



# On sample average approximation algorithms for determining the optimal importance sampling parameters in pricing financial derivatives on Lévy processes



Guangxin Jiang<sup>a,\*</sup>, Chenglong Xu<sup>c</sup>, Michael C. Fu<sup>b</sup>

<sup>a</sup> Department of Economics and Finance, City University of Hong Kong, Kowloon, Hong Kong

<sup>b</sup> Robert H. Smith School of Business and Institute for Systems Research, University of Maryland, College Park, MD, 20742, USA

<sup>c</sup> Department of Mathematics, Tongji University, Shanghai, 200092, China

## ARTICLE INFO

### Article history:

Received 25 April 2015

Received in revised form

11 November 2015

Accepted 11 November 2015

Available online 22 November 2015

### Keywords:

Importance sampling

Sample average approximation

Newton iteration

Infinitesimal perturbation analysis

Lévy processes

## 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, are provided. A numerical example illustrates the effectiveness of the approach.

© 2015 Elsevier B.V. All rights reserved.

## 1. Introduction

Lévy processes can model critical characteristics observed in financial markets, such as heavy tails and jumps; see Cont [3] and Schoutens [20] for a survey. Due to the complicated nature of Lévy process models, Monte Carlo simulation is routinely used to price the derivatives [2,7] and estimate Greeks [6,9,16]. Although Monte Carlo simulation can solve high-dimensional problems, variance reduction techniques are often needed to improve the computational efficiency; see [7] for a survey. For Lévy processes models, variance reduction techniques used in derivatives pricing include importance sampling [14], control variates [4], stratified sampling [15], 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. [10]). 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 [23] used stochastic approximation (SA) to solve the resulting stochastic optimization problem in the geometric Brownian motion setting (also refer to Lapeyre and Lelong [18]), and Kawai [13] and Kawai [14] 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) [21,17]. SAA combined with a good deterministic algorithm has proven effective for solving certain classes of stochastic optimization problems where the classical SA numerical procedures perform poorly [19]. 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 [5] for survey. Jourdain and Lelong [12] and Badouraly-Kassim et al. [1] 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.

\* Corresponding author.

E-mail address: [guajiang@cityu.edu.hk](mailto:guajiang@cityu.edu.hk) (G. Jiang).

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. In Section 4, we provide a numerical example, an Asian call option driven by a normal inverse Gaussian (NIG) process, to show the effectiveness of the algorithms in finding the optimal parameter and reducing the variance.

## 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)  $X_0 = \mathbf{0}$  a.s.
- (2)  $\mathbf{X}$  has independent and stationary increments.
- (3)  $\mathbf{X}$  is stochastically continuous, i.e.,  $\lim_{t \rightarrow s} \mathbb{P}(\|\mathbf{X}_t - \mathbf{X}_s\| > a) = 0$ , 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

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

where  $u \in \mathbb{R}^d$ ,  $b \in \mathbb{R}^d$ ,  $A$  is a non-negative definite symmetric  $d \times d$  matrix, and  $\nu$  is a Lévy measure on  $\mathbb{R}^d - \{0\}$ . A measure  $\nu$  is a Lévy measure if it satisfies  $\int_{\mathbb{R}^d - \{0\}} (|y|^2 \wedge 1) \nu(dy) < \infty$ . Let  $\Lambda$  denote the process parameters, e.g.,  $\Lambda = (\alpha, \beta, \delta)$  for the NIG 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  $\mathbb{E}_{\mathbb{P}}[F(X)]$ .

### 2.2. Importance sampling and 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 [8,23]. For Lévy processes, the Esscher measure change is commonly used [14].

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

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

where  $\lambda \in C \subseteq \mathbb{R}^d$  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

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

where  $\mathcal{F}_t$  is the natural filtration of  $\{X_t, t \geq 0\}$ . Note that  $\lambda = \mathbf{0}$  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  $\mathbb{E}_{\mathbb{P}} [F(X)]$ ,

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

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

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

We call  $\lambda^*$  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 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

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

where

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

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

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

In SAA, we generate independent and identically distributed (i.i.d.) paths of  $X$  denoted by  $X^1, X^2, \dots, X^n$ , 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  $\pi^*$  and  $\Pi_n^*$  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  $\mathbb{D}(\pi^*, \Pi_n^*) \rightarrow 0$  and  $v_n^* \rightarrow v^*$  as  $n \rightarrow \infty$  under some mild conditions.

