On Sample Average Approximation Algorithms for Determining the Optimal Importance Sampling Parameters in Pricing Financial Derivatives on Lévy Processes ☆

Guangxin Jiang[a,*], Michael C. Fu[b], Chenglong Xu[c]

[a]*Department of Economics and Finance, City University of Hong Kong, Kowloon, Hong Kong*
[b]*Robert H. Smith School of Business and Institute for Systems Research, University of Maryland, College Park, Maryland, 20742, USA*
[c]*Department of Mathematics, Tongji University, Shanghai, 200092, China*

Abstract

We formulate the problem of determining the optimal importance sampling measure change for pricing financial derivatives under Lévy processes as a parametric optimization problem, and propose a solution approach using sample average approximation (SAA) with Newton iteration to find the optimal parameters in the Esscher probability measure change. Theoretical results, such as convergence rate of the optimal solutions and the effectiveness of the algorithms, are provided.

Keywords: importance sampling; sample average approximation; Newton iteration; infinitesimal perturbation analysis; Lévy processes

*Corresponding author
Email address: guajiang@cityu.edu.hk (Guangxin Jiang)

1. Introduction

Lévy processes can model critical characteristics observed in financial markets, such as heavy tails and jumps; see Cont [1] and Schoutens [2] for a survey. Due to the complicated nature of Lévy process models, Monte Carlo simulation is routinely used to price the derivatives (Carr and Madan [3], Glasserman [4]) and estimate Greeks (Fu [5], Glasserman and Liu [6], Kawai [7]). Although Monte Carlo simulation can solve high-dimensional problems, variance reduction techniques are often needed to improve the computational efficiency; see Glasserman [4] for a survey. For Lévy processes models, variance reduction techniques used in derivatives pricing include importance sampling (Kawai [8]), control variates (Dingeç and Hörmann [9]), stratified sampling (Kawai [10]), etc.

In this paper, we focus on importance sampling. The basic idea of importance sampling is to find a new probability measure such that the variance of the importance sampling estimator under the new probability measure should be smaller or even zero (the perfect probability measure, cf. Homem-de-Mello [11]). In practice, the perfect probability measure is generally unavailable, and the importance sampling problem can be formulated as a parametric optimization problem. Based on Monte Carlo sampling techniques, Su and Fu [12] used stochastic approximation (SA) to solve the resulting stochastic optimization problem in the geometric Brownian motion setting (also refer to Lapeyre and Lelong [13]), and Kawai [14] and Kawai [8] extended the approach to Lévy processes.

However, in SA, the proper choice of step sizes is difficult. We consider an alternative approach: sample average approximation (SAA) (Shapiro [15], Kim et al. [16]). SAA combined with a good deterministic algorithm has proven effective for solving certain classes of stochastic optimization problems where the

2

classical SA numerical procedures perform poorly (Nemirovski et al. [17]). In this paper, we propose a new gradient-based search optimization procedure combining SAA with Newton iteration to find the optimal parameter of the Esscher measure change. For gradient estimation, see Fu [18] for survey. Jourdain and Lelong [19] and Badouraly-Kassim et al. [20] used Newton iteration algorithm on SAA to find the optimal parameters in Black-Scholes model and jump diffusion models, respectively, by exploiting the special structures of these two models.

In our work, we consider more general Lévy process models. We show that importance sampling framework naturally satisfies the requirements for using SAA, and provide convergence proofs based on the classical results in SAA. The rest of the paper is organized as follows. In Section 2, we briefly review Lévy processes and the Monte Carlo method with importance sampling. Section 3 introduces the proposed SAA-Newton iteration method and analyzes the solution properties. Based on these, an efficient algorithm is given. We also provide a simple way to verify the conditions in the theorems. Section 4 concludes.

2. Preliminaries

2.1. Lévy processes models

Let $\mathbf{X} = \{X_t, t \geq 0\}$ be a d-dimensional stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, satisfying the following conditions:

(1) $X_0 = 0$ a.s.

(2) $\mathbf{X}$ has independent and stationary increments.

(3) $\mathbf{X}$ is stochastically continuous, i.e., $\lim_{t \to s} \mathbb{P}(||X_t - X_s|| > a) = 0$, for all $a > 0$ and $s \geq 0$, where $|| \cdot ||$ is the Euclidean norm.

Then $\mathbf{X}$ is a Lévy process.

Let $\langle \cdot, \cdot \rangle$ denote the inner product. X_t is infinitely divisible for each $t \geq 0$, and through the Lévy-Khintchine formula, its characteristic function can be written as

$$
\begin{aligned}
\Phi_{X_t}(u) \quad &= \mathbb{E}_P \left[e^{i\langle X_t, u \rangle} \right] = \exp \left\{ t \left(i\langle b, u \rangle - \frac{1}{2} \langle u, Au \rangle \right. \right. \\
&\quad + \left. \left. \int_{\mathbb{R}^d - \{0\}} \left[e^{i\langle u, y \rangle} - 1 - i\langle u, y \rangle \mathbb{1}_{B_1(0)}(y) \nu(dy) \right] \right) \right\},
\end{aligned}
$$

where $u \in \mathbb{R}^d$, $b \in \mathbb{R}^d$, A is a non-negative definite symmetric $d \times d$ matrix, and ν is a Lévy measure on $\mathbb{R}^d - \{0\}$. A measure ν is a *Lévy measure* if it satisfies $\int_{\mathbb{R}^d - \{0\}} \left(|y|^2 \wedge 1 \right) \nu(dy) < \infty$. Let Λ denote the process parameters, e.g., $\Lambda = (\sigma, \nu, \theta)$ for the VG process and $\Lambda = (\mu, \sigma)$ for Brownian motion, which determine the corresponding processes.

Let $F(X)$ be the payoff of the financial derivative, given by

$$
F(X) = F(X_t; 0 \leq t \leq T),
$$

where T is the maturity. The goal is to calculate the discounted expectation of $F(X)$.

