Its vast computational cost would then make such Monte Carlo simulations too expensive to be of practical use. Let us here present such a case in which our method would be the only choice.

Consider $\{X_t : t \in [0, 1]\}$, where $\{X_t : t \geq 0\}$ is a CGMY process and $\{Y_t : t \in [0, 1]\}$ is a Gamma process independent of $\{X_t : t \geq 0\}$. We call the transformation of X to X_Y *subordination* by the *subordinator* Y . The subordination has recently attracted much attention in financial modelling. See, for example, Cont and Tankov [4]. Now, as in the last example, let us again consider a lookback type,

$$\mathbb{E}_{\mathbb{P}} \left[\left(\sup_{t \in [0, 1]} \exp(X_t) - \inf_{t \in [0, 1]} \exp(X_t) - K \right)^+ \right] \quad (=:\mathbb{E}_{\mathbb{P}}[F(X, Y)]). \quad (4.8)$$

Here, we need not generate sample paths of the subordinator but only its terminal values Y_1 , since by Y -pointwise,

$$\sup_{t \in [0, 1]} \exp(X_t) - \inf_{t \in [0, 1]} \exp(X_t) = \sup_{t \in [0, Y_1]} \exp(X_t) - \inf_{t \in [0, Y_1]} \exp(X_t),$$

and then by fixing Y_1 , we get

$$\{X_t : t \in [0, Y_1]\} \stackrel{\mathcal{L}}{=} \left\{ \sum_{i=1}^{\infty} \left[H(\Gamma_i/Y_1, V_i) \mathbb{1}(T_i \leq t) - c_i \frac{t}{Y_1} \right] : t \in [0, Y_1] \right\},$$

where $\{T_i\}_{i \geq 1}$ is now a sequence of iid uniform random variables on $[0, Y_1]$, instead of $[0, T]$. Meanwhile, in the increment-based sample paths framework, each Monte Carlo summand requires a density of its own because of the randomness of the terminal value Y_1 , and so it is unrealistic to implement the increment-based MC for this example.

Now, in view of the trivial inequality, for $t_1 < t_2$,

$$\sup_{t \in [0, t_1]} \exp(X_t) - \inf_{t \in [0, t_1]} \exp(X_t) \leq \sup_{t \in [0, t_2]} \exp(X_t) - \inf_{t \in [0, t_2]} \exp(X_t),$$

we want the terminal value Y_1 to end up with greater values. Since

$$\nu_{\mathbb{Q}_\lambda}(dz) = e^{\lambda z} \nu_{\mathbb{P}}(dz) = a \frac{e^{-\frac{1-b\lambda}{b}z}}{z} dz,$$

$\{Y_t : t \in [0, 1]\}$ is again a Gamma process after the Esscher transform as long as $\lambda \in (-\infty, 1/b)$, and the new probability measure \mathbb{Q}_λ is defined via the Radon-Nikodym derivative, \mathbb{Q}_λ -a.s.,

$$\frac{d\mathbb{P}}{d\mathbb{Q}_\lambda} \Big|_{\mathcal{F}_t} = \frac{e^{-\lambda Y_t}}{\mathbb{E}_{\mathbb{Q}_\lambda}[e^{-\lambda Y_t}]}, \quad t \in [0, 1],$$

where $(\mathcal{F}_t)_{t \in [0, 1]}$ is the natural filtration of $\{Y_t : t \in [0, 1]\}$.

Here, we fix $(C, G, M, Y) = (0.02, 1.0, 1.0, 1.8)$ for the CGMY process $\{X_t : t \geq 0\}$ and $(a, b) = (1, 1)$ for the original Gamma process $\{Y_t : t \in [0, 1]\}$, and perform the Esscher transform on the Gamma subordinator Y with the Esscher parameter $\lambda = 0.3$ and 0.6 . To be more precise, we will compare the original estimator

$$\frac{1}{n} \sum_{i=1}^n F(X^{\mathbb{P}, i}, Y^{\mathbb{P}, i}),$$

with the transformed one

$$\frac{1}{n} \sum_{i=1}^n \frac{e^{-\lambda Y_1^{\mathbb{Q}_\lambda, i}}}{\mathbb{E}_{\mathbb{Q}_\lambda}[e^{-\lambda Y_1}]} F(X^{\mathbb{Q}_\lambda, i}, Y^{\mathbb{Q}_\lambda, i}) = \frac{1}{n} \sum_{i=1}^n \frac{e^{-\lambda Y_1^{\mathbb{Q}_\lambda, i}}}{\mathbb{E}_{\mathbb{Q}_\lambda}[e^{-\lambda Y_1}]} F(X^{\mathbb{P}, i}, Y^{\mathbb{Q}_\lambda, i}),$$

where the last equality holds since X is independent of the filtration $(\mathcal{F}_t)_{t \in [0, 1]}$. For each i , we sample $Y_1^{\mathbb{P}, i}$ and $Y_1^{\mathbb{Q}_\lambda, i}$ via the transformation method with a common uniform random variable.