**Lemma 1** (Theorem 9, Kim et al. [17]). *Suppose that there exists a compact subset  $C \subset \mathbb{R}^d$  such that:*

- (i)  $\pi^*$  is non-empty and contained in  $C$ ,
- (ii)  $\{f_n(x)\}$  converges uniformly to  $f(x)$  on  $C$  a.s. as  $n \rightarrow \infty$ , and
- (iii) for sufficiently large  $n$ ,  $\Pi_n^*$  is non-empty and contained in  $C$  a.s.

Then  $v_n^* \rightarrow v^*$  a.s. as  $n \rightarrow \infty$ . Furthermore, if  $f$  is continuous on  $C$ , then  $\mathbb{D}(\pi^*, \Pi_n^*) \rightarrow 0$  a.s. as  $n \rightarrow \infty$ .

Let  $\mathcal{E}_1 = \left\{ \lambda \in C : \mathbb{E}_{\mathbb{P}} [e^{-2p(\lambda, X_T)}] < \infty \text{ and } \mathbb{E}_{\mathbb{P}} [F(X)^{4p}] < \infty \text{ for } p > 1 \right\}$ , then we have the following theorem.

**Theorem 1.** *If  $\pi^*$  and  $\Pi_n^*$  are non-empty and contained in  $\mathcal{E}_1$ ,  $\lambda \in \mathcal{E}_1$ , then  $v_n^* \rightarrow v^*$  and  $\mathbb{D}(\pi^*, \Pi_n^*) \rightarrow 0$  a.s. as  $n \rightarrow \infty$ .*

**Proof.** Note that  $\lambda$  appears in the exponential function, and it is easy to verify that  $g(\cdot, \lambda)$  is continuous with respect to (w.r.t.)  $\lambda$  on  $\mathcal{E}_1$ . By Lemma 1, it suffices to prove  $\{f_n(\lambda)\}$  converges uniformly to  $f(\lambda)$  on  $\mathcal{E}_1$ , a.s. as  $n \rightarrow \infty$  and  $f(\lambda)$  is continuous on  $\mathcal{E}_1$ .

By the Cauchy–Schwarz inequality, since  $\forall \lambda \in \mathcal{E}_1$ ,

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

Since  $g(x, \lambda) > 0$  for all  $x \in \mathbb{R}^d$ ,  $\sup_{\lambda \in \mathcal{E}_1} \mathbb{E}_{\mathbb{P}} [g(X, \lambda)^p] = \mathbb{E}_{\mathbb{P}} [(g(X, \lambda))^p] < \infty$ . Then by Lemma 3, Chapter 6, Shiryaev [22],  $g(X, \lambda)$  is uniformly integrable for all  $\lambda \in \mathcal{E}_1$ .

We first consider the continuity of  $f(\lambda)$ ,  $\forall \lambda_1, \lambda_2 \in \mathcal{E}_1$ . Since  $g(X, \lambda)$  is continuous w.r.t.  $\lambda$ ,

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

where the interchange of limit and integration is due to the uniform integrability of  $g(X, \lambda)$ . Therefore,  $f(\lambda)$  is continuous on  $\mathcal{E}_1$ .

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,

$$\begin{aligned} \sup_{\lambda \in \mathcal{E}_1} \left| \frac{1}{n} \sum_{i=1}^n g(X^i, \lambda) - \mathbb{E}[g(X, \lambda)] \right| &= \sup_{\lambda \in \mathcal{E}_1} |f_n(\lambda) - f(\lambda)| \rightarrow 0, \\ \text{a.s. as } n \rightarrow \infty. \end{aligned}$$

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

**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  $\lambda$  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.  $\lambda$ . Define

$$\begin{aligned} \mathcal{E}_2 &= \left\{ \lambda \in C : \mathbb{E}_{\mathbb{P}} [e^{-p(\lambda, X_T)} F(X)^{2p}] < \infty \text{ and} \right. \\ &\quad \left. \mathbb{E}_{\mathbb{P}} [e^{-p(\lambda, X_T)} F(X)^{2p} \|X_T\|^{2p}] < \infty \text{ for } p > 1 \right\}. \end{aligned}$$

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  $\forall \lambda \in \mathcal{E}_2$ , 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