2.2. *Importance sampling and a parametric optimization problems*

To improve computing efficiency, variance reduction techniques are routinely used with Monte Carlo simulation. In this paper, we focus on importance sampling. In the Black-Scholes model, the drift is changed (Glasserman et al. [21], Su and Fu [12]). For Lévy processes, the Esscher measure change is commonly used (Kawai [8]).

Let $\varphi_t(\lambda)$ denote the cumulant generating function of X_t, i.e.,

$$
\varphi_t(\lambda) = \log \mathbb{E}_{\mathbb{P}} \left[e^{\langle \lambda, X_t \rangle} \right] = \log \Phi_{X_t}(-i\lambda),
$$

where $\lambda \in C \subseteq \mathbf{R}^d$ and C is a nonempty convex compact set. Given another probability measure $\mathbb{P}_\lambda$ that is absolutely continuous w.r.t. $\mathbb{P}$, the Radon-Nikodym derivative is given by

$$\left.\frac{d\mathbb{P}_\lambda}{d\mathbb{P}}\right|_{\mathcal{F}_t} = \frac{e^{\langle\lambda,X_t\rangle}}{\mathbb{E}_\mathbb{P}\left[e^{\langle\lambda,X_t\rangle}\right]} = e^{\langle\lambda,X_t\rangle-\varphi_t(\lambda)},$$

where $\mathcal{F}_t$ is the natural filtration of $\{X_t, t \geq 0\}$. Note that $\lambda = 0$ corresponds to the initial probability measure $\mathbb{P}$. Suppose that $\mathbb{E}_{\mathbb{P}_\lambda}[F(X)] < \infty$, then applying the Esscher measure change to $\mathbb{E}_\mathbb{P}[F(X)]$,

$$
\begin{aligned}
V : \quad &= \quad \mathbb{E}_\mathbb{P}[F(X)] = \mathbb{E}_{\mathbb{P}_\lambda}\left[\left.\frac{d\mathbb{P}}{d\mathbb{P}_\lambda}\right|_{\mathcal{F}_T} F(X)\right] \\
&= \quad \mathbb{E}_{\mathbb{P}_\lambda}\left[\left(\left.\frac{d\mathbb{P}_\lambda}{d\mathbb{P}}\right|_{\mathcal{F}_T}\right)^{-1} F(X)\right] \\
&= \quad \mathbb{E}_{\mathbb{P}_\lambda}\left[e^{-\langle\lambda,X_T\rangle+\varphi_T(\lambda)}F(X)\right].
\end{aligned}
$$

The variance of $F(X)$ under $\mathbb{P}_\lambda$ is given by

$$
\begin{aligned}
\text{Var}(F(X),\lambda): \quad &= \quad \mathbb{E}_{\mathbb{P}_\lambda}\left[\left(\left.\frac{d\mathbb{P}}{d\mathbb{P}_\lambda}\right|_{\mathcal{F}_T}\right)^2 F(X)^2\right] - V^2 \\
&= \quad \mathbb{E}_\mathbb{P}\left[\left(\left.\frac{d\mathbb{P}_\lambda}{d\mathbb{P}}\right|_{\mathcal{F}_T}\right)^{-1} F(X)^2\right] - V^2 \\
&= \quad \mathbb{E}_\mathbb{P}\left[e^{-\langle\lambda,X_T\rangle+\varphi_T(\lambda)}F(X)^2\right] - V^2. \quad (1)
\end{aligned}
$$

We call λ^* that minimizes the variance $\text{Var}(F(X),\lambda)$, the *optimal parameter*, i.e.,

$$\lambda^* \in \arg\min_{\lambda \in C} \text{Var}(F(X),\lambda). \quad (2)$$

Remark 1. Finding the optimal probability measure requires solving the corresponding stochastic optimization problem. In this paper, we restrict the candidate

5

probability measures to a parametric class to simplify the optimization problem. Thus, "optimal" means the optimal in this restricted class, which is not necessarily optimal over all probability measures. In the next section, we show that the objective function, which applies the Esscher measure change, has good regularity conditions, so that the corresponding optimization problem can be solved efficiently by the SAA approach.

3. Main Results

3.1. The SAA-Newton method

In this section, we first use SAA to change the stochastic optimization problem to a deterministic optimization problem, then apply the Newton iteration method (also known as the Newton-Raphson method), a common method for finding roots of a real-valued differentiable function system in numerical analysis and the optimal solutions in deterministic optimization. Denote

$$f(\lambda) := \mathbb{E}_{\mathbb{P}}\left[g(X, \lambda)\right],$$

where

$$g(X, \lambda) := e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} F(X)^2.$$

Generally, the function $f(\lambda)$ cannot be observed or computed directly, but we can observe $g(X, \lambda)$, where $g(\cdot, \cdot)$ is a deterministic real-valued function, and X is simulated. We assume that $|\varphi_T(\lambda)| < \infty$ for all $\lambda \in C$. Then by (1), solving the optimization problem (2) is equivalent to solving the problem

$$\min_{\lambda \in C} f(\lambda). \tag{3}$$

6

In SAA, we generate independent and identically distributed (i.i.d.) paths of X denoted by $X^1, X^2, \ldots, X^n$, and let

$$f_n(\lambda) = \frac{1}{n} \sum_{i=1}^{n} g(X^i, \lambda).$$

Then, we formulate a deterministic optimization problem

$$\min_{\lambda \in C} f_n(\lambda), \tag{4}$$

which can be solved by a deterministic algorithm such as Newton iteration method.

Let π^* and Π_n^* denote the set for optimal solutions for the original problem (3) and the SAA problem (4), respectively, and v^* and v_n^* denote the optimal respective values. Let $\mathbb{D}(A, B)$ denote the distance between set A and set B, i.e., $\mathbb{D}(A, B) = \sup\{d(x, B) : x \in A\}$, where $d(x, B) = \inf_{y \in B} \|x - y\|$. Based on the following lemma, we show that $\mathbb{D}(\pi^*, \Pi_n^*) \to 0$ and $v_n^* \to v^*$ as $n \to \infty$ under some mild conditions.