Table 3 summarizes variance ratios estimated from 20,000 Monte Carlo summands, while the right of Figure 3 shows a typical convergence with $K = 2.5$, where the number of non-zeros is only 49 out of 20,000 via the original subordinator, while 128 via the transformed one and 578 via the more transformed one. We can see that the transformation \mathbb{Q}_λ with positive λ contributes to a faster convergence.

5 Conclusion

In this paper, we have developed a new importance sampling method based on the density transformation of Lévy processes in the form of the jumps-based sample paths. The relatively unrestricted control on the Lévy measure in the density transformation of Lévy processes enables us to change shapes of the marginal density, unlike the drift is the only transformable parameter in the Brownian

K	0.0	0.5	1.0	1.5	2.0	2.5
Ratio1	0.58296	0.70519	0.72995	0.97790	0.65842	0.33633
Ratio2	0.99009	0.69830	0.68483	0.89271	0.41816	0.30966

$$\text{Ratio1} = \frac{(\text{transformed}(\lambda=0.3))}{(\text{original})} \text{ and } \text{Ratio2} = \frac{(\text{more-transformed}(\lambda=0.6))}{(\text{original})}$$

Table 3: Empirical variance ratios

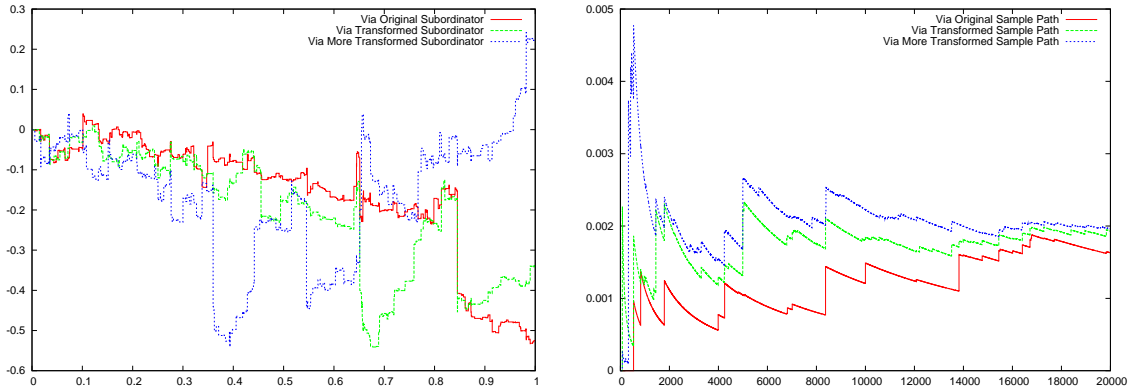


Figure 3: Typical sample paths of subordinated CGMY process $\{X_{Y_t} : t \in [0, 1]\}$ (left) and a typical convergence in Monte Carlo simulations with $K = 2.5$ (right)

case. In actual numerical procedures, the discretization error may be eased through approximating discarded small jumps by a Brownian motion, while the unit mean of Radon-Nikodym derivative sequences can be achieved by instead using its self-normalized version.

The numerical examples show that using our method with a “right” choice of the density transformation leads to very efficient simulations, especially for random variables with complex path-dependent structure. We expect that our method plays an active part in a variety of intricate financial derivatives modelling in the future. Meanwhile, we also find our method still behind the increment-based Monte Carlo in terms of the CPU time, especially for simple Monte Carlo simulations. This is mainly due to the cost of sample paths generation via the series representation, and its cost improvement is left as a future research.

References

- [1] Asmussen, S. and Rosiński, J. (2001) Approximations of small jumps of Lévy processes with a view towards simulations, *J. Appli. Probab.*, **38**(2) 482-493.

- [2] Carr, P., Geman, H., Madan, D.B. and Yor, M. (2002) The fine structure of asset returns: An empirical investigation, *J. Business*, **75**, 303-325.
- [3] Cohen, S. and Rosiński, J. (2007) Gaussian approximation of multivariate Lévy processes with applications to simulation of tempered and operator stable processes, To appear in *Bernoulli*.
- [4] Cont, R. and Tankov, P. (2003) *Financial Modelling With Jump Processes*, Chapman & Hall / CRC Press.
- [5] Rosiński, J. (2001) Series representations of Lévy processes from the perspective of point processes, *Lévy Processes - Theory and Applications*, Eds. O.E. Barndorff-Nielsen, T. Mikosch and S.I. Resnick, Birkhäuser, Boston, 401-415.
- [6] Rosiński, J. (2007) Tempering stable processes, To appear in *Stochastic Processes and their Applications*.
- [7] Sato, K. (1999) *Lévy processes and infinitely divisible distributions*, Cambridge University Press.
- [8] Schoutens, W. and Teugels, J.L. (1998) Lévy processes, polynomials and martingales, *Communications in Statistics: Stochastic Models*, **14**, 335-349.