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

Using the Cauchy–Schwarz inequality, since  $\lambda \in \mathcal{E}_2$ ,

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

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

$$\sup_{\lambda \in \mathcal{E}_2} \mathbb{E}_{\mathbb{P}} \left[ ((\|\nabla \varphi_T(\lambda) - X_T\|) e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} F(X)^2)^p \right] < \infty.$$

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

$$\begin{aligned} \sup_{\lambda \in \mathcal{E}_2} \mathbb{E} \left[ (\|Hess[\varphi_T(\lambda)] + (\nabla \varphi_T(\lambda) - X_T) (\nabla \varphi_T(\lambda) - X_T)'\| \right. \\ \left. \times e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} F(X)^2)^p \right] < \infty \end{aligned}$$

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

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

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

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

**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  $\lambda^*$  denote the unique optimal solution of the original optimization problem (3), and  $\lambda_n^*$  denote the optimal solution of the SAA problem (4). Furthermore, define

$$\mathcal{E}_3 = \left\{ \lambda \in C : \mathbb{E}_{\mathbb{P}} \left[ e^{-2(\lambda, X_T)} F(X)^4 \right] < \infty \text{ and } \mathbb{E}_{\mathbb{P}} \left[ e^{-2(\lambda, X_T)} F(X)^4 \|X_T\|^2 \right] < \infty \right\}.$$

Note that if  $\lambda \in \mathcal{E}_3$ , by the Cauchy–Schwarz inequality,

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

In addition, if  $X_T$  has finite variance,

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

which implies  $\lambda \in \mathcal{E}_2$ . 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  $\|\nabla\varphi(\lambda)\| < \infty$ ,  $\|\text{Hess}[\varphi(\lambda)]\| < \infty$ , and the following conditions hold:

- (i)  $\lambda^* \in \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ .
- (ii)  $\mathbb{E}_{\mathbb{P}} [\|\text{Hess}_{\lambda} [g(X, \lambda)]\|] < \infty$ .

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  $\mathbb{E}_{\mathbb{P}} [\|\text{Hess}_{\lambda} [g(X, \lambda)]\|] < \infty$ , then  $g(X, \lambda)$  and  $\nabla_{\lambda} g(X, \lambda)$  are Lipschitz continuous on  $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ , and  $g(X, \lambda)$  is continuously differentiable w.r.t.  $\lambda$  on  $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ .

Consider the convergence of the optimal values. Since  $\lambda \in \mathcal{E}_3$ ,

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

and  $|\varphi_T(\lambda)| < \infty$ , so  $\mathbb{E}[(g(X, \lambda))^2] < \infty$ . Then by Theorem 11 in [17],  $\|v_n^* - v^*\| = O_p(n^{-1/2})$ .

Now, consider the convergence of optimal solutions. Since  $\lambda \in \mathcal{E}_3$  and  $|\varphi_T(\lambda)| < \infty$ , we also have

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

then by the Cauchy–Schwarz inequality

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

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

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

By condition (ii),  $\mathbb{E}_{\mathbb{P}} [\|\text{Hess}_{\lambda} [g(X, \lambda)]\|] < \infty$ , and by Lemma 2,  $\text{Hess}[f(\lambda)]$  is strictly positive definite. Then by Theorem 12 in [17],  $\|\lambda_n^* - \lambda^*\| = O_p(n^{-1/2})$ .  $\square$

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:

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

**Theorem 3.** Suppose  $\|\nabla\varphi(\lambda)\| < \infty$ ,  $\|\text{Hess}[\varphi(\lambda)]\| < \infty$ , and  $\text{Hess}^{-1}[f_n(\lambda_n^*)]$  exists a.s. for  $\lambda_n^* \in \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ . If there exist positive constants  $K, L$  and  $R$ , such that  $\|\text{Hess}[f_n(\lambda)]\| \leq K$  a.s., and for  $\forall \lambda_1, \lambda_2 \in B(\lambda_n^*, R)$ ,

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

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  $\lambda_n^*$  a.s. as  $k \rightarrow \infty$ .

**Proof.** We first prove that  $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$  is a convex set. Consider  $\mathcal{E}_1$ . For any  $\lambda_1, \lambda_2 \in \mathcal{E}_1$  and  $p + q = 1, p, q > 0$ , by the Hölder inequality,

$$\mathbb{E}_{\mathbb{P}} \left[ e^{-2(p\lambda_1 + q\lambda_2, X_T)} \right] \leq \mathbb{E}_{\mathbb{P}} \left[ e^{-2(\lambda_1, X_T)} \right]^p \mathbb{E}_{\mathbb{P}} \left[ e^{-2(\lambda_2, X_T)} \right]^q < \infty.$$

Therefore  $\mathcal{E}_1$  is convex. Similarly,  $\mathcal{E}_2$  and  $\mathcal{E}_3$  are both convex, and  $\mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$  is convex.