Lemma 1 (Theorem 9, Kim et al. [16]). *Suppose that there exists a compact subset $C \subset \mathbf{R}^d$ such that:*
(i) π^* is non-empty and contained in C,
(ii) $\{f_n(x)\}$ converges uniformly to $f(x)$ on C a.s. as $n \to \infty$, and
(iii) for sufficiently large n, Π_n^* is non-empty and contained in C a.s.
Then $v_n^* \to v^*$ a.s. as $n \to \infty$. Furthermore, if f is continuous on C, then $\mathbb{D}(\pi^*, \Pi_n^*) \to 0$ a.s. as $n \to \infty$.

Let $\Xi_1 = \left\{ \lambda \in C : \mathbb{E}_\mathbb{P}\left[e^{-2p\langle \lambda, X_T \rangle}\right] < \infty \text{ and } \mathbb{E}_\mathbb{P}[F(X)^{4p}] < \infty \text{ for } p > 1 \right\}$, then we have the following theorem.

Theorem 1. *If π^* and Π_n^* are non-empty and contained in Ξ_1, $\lambda \in \Xi_1$, then $v_n^* \to v^*$ and $\mathbb{D}(\pi^*, \Pi_n^*) \to 0$ a.s. as $n \to \infty$.*

Proof. Note that λ appears in the exponential function, and it is easy to verify that $g(\cdot, \lambda)$ is continuous with respect to (w.r.t.) λ on Ξ_1. By Lemma 1, it suffices to prove $\{f_n(\lambda)\}$ converges uniformly to $f(\lambda)$ on Ξ_1, a.s. as $n \to \infty$ and $f(\lambda)$ is continuous on Ξ_1.

By the Cauchy-Schwarz inequality, since $\forall \lambda \in \Xi_1$,

$$
\begin{aligned}
\left(\mathbb{E}_{\mathbb{P}}\left[(g(X, \lambda))^p\right]\right)^2 &= \left(\mathbb{E}_{\mathbb{P}}\left[e^{-p\langle \lambda, X_T\rangle + p\varphi_T(\lambda)} F(X)^{2p}\right]\right)^2 \\
&= e^{2p\varphi_T(\lambda)} \left(\mathbb{E}_{\mathbb{P}}\left[e^{-p\langle \lambda, X_T\rangle} F(X)^{2p}\right]\right)^2 \\
&\leq e^{2p\varphi_T(\lambda)} \mathbb{E}_{\mathbb{P}}\left[e^{-2p\langle \lambda, X_T\rangle}\right] \mathbb{E}_{\mathbb{P}}\left[F(X)^{4p}\right] < \infty.
\end{aligned}
$$

Since $g(x, \lambda) > 0$ for all $x \in \mathbb{R}^d$, $\sup_{\lambda \in \Xi_1} \mathbb{E}_{\mathbb{P}}\left[|g(X, \lambda)|^p\right] = \mathbb{E}_{\mathbb{P}}\left[(g(X, \lambda))^p\right] < \infty$. Then by Lemma 3, Chapter 6, Shiryaev [22], $g(X, \lambda)$ is uniformly integrable for all $\lambda \in \Xi_1$.

We first consider the continuity of $f(\lambda)$, $\forall \lambda_1, \lambda_2 \in \Xi_1$. Since $g(X, \lambda)$ is continuous w.r.t. λ,

$$
\begin{aligned}
\lim_{||\lambda_1 - \lambda_2|| \to 0} f(\lambda_1) - f(\lambda_2) &= \lim_{||\lambda_1 - \lambda_2|| \to 0} \mathbb{E}\left[g(X, \lambda_1) - g(X, \lambda_2)\right] \\
&= \mathbb{E}\left[\lim_{||\lambda_1 - \lambda_2|| \to 0} g(X, \lambda_1) - g(X, \lambda_2)\right] = 0,
\end{aligned}
$$

where the interchange of limit and integration is due to the uniform integrability of $g(X, \lambda)$. Therefore, $f(\lambda)$ is continuous on Ξ_1.

Now, we consider the uniform convergence of $\{f_n(x)\}$. Since

$$
\mathbb{E}[|g(X, \lambda)|] = \mathbb{E}[g(X, \lambda)] < \infty,
$$

by the uniform law of large numbers,

$$
\sup_{\lambda \in \Xi_1} \left|\frac{1}{n} \sum_{i=1}^n g(X^i, \lambda) - \mathbb{E}[g(X, \lambda)]\right| = \sup_{\lambda \in \Xi_1} |f_n(\lambda) - f(\lambda)| \to 0, a.s. \text{ as } n \to \infty.
$$

Then, $\{f_n(x)\}$ is converges uniformly to $f(\lambda)$, and the theorem is proved. $\qquad \square$

Remark 2. Theorem 1 connects the optimal values/solutions of the SAA problems with those of the original problem, also establishing that the optimal solutions for SAA are consistent estimators of the optimal solutions for the original problem.

If $f(\lambda)$ is convex, the optimal solution is unique. To find the optimal solution, we use $\nabla f(\lambda)$ and $Hess[f(\lambda)]$, which are estimated using the infinitesimal perturbation analysis (IPA) estimators $\nabla_\lambda g(X, \lambda)$ and $Hess_\lambda[g(X, \lambda)]$, respectively. The parameter λ only appears in the function $e^{-\langle\lambda, X_T\rangle + \varphi_T(\lambda)}$, so the payoff function $F(X)$ does not affect the continuity of $g(X, \lambda)$ w.r.t. λ. Define

$$\Xi_2 = \left\{\lambda \in C : \mathbb{E}_\mathbb{P}\left[e^{-p\langle\lambda, X_T\rangle}F(X)^{2p}\right] < \infty \text{ and}\right.$$
$$\left.\mathbb{E}_\mathbb{P}\left[e^{-p\langle\lambda, X_T\rangle}F(X)^{2p}\|X_T\|^{2p}\right] < \infty \text{ for } p > 1\right\}.$$

The following lemma provides conditions guaranteeing the unbiasedness of the estimators and the convexity of $f(\lambda)$. For matrices, $\|\cdot\|$ will denote the Frobenius norm.

Lemma 2. *For $\forall\lambda \in \Xi_2$, if $\|\nabla\varphi_T(\lambda)\| < \infty$ and $\|Hess[\varphi_T(\lambda)]\| < \infty$, then $\nabla f(\lambda) = \mathbb{E}[\nabla_\lambda g(X, \lambda)]$ and $Hess[f(\lambda)] = \mathbb{E}[Hess_\lambda[g(X, \lambda)]]$. Moreover, $Hess[f(\lambda)] > 0$, i.e., $f(\lambda)$ is strictly convex.*

Proof. The gradient and Hessian matrix of $g(X, \lambda)$ are given respectively by

$$\nabla_\lambda g(X, \lambda) = (\nabla\varphi_T(\lambda) - X_T)\, e^{-\langle\lambda, X_T\rangle + \varphi_T(\lambda)}F(X)^2,$$
$$Hess_\lambda[g(X, \lambda)] = Hess[\varphi_T(\lambda)]$$
$$+ (\nabla\varphi_T(\lambda) - X_T)(\nabla\varphi_T(\lambda) - X_T)'\, e^{-\langle\lambda, X_T\rangle + \varphi_T(\lambda)}F(X)^2.$$

Using the Cauchy-Schwarz inequality, since $\lambda \in \Xi_2$,

$$\mathbb{E}_\mathbb{P}\left[\left(e^{-\langle\lambda, X_T\rangle}F(X)^2\|X_T\|\right)^p\right]$$
$$\leq \left(\mathbb{E}_\mathbb{P}\left[e^{-p\langle\lambda, X_T\rangle}F(X)^{2p}\right]\mathbb{E}_\mathbb{P}\left[e^{-p\langle\lambda, X_T\rangle}F(X)^{2p}\|X_T\|^{2p}\right]\right)^{\frac{1}{2}} < \infty.$$

9

Since $\|\nabla \varphi_T(\lambda)\| < \infty$, also $|\varphi(\lambda)| < \infty$, then,

$$\sup_{\lambda \in \Xi_2} \mathbb{E}_{\mathbb{P}} \left[\left(\left(\|\nabla \varphi_T(\lambda) - X_T\| \right) e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} F(X)^2 \right)^p \right] < \infty.$$

Therefore $\nabla_\lambda g(X, \lambda)$ is uniformly integrable for each component, so $\nabla f(\lambda) = \mathbb{E}[\nabla_\lambda g(X, \lambda)]$, i.e., the IPA estimator is unbiased. Similarly, we can prove that

$$\sup_{\lambda \in \Xi_2} \mathbb{E} \left[\left(\| Hess[\varphi_T(\lambda)] \right. \right.$$
$$\left. \left. + \left(\nabla \varphi_T(\lambda) - X_T \right) \left(\nabla \varphi_T(\lambda) - X_T \right)' \| e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} F(X)^2 \right)^p \right] < \infty.$$

so $Hess[f(\lambda)] = \mathbb{E}\left[Hess_\lambda[g(X, \lambda)] \right]$.

In addition, for any $y \in \mathbb{R}^d (y \neq 0)$,

$$y Hess[f(\lambda)]y' = y\mathbb{E}_{\mathbb{P}}\left[Hess_\lambda[g(X, \lambda)] \right] y' = \mathbb{E}_{\mathbb{P}}\left[y Hess_\lambda[g(X, \lambda)] y' \right]$$
$$= \mathbb{E}_{\mathbb{P}} \left[\left(y Hess[\varphi_T(\lambda)] y' \right. \right.$$
$$\left. \left. + \langle y, \nabla \varphi_T(\lambda) - X_T \rangle^2 \right) e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} F(X)^2 \right] > 0.$$

Thus, $Hess[f(\lambda)]$ is strictly positive definite. $\qquad\square$

Remark 3. (i) $\nabla_\lambda g(X, \lambda)$ and $Hess_\lambda[g(X, \lambda)]$ are unbiased IPA estimators for $\nabla f(\lambda)$ and $Hess[f(\lambda)]$. If $\nabla_\lambda g(X, \lambda)$ and $Hess_\lambda[g(X, \lambda)]$ have simple expressions, and $\mathbb{E}_{\mathbb{P}}[\nabla_\lambda g(X, \lambda)]$ and $\mathbb{E}_{\mathbb{P}}[Hess_\lambda[g(X, \lambda)]]$ are easy to compute analytically, then we can use a deterministic gradient-based method, e.g., Newton iteration, to solve the problem directly, and do not need the SAA framework. (ii) Since $f(\lambda)$ is strictly positive definite, if there exists a local optimal solution, it is the global optimal solution, i.e., if there exists an optimal solution, it is the unique optimal solution.