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

$$\begin{aligned} & \|\text{Hess}^{-1}[f_n(\lambda_n^*)] (\text{Hess}[f_n(\lambda_{n,0})] - \text{Hess}[f_n(\lambda_n^*)])\| \\ & \leq \|\text{Hess}^{-1}[f_n(\lambda_n^*)]\| \|\text{Hess}[f_n(\lambda_{n,0})] - \text{Hess}[f_n(\lambda_n^*)]\| \\ & \leq CLr \leq \frac{1}{2} \quad \text{a.s.} \end{aligned} \quad (7)$$

Then,

$$\begin{aligned} & \|\text{Hess}^{-1}[f_n(\lambda_{n,0})]\| \\ & \leq \frac{\|\text{Hess}^{-1}[f_n(\lambda_n^*)]\|}{1 - \|\text{Hess}^{-1}[f_n(\lambda_n^*)] (\text{Hess}[f_n(\lambda_{n,0})] - \text{Hess}[f_n(\lambda_n^*)])\|} \\ & \leq 2C \quad \text{a.s.}, \end{aligned} \quad (8)$$

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

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

$$\begin{aligned} \lambda_{n,k+1} - \lambda_n^* &= \lambda_{n,k} - \lambda_n^* - \text{Hess}^{-1}[f_n(\lambda_{n,k})] \\ & \quad \times (\nabla f_n(\lambda_{n,k}) - \nabla f_n(\lambda_n^*)) \quad \text{a.s.}, \end{aligned}$$

so

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

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

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

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

**Remark 4.** (i) By Theorem 1, we know that when  $n \rightarrow \infty$ ,  $\lambda_n^* \rightarrow \lambda^*$ ; therefore, the sequence  $\{\lambda_{n,k}\}$  converges to  $\lambda^*$  a.s. as  $n \rightarrow \infty$  and  $k \rightarrow \infty$ . 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  $\text{Var}(F(X)) < \infty$ , 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  $\lambda^* \in \mathcal{E}_1 \cap \mathcal{E}_2$  such that  $\mathbb{E}_{\mathbb{P}}[f(\lambda^*)] \leq \mathbb{E}_{\mathbb{P}}[f(0)]$ . In other words, even if the optimal solution is not contained in the set, we can still find other points  $\lambda^*$  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  $Hess[f_n(\lambda)]$  can be obtained by

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

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

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

Then Algorithm 1 finds the optimal parameter in the Esscher measure change.

**Algorithm 1** SAA–Newton method to find the optimal parameter

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

**Initialization:** initial point  $\lambda_0$ ;  $k = 0$ .

- 1: generate and store Lévy process paths  $\{X_t^1, 0 \leq t \leq T\}$ ,  $\{X_t^2, 0 \leq t \leq T\}$ ,  $\dots$ ,  $\{X_t^n, 0 \leq t \leq T\}$  under the **original** probability measure;
- 2: compute payoffs  $\{F(X^i), i = 1, 2, \dots, n\}$ ;
- 3: **repeat**
- 4:   compute  $\nabla f_n(\lambda_k)$  and  $Hess[f_n(\lambda_k)]$ ;
- 5:   solve  $Hess[f_n(\lambda_k)] \Delta \lambda_k = -\nabla f_n(\lambda_k)$ ;
- 6:   set  $\lambda_{k+1} = \lambda_k + \Delta \lambda_k$ ;
- 7:   set  $k = k + 1$ ;
- 8: **until**  $(\|\lambda_{k+1} - \lambda_k\| / \|\lambda_{k+1}\| \leq \rho)$

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

After obtaining the estimated  $\lambda_n^*$ , we can change the probability measure using the 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 [20]):

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

where  $\Phi_{X_T}^\lambda$  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 [8,7,23]. In our framework, if the original process parameters are  $(\mu, \sigma)$  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 Eq. (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 NIG process are  $(\alpha, \beta, \delta)$  under the original probability measure, and the characteristic function is given by

$$\Phi_{X_t}^{NIG}(u) = \exp \left( -\delta t \left( \sqrt{\alpha^2 - (\beta + iu)^2} - \sqrt{\alpha^2 - \beta^2} \right) \right),$$

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

$$\Phi_{X_t}^{NIG}(u) = \exp \left( -\delta t \left( \sqrt{\alpha^2 - (\beta + \lambda^* + iu)^2} - \sqrt{\alpha^2 - (\beta + \lambda^*)^2} \right) \right),$$

and the new process parameters for the Esscher measure change are  $(\alpha, \beta + \lambda^*, \delta)$ .