Let λ^* denote the unique optimal solution of the original optimization problem (3), and λ_n^* denote the optimal solution of the SAA problem (4). Furthermore, define

$$
\begin{aligned}
\Xi_3 = \quad & \left\{ \lambda \in C : \mathbb{E}_\mathbb{P} \left[e^{-2\langle \lambda, X_T \rangle} F(X)^4 \right] < \infty \text{ and} \right. \\
& \left. \mathbb{E}_\mathbb{P} \left[e^{-2\langle \lambda, X_T \rangle} F(X)^4 \|X_T\|^2 \right] < \infty \right\}.
\end{aligned}
$$

Note that if $\lambda \in \Xi_3$, by the Cauchy-Schwarz inequality,

$$
\left(\mathbb{E}_\mathbb{P} \left[e^{-\langle \lambda, X_T \rangle} F(X)^2 \right] \right) \leq \mathbb{E}_\mathbb{P} \left[e^{-2\langle \lambda, X_T \rangle} F(X)^4 \right] < \infty.
$$

In addition, if X_T has finite variance,

$$
\begin{aligned}
& \left(\mathbb{E}_\mathbb{P} \left[e^{-\langle \lambda, X_T \rangle} F(X)^2 \|X_T\|^2 \right] \right)^2 \\
& \leq \mathbb{E}_\mathbb{P} \left[e^{-2\langle \lambda, X_T \rangle} F(X)^4 \|X_T\|^2 \right] \mathbb{E}_\mathbb{P} \left[\|X_T\|^2 \right] < \infty,
\end{aligned}
$$

which implies $\lambda \in \Xi_2$. Let $X_n = O_p(Y_n)$ denote that $\{X_n/Y_n\}$ is bounded in probability, i.e., for any $\epsilon > 0$, there exists $M > 0$ such that $\Pr\{|X_n/Y_n| > M\} < \epsilon$ for all n. The following theorem presents the convergence rate of the optimal solution.

Theorem 2. *Suppose that $\|\nabla\varphi(\lambda)\| < \infty$, $\|Hess[\varphi(\lambda)]\| < \infty$, and the following conditions hold:*

(i) $\lambda^* \in \Xi_1 \cap \Xi_2 \cap \Xi_3$.

(ii) $\mathbb{E}_\mathbb{P}\left[\|Hess_\lambda\left[g(X, \lambda)\right]\|\right] < \infty$.

Then $\|v_n^* - v^*\| = O_p(n^{-1/2})$ and $\|\lambda_n^* - \lambda^*\| = O_p(n^{-1/2})$.

Proof. Since $\mathbb{E}_\mathbb{P}[\|\nabla_\lambda g(X, \lambda)\|] < \infty$ and $\mathbb{E}_\mathbb{P}\left[\|Hess_\lambda\left[g(X, \lambda)\right]\|\right] < \infty$, then $g(X, \lambda)$ and $\nabla_\lambda g(X, \lambda)$ are Lipschitz continuous on $\Xi_1 \cap \Xi_2 \cap \Xi_3$, and $g(X, \lambda)$ is continuously differentiable w.r.t. λ on $\Xi_1 \cap \Xi_2 \cap \Xi_3$.

Consider the convergence of the optimal values. Since $\lambda \in \Xi_3$,

$$\mathbb{E}_{\mathbb{P}}\left[e^{-2\langle\lambda,X_T\rangle}F(X)^4\right] < \infty,$$

and $|\varphi_T(\lambda)| < \infty$, so $\mathbb{E}[(g(X,\lambda))^2] < \infty$. Then by Theorem 11 in Kim et al. [16], $\|v_n^* - v^*\| = O_p(n^{-1/2})$.

Now, consider the convergence of optimal solutions. Since $\lambda \in \Xi_3$ and $|\varphi_T(\lambda)| < \infty$, we also have

$$\mathbb{E}_{\mathbb{P}}\left[\|X_T\|^2 e^{-2\langle\lambda,X_T\rangle+2\varphi_T(\lambda)}F(X)^4\right] < \infty,$$

then by the Cauchy-Schwarz inequality

$$\left(\mathbb{E}_{\mathbb{P}}\left[\|X_T\|e^{-2\langle\lambda,X_T\rangle+2\varphi_T(\lambda)}F(X)^4\right]\right)^2$$
$$\leq \mathbb{E}_{\mathbb{P}}\left[\|X_T\|^2 e^{-2\langle\lambda,X_T\rangle+2\varphi_T(\lambda)}F(X)^4\right]\mathbb{E}_{\mathbb{P}}\left[e^{-2\langle\lambda,X_T\rangle}F(X)^4\right] < \infty.$$

By the condition $\|\nabla\varphi_T(\lambda)\| < \infty$, therefore,

$$\mathbb{E}_{\mathbb{P}}\left[\|\nabla_\lambda g(X,\lambda)\|^2\right] = \mathbb{E}_{\mathbb{P}}\left[\left\|\left(\nabla\varphi_T\left(\lambda\right) - X_T\right)e^{-\langle\lambda,X_T\rangle+\varphi_T(\lambda)}F(X)^2\right\|^2\right]$$
$$= \mathbb{E}_{\mathbb{P}}\left[\|\nabla\varphi_T\left(\lambda\right) - X_T\|^2 e^{-2\langle\lambda,X_T\rangle+2\varphi_T(\lambda)}F(X)^4\right] < \infty.$$

By condition (ii), $\mathbb{E}_{\mathbb{P}}\left[\|Hess_\lambda\left[g(X,\lambda)\right]\|\right] < \infty$, and by Lemma 2, $Hess[f(\lambda)]$ is strictly positive definite. Then by Theorem 12 in Kim et al. [16], $\|\lambda_n^* - \lambda^*\| = O_p(n^{-1/2})$.   □

Let $B(x,r) = \{y : \|y - x\| \leq r\}$ denote the closed ball of radius r around x. We will prove that the SAA problem can be solved by the following Newton iteration procedure:

$$\lambda_{n,k+1} = \lambda_{n,k} + \Delta\lambda_{n,k},$$
$$\text{with} \quad Hess[f_n(\lambda_{n,k})]\Delta\lambda_{n,k} = -\nabla f_n(\lambda_{n,k}). \tag{5}$$

Theorem 3. *Suppose $\|\nabla\varphi(\lambda)\| < \infty$, $\|Hess[\varphi(\lambda)]\| < \infty$, and $Hess^{-1}[f_n(\lambda_n^*)]$ exists a.s. for $\lambda_n^* \in \Xi_1 \cap \Xi_2 \cap \Xi_3$. If there exist positive constants K, L and R, such that $\|Hess[f_n(\lambda)]\| \le K$ a.s., and for $\forall \lambda_1, \lambda_2 \in B(\lambda_n^*, R)$,*

$$\|Hess[f_n(\lambda_1)] - Hess[f_n(\lambda_2)]\| \le L\|\lambda_1 - \lambda_2\| \quad a.s. \tag{6}$$

Then there exists $r > 0$ such that for any $\lambda_{n,0} \in B(\lambda_n^*, r)$, the sequence $\{\lambda_{n,k}\}$ derived from the iteration procedure (5) converges to λ_n^* a.s. as $k \to \infty$.

Proof. We first prove that $\Xi_1 \cap \Xi_2 \cap \Xi_3$ is a convex set. Consider Ξ_1. For any $\lambda_1, \lambda_2 \in \Xi_1$ and $p + q = 1, p, q > 0$, by the Hölder inequality,

$$\mathbb{E}_{\mathbb{P}}\left[e^{-2\langle p\lambda_1 + q\lambda_2, X_t\rangle}\right] \le \mathbb{E}_{\mathbb{P}}\left[e^{-2\langle\lambda_1, X_t\rangle}\right]^p \mathbb{E}_{\mathbb{P}}\left[e^{-2\langle\lambda_2, X_t\rangle}\right]^q < \infty.$$

Therefore Ξ_1 is convex. Similarly, Ξ_2 and Ξ_3 are both convex, and $\Xi_1 \cap \Xi_2 \cap \Xi_3$ is convex.

Then, we consider the convergence of Newton iteration via induction. For any $\lambda_{n,0} \in B(\lambda^*, r)$, by Equation (6), letting $r = \min(R, 1/(2CL))$, we have

$$\begin{aligned}
&\|Hess^{-1}[f_n(\lambda_n^*)]\left(Hess[f_n(\lambda_{n,0})] - Hess[f_n(\lambda_n^*)]\right)\| \\
&\le \|Hess^{-1}[f_n(\lambda_n^*)]\|\|\left(Hess[f_n(\lambda_{n,0})] - Hess[f_n(\lambda_n^*)]\right)\| \\
&\le CLr \le \frac{1}{2} \; a.s.
\end{aligned} \tag{7}$$

Then,

$$\begin{aligned}
&\|Hess^{-1}[f_n(\lambda_{n,0})]\| \\
&\le \frac{\|Hess^{-1}[f_n(\lambda_n^*)]\|}{1 - \|Hess^{-1}[f_n(\lambda_n^*)]\left(Hess[f_n(\lambda_{n,0})] - Hess[f_n(\lambda_n^*)]\right)\|} \le 2C \; a.s.,
\end{aligned} \tag{8}$$

and $Hess^{-1}[f_n(\lambda_{n,0})]$ exists a.s.

Suppose that $\lambda_{n,k} \in B(\lambda_n^*, r)$. Similar to Equations (7) and (8), we know that $Hess^{-1}[f_n(\lambda_{n,k})]$ exists a.s. and $\|Hess^{-1}[f_n(\lambda_{n,k})]\| \le 2C$ a.s. Now we prove $\lambda_{n,k+1} \in B(\lambda_n^*, r)$. By (5),

$$\lambda_{n,k+1} - \lambda_n^* = \lambda_{n,k} - \lambda_n^* - Hess^{-1}[f_n(\lambda_{n,k})] \left(\nabla f_n(\lambda_{n,k}) - \nabla f_n(\lambda_n^*)\right) \ a.s.,$$

so

$$
\begin{aligned}
&\|\lambda_{n,k+1} - \lambda_n^*\| \\
&\le \|Hess^{-1}[f_n(\lambda_{n,k})]\| \|\nabla f_n(\lambda_{n,k}) - \nabla f_n(\lambda_n^*) - Hess[f_n(\lambda_{n,k})](\lambda_{n,k} - \lambda_n^*)\| \\
&\le 2C\frac{L}{2}\|\lambda_n^* - \lambda_{n,k}\|^2 \ a.s.
\end{aligned}
$$

Since $\lambda_{n,k} \in B(\lambda_n^*, r)$, $\|\lambda_n^* - \lambda_{n,k}\| \le r \le 1/(2CL)$ a.s., then

$$\|\lambda_{n,k+1} - \lambda_n^*\| \le \frac{1}{2}\|\lambda_{n,k} - \lambda_n^*\| \ a.s.$$

Therefore $\lambda_{n,k+1} \in B(\lambda_n^*, r)$, and $\{\lambda_{n,k}\}$ converges to λ_n^* a.s. as $k \to \infty$. $\qquad\square$

Remark 4. (i) By Theorem 1, we know that when $n \to \infty$, $\lambda_n^* \to \lambda^*$; therefore, the sequence $\{\lambda_{n,k}\}$ converges to λ^* a.s. as $n \to \infty$ and $k \to \infty$. In practice, since Newton iteration converges quickly, we allocate a much larger share of the computation budget to the approximation samples of SAA than to the iterations, i.e., $n \gg k$. (ii) Suppose that $\mathrm{Var}(F(X)) < \infty$, and note that when $\lambda = 0$, the probability measure is not changed, so $0 \in C$. By the convexity of $f(\lambda)$, we know that there must exist $\lambda^\star \in \Xi_1 \cap \Xi_2$ such that $\mathbb{E}_{\mathbb{P}}[f(\lambda^\star)] \le \mathbb{E}_{\mathbb{P}}[f(0)]$. In other words, even if the optimal solution is not contained in the set, we can still find other points $\lambda^\star$ that reduce the variance under the new probability measure.