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  $\Lambda$  under the **new** probability measure with parameter  $\lambda^*$ .

- 1: generate and store Lévy process paths  $\{X_t^1, 0 \leq t \leq T\}$ ,  $\{X_t^2, 0 \leq t \leq T\}$ ,  $\dots$ ,  $\{X_t^N, 0 \leq t \leq T\}$  under the **new** probability measure;
- 2: compute  $\left\{ V^i = F(X^i) e^{-\langle \lambda^*, X_T^i \rangle + \varphi_T(\lambda^*)}, i = 1, 2, \dots, N \right\}$ .

**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  $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$ , which are difficult to specify directly. However, we demonstrate how to easily obtain a subset  $\tilde{C} \subseteq \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ , and by Remark 4(ii), even if the optimal solution is not in  $\tilde{C}$ , Algorithm 1 will still find a  $\lambda^*$  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  $\mathbb{E}_{\mathbb{P}} \left[ \exp(-2\langle \lambda, X' \rangle) F(X')^4 \|X'\|^2 \right]$ . First, it is easy to verify that  $\mathbb{E}[X_i^4] < \infty$ , where  $X_i$  is the  $i$ th component for the random vector  $X'$ . By the Cauchy–Schwarz inequality, it suffices to find the set  $\tilde{C}$  to ensure  $\mathbb{E}_{\mathbb{P}} \left[ \exp(-4\langle \lambda, X' \rangle) F(X')^8 \right] < \infty$  for all  $\lambda \in \tilde{C}$ , by Eq. (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  $\tilde{C} \subseteq \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ .

4. Numerical example

In this numerical example, we consider the convergence of the SAA–Newton method w.r.t. different initial points, and the variance reduction of the importance sampling measure change. Additional simulation experiments can be found in [11]. Consider an Asian call option whose underlying asset is driven by a discrete NIG process, i.e.,

$$S_{t_i} = S_0 e^{at_i + X_{t_i}}, \quad 0 = t_0 \leq \dots \leq t_i \leq \dots \leq t_M = T,$$

where  $X'_i, i = 1, \dots, M$  are independent NIG random variables with parameters  $(\alpha_i, \beta_i, \delta_i)$ , and  $X_{t_i} = \sum_{j=1}^i X'_j$  is still a NIG random variable. The payoff is then given by

$$F(X') = e^{-rT} \left( \frac{1}{M} \sum_{i=1}^M S_0 e^{at_i + \sum_{j=1}^i X'_j} - K \right)^+.$$

We use the SAA–Newton method (Algorithm 1) to find the optimal parameter vector  $\lambda^* = (\lambda_{[1]}^*, \lambda_{[2]}^*, \dots, \lambda_{[M]}^*)$ , where  $\lambda_{[i]}^*$  is the

**Table 1**Convergence w.r.t. different initial points ( $n = 10^7$ ,  $\rho = 10^{-8}$ ).

	$\lambda_{[1]}^*$	$\lambda_{[2]}^*$	$\lambda_{[3]}^*$	$\lambda_{[4]}^*$	$\lambda_{[5]}^*$	$\lambda_{[6]}^*$	$\lambda_{[7]}^*$	$\lambda_{[8]}^*$	$\lambda_{[9]}^*$	$\lambda_{[10]}^*$	$\lambda_{[11]}^*$	$\lambda_{[12]}^*$	Steps
$\lambda^*$	1.75	1.73	1.70	1.66	1.61	1.54	1.45	1.33	1.17	0.98	0.73	0.42	N/A
Case 1	1.75	1.73	1.70	1.66	1.61	1.55	1.46	1.33	1.19	0.98	0.73	0.40	13
Case 2	1.75	1.73	1.70	1.66	1.61	1.54	1.46	1.32	1.16	0.98	0.73	0.35	9
Case 3	1.75	1.72	1.71	1.67	1.61	1.55	1.36	1.32	1.17	1.00	0.71	0.38	10
Case 4	1.75	1.73	1.70	1.66	1.61	1.55	1.44	1.33	1.17	0.95	0.80	0.48	10

**Table 2**Prices, 95%-confidence intervals, and variance reduction ratios, with  $(\alpha_i, \beta_i, \delta_i) = (2, 0.2, 0.8)$ .