3.2. Algorithm for the SAA-Newton method and importance sampling

For the SAA-Newton method, $\nabla f_n(\lambda)$ and $Hess\left[f_n(\lambda)\right]$ can be obtained by

$$\nabla f_n(\lambda) = \frac{1}{n} \sum_{j=1}^{n} \left[\left(\nabla \varphi_T(\lambda) - X_T^j \right) e^{-\left\langle \lambda, X_T^j \right\rangle + \varphi_T(\lambda)} F\left(X^j\right)^2 \right],$$

$$Hess\left[f_n(\lambda)\right] = \frac{1}{n} \sum_{j=1}^{n} \left[\left(Hess\left[\varphi_T(\lambda)\right] \right. \right.$$

$$\left. \left. + \left(\nabla \varphi_T(\lambda) - X_T^j \right) \left(\nabla \varphi_T(\lambda) - X_T^j \right)' \right) e^{-\left\langle \lambda, X_t^j \right\rangle + \varphi_T(\lambda)} F\left(X^j\right)^2 \right].$$

Then Algorithm 1 finds the optimal parameter in the Esscher measure change. After obtaining the estimated λ_n^*, we can change the probability measure using the

---

Algorithm 1 SAA-Newton method to find the optimal parameter

Input: number of samples n in SAA; maturity T; Lévy processes parameter Λ under the **original** probability measure; termination tolerance of Newton iteration ρ.

Initialization: initial point λ_0; $k = 0$.

1: generate and store Lévy process paths $\{X_t^1, 0 \le t \le T\}, \{X_t^2, 0 \le t \le T\}, \ldots \{X_t^n, 0 \le t \le T\}$ under the **original** probability measure;

2: compute payoffs $\{F(X^i), i = 1, 2, \ldots, n\}$;

3: **repeat**

4:      compute $\nabla f_n(\lambda_k)$ and $Hess[f_n(\lambda_k)]$;

5:      solve $Hess[f(\lambda_k)]\Delta\lambda_k = -\nabla f(\lambda_k)$;

6:      set $\lambda_{k+1} = \lambda_k + \Delta\lambda_k$;

7:      set $k = k + 1$;

8: **until** $(\|\lambda_{k+1} - \lambda_k\| \le \rho)$

Output: estimated optimal parameter $\lambda_n^* = \lambda_{k+1}$.

---

following relationship between the characteristic functions of the original probability measure $\mathbb{P}$ and the new probability measure $\mathbb{P}_\lambda$ (Section 6.2.2, Schoutens [2]):

$$\Phi^\lambda_{X_T}(u) = \frac{\Phi_{X_T}(u - i\lambda)}{\Phi_{X_T}(-i\lambda)}, \tag{9}$$

where $\Phi^\lambda_{X_T}$ is the characteristic function under the new probability measure $\mathbb{P}_\lambda$.

In Brownian motion, a drift change is the most frequently used method for changing the probability measure; see Glasserman [4], Glasserman et al. [21] and Su and Fu [12]. In our framework, if the original process parameters are (μ, σ) in Brownian motion, after the Esscher measure change, we can obtain the characteristic function of the new process, then obtain the new process parameters $(\mu + \lambda^*, \sigma)$, which indicates the drift change is a special case of the Esscher measure change in Brownian motion. Similarly, for other processes, we can obtain the new process parameters by the same procedure, i.e., by Equation (9), the characteristic function of X_T under the new probability measure is obtained, then we can compute the process parameters under the new probability measure. For example, if the process parameters of the VG process are (σ, ν, θ) under the original probability measure, and the characteristic function is given by

$$\Phi_{X_T^{VG}}(u) = \left(1 - iu\theta\nu + \frac{1}{2}\sigma^2\nu u^2\right)^{-T/\nu},$$

then by Equation (9), the new characteristic function is given by

$$\Phi^\lambda_{X_T^{VG}}(u) = \left(1 - \frac{iu\nu(\theta + \lambda\sigma^2)}{1 - \lambda\theta\nu - 1/2\sigma^2\nu\lambda^2} + \frac{1}{2}\frac{\sigma^2}{1 - \lambda\theta\nu - 1/2\sigma^2\nu\lambda^2}\nu u^2\right)^{-T/\nu},$$

and the new process parameters for the Esscher measure change are

$$\left(\sigma/\sqrt{1 - \lambda\theta\nu - 1/2\sigma^2\nu\lambda^2}, \nu, (\theta + \lambda\sigma^2)/(1 - \lambda\theta\nu - 1/2\sigma^2\nu\lambda^2)\right).$$

Finally, we give the algorithm for importance sampling using the Esscher measure change.

Algorithm 2 Importance sampling by Esscher measure change

Input: number of simulation samples N in pricing; maturity T; Lévy processes parameter Λ under the **new** probability measure with the optimal parameter λ_n^*.

1: generate and store Lévy process paths $\{X_t^1, 0 \le t \le T\}$, $\{X_t^2, 0 \le t \le T\}$, ..., $\{X_t^N, 0 \le t \le T\}$ under the **new** probability measure;

2: compute $\left\{V^i = F\left(X^i\right) e^{-\left\langle \lambda_n^*, X_T^i \right\rangle + \varphi_T(\lambda_n^*)}, i = 1, 2, \ldots, N\right\}$.

Output: estimated payoff $\hat{V} = 1/N \sum_{i=1}^{N} V^i$.

3.3. Verifying the conditions

Verifying the conditions in the theorems requires finding the sets Ξ_1, Ξ_2, Ξ_3, which are difficult to specify directly. However, we demonstrate how to easily obtain a subset $\tilde{C} \subseteq \Xi_1 \cap \Xi_2 \cap \Xi_3$, and by Remark 4ii, even if the optimal solution is not in $\tilde{C}$, Algorithm 1 will still find a $\lambda^\star$ to reduce the variance, when the search for the optimal parameters is restricted to $\tilde{C}$. For the NIG example, the density function is given by

$$f(x; \alpha, \beta, \delta) = \frac{\alpha \delta}{\pi} \exp\left(\delta \sqrt{\alpha^2 - \beta^2} + \beta x\right) \frac{K_1\left(\alpha \sqrt{\delta^2 + x^2}\right)}{\sqrt{\delta^2 + x^2}},$$

where $K_n(\cdot)$ is a modified Bessel function of the third kind with index n. Consider $\mathbb{E}_{\mathbb{P}}\left[\exp\left(-2\langle \lambda, X' \rangle\right) F(X')^4 \|X'\|^2\right]$. First, it is easy to verify that $\mathbb{E}[X_i'^4] < \infty$, where X_i' is the ith component for the random vector X'. By the Cauchy-Schwartz inequality, it suffices to find the set $\tilde{C}$ to ensure $\mathbb{E}_{\mathbb{P}}\left[\exp\left(-4\langle \lambda, X' \rangle\right) F(X')^8\right] <$

∞ for all $\lambda \in \tilde{C}$, by Equation (3), which is equivalent to verifying

$$\int_{\mathbb{R}} e^{-4\lambda_i x_i} e^{8x_i} f(x; \alpha, \beta, \delta) dx = \int_{\mathbb{R}} f(x; \alpha, \beta + 8 - 4\lambda_i, \delta) dx < \infty.$$

So $\tilde{C} = (-(\alpha + \beta)/4 + 2, (\alpha - \beta)/4 + 2)^M$, where the superscript denotes the dimension. It is easy to verify that $\tilde{C} \subseteq \Xi_1 \cap \Xi_2 \cap \Xi_3$.

For the VG example, since $F(X) \leq 1$, we can calculate $\tilde{C} = (-\theta/\sigma^2 - \sqrt{2\sigma^2 + \theta^2\nu/(\sigma^4\nu)}, -\theta/\sigma^2 + \sqrt{2\sigma^2 + \theta^2\nu/(\sigma^4\nu)})^M$. For numerical examples, refer to Jiang et al. [23].

4. Conclusion

In this paper, we propose a new method, the SAA-Newton method, to find the optimal importance sampling parameters based on the Esscher measure change for Lévy processes. The optimization problem of minimizing the variance under the new probability measure is formulated in the SAA framework, for which there exists a unique optimal solution (parameter) under some mild conditions. Furthermore, the corresponding deterministic problem can be solved by Newton iteration. Theoretical proofs, such as the relationship between the SAA solutions and the true optimum and the convergence rate to the optimum, are provided.

References

[1] R. Cont, Empirical properties of asset returns: stylized facts and statistical issues, Quant. Finance. 1 (2001) 223–236.

[2] W. Schoutens, Lévy Processes in Finance: Pricing Financial Derivatives, John Wiley Sons, Chichester, 2003.

[3] P. Carr, D. B. Madan, Representing the CGMY and Meixner Lévy processes as time changed Brownian motions, J. Comput. Finance. 2 (2008) 27–47.

[4] P. Glasserman, Monte Carlo Methods in Financial Engineering, Springer, New York, 2003.

[5] M. C. Fu, Variance-Gamma and Monte Carlo, in: M. C. Fu, R. A. Jarrow, J. Y. Yen, R. J. Elliott (Eds.), Advances in Mathematical Finance, Birkhäuser, Boston, 21–34, 2007.

[6] P. Glasserman, Z. J. Liu, Sensitivity estimates from characteristic functions, Oper. Res. 58 (2010) 1611–1623.

[7] R. Kawai, Likelihood ratio gradient estimation for Meixner distribution and Lévy processes, Comput. Stat. 27 (2012) 739–755.

[8] R. Kawai, Optimal importance sampling parameter search for Lévy processes via stochastic approximation, SIAM J. Numer. Anal. 47 (2008) 293–307.

[9] K. D. Dingeç, W. Hörmann, A general control variate method for option pricing under Lévy processes, Eur. J. Oper. Res. 221 (2012) 368–377.

[10] R. Kawai, Asymptotically optimal allocation of stratified sampling with adaptive variance reduction by strata, ACM Trans. Model Comput. Simul. 20 (2010) 1–17.

[11] T. Homem-de-Mello, A study on the cross-entropy method for rare-event probability estimation, INFORMS J. Comput. 19 (2007) 381–394.

[12] Y. Su, M. C. Fu, Optimal importance sampling in securities pricing, J. of Comput. Finance. 5 (2002) 27–50.

[13] B. Lapeyre, J. Lelong, A framework for adaptive Monte Carlo procedures, Monte Carlo Methods Appl. 17 (2011) 77–98.

[14] R. Kawai, An importance sampling method based on the density transformation of Lévy processes, Monte Carlo Methods Appl. 12 (2007) 171–186.

[15] A. Shapiro, Monte Carlo sampling methods, in: A. Ruszczynski, A. Shapiro (Eds.), Handbooks in Operations Research and Management Science, vol. 10, chap. 6, Elsevier, New York, 353–425, 2003.

[16] S. Kim, R. Pasupathy, S. G. Henderson, A guide to sample-average approximation, in: M. C. Fu (Ed.), Handbook of Simulation Optimization, chap. 8, Springer, New York, 207–241, 2014.

[17] A. Nemirovski, A. Juditsky, G. Lan, A. Shapiro, Robust stochastic approximation approach to stochastic programming, SIAM J. Optimiz. 19 (2009) 1574–1609.

[18] M. C. Fu, Gradient estimation, in: S. G. Henderson, B. L. Nelson (Eds.), Handbooks in Operations Research and Management Science, vol. 13, chap. 19, Springer, Amsterdam, 575–616, 2006.

[19] B. Jourdain, J. Lelong, Robust adaptive importance sampling for normal random vectors, Ann. Appl. Probab. 19 (2009) 1687–1718.

[20] L. Badouraly-Kassim, J. Lelong, I. Loumrhari, Importance sampling for jump processes and applications to finance, J. Comput. Finance. In press.

[21] P. Glasserman, P. Heidelberger, P. Shahabuddin, Asymptotic optimal importance sampling and stratification for pricing path-dependent options, Math. Finance. 9 (1999) 117–152.

[22] A. N. Shiryaev, Probability, Springer, Berlin, 1995.

[23] G. Jiang, M. C. Fu, C. Xu, Optimal importance sampling for simulation of Lévy processes, submitted to Winter Simulation Conference 2015, 2015.