N	MC price	IS price	VR ratio
1 000	10.08 $\pm$ 0.51	10.31 $\pm$ 0.14	15.8
5 000	10.26 $\pm$ 0.24	10.29 $\pm$ 0.05	21.7
10 000	10.33 $\pm$ 0.16	10.31 $\pm$ 0.04	24.2
50 000	10.27 $\pm$ 0.07	10.31 $\pm$ 0.02	22.3

optimal parameter of  $X'_i$  in the measure change. Let  $(\alpha_i, \beta_i, \delta_i) = (2, 0.2, 0.8)$ ,  $i = 1, \dots, M$ , for the original probability measure  $((\alpha_i, \beta_i + \lambda_{[i]}^*, \delta_i)$  for the new probability measure),  $a = 0$ , interest rate  $r = 0.02$ , initial price  $S_0 = 100$ , strike price  $K = 120$ , the number of observation points  $M = 12$ , maturity  $T = 1$ .

We investigate the sensitivity of the SAA–Newton method to the initial point. Case 1 uses  $\lambda_0 = (0, 0, \dots, 0)$ ; Case 2 uses  $\lambda_0 = (1, 1, \dots, 1)$ ; Case 3 uses  $\lambda_0 = (10, 10, \dots, 10)$ ; Case 4 uses  $\lambda_0 = (1, 2, \dots, 12)$ . The optimal parameter vector  $\lambda^*$  is obtained by an extensive brute-force search. Table 1 indicates that the algorithm converges quickly for all four initial points, and the estimated parameters are all close to the optimal values.

In practice, using very large  $n$  and very small  $\rho$  is computationally inefficient, so for option pricing, we use  $n = 1000$ ,  $\rho = 10^{-4}$ . Using Algorithm 2, Table 2 shows the estimated variance reduction of importance sampling, where “MC price” denotes the estimate price via classical Monte Carlo simulation, “IS price” denotes the estimate price via importance sampling, and “VR ratio” is the variance reduction ratio defined as the variance of classical Monte Carlo simulation divided by the variance of importance sampling.

## Acknowledgments

The authors thank the editors and referees for their valuable comments. This work was supported in part by the National Science Foundation (NSF) under Grants CMMI-0856256, CMMI-1362303, CMMI-1434419, and EECs-0901543, by the Air Force of Scientific Research (AFOSR) under Grant FA9550-15-10050, and by the National Natural Science Foundation of China (Project 11171256).

## References

- [1] L. Badouraly-Kassim, J. Lelong, I. Loumrhari, Importance sampling for jump processes and applications to finance, *J. Comput. Finance* (2015) in press.
- [2] P. Carr, D.B. Madan, Representing the CGMY and Meixner Lévy processes as time changed Brownian motions, *J. Comput. Finance* 2 (2008) 27–47.
- [3] R. Cont, Empirical properties of asset returns: stylized facts and statistical issues, *Quant. Finance* 1 (2001) 223–236.
- [4] K.D. Dinger, W. Hörmann, A general control variate method for option pricing under Lévy processes, *European J. Oper. Res.* 221 (2012) 368–377.
- [5] M.C. Fu, Gradient estimation, in: S.G. Henderson, B.L. Nelson (Eds.), *Handbooks in Operations Research and Management Science*, Vol. 13, Springer, Amsterdam, 2006, pp. 575–616. (Chapter 19).
- [6] 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, 2007, pp. 21–34.
- [7] P. Glasserman, *Monte Carlo Methods in Financial Engineering*, Springer, New York, 2003.
- [8] P. Glasserman, P. Heidelberger, P. Shahabuddin, Asymptotic optimal importance sampling and stratification for pricing path-dependent options, *Math. Finance* 9 (1999) 117–152.
- [9] P. Glasserman, Z.J. Liu, Sensitivity estimates from characteristic functions, *Oper. Res.* 58 (2010) 1611–1623.
- [10] T. Homem-de-Mello, A study on the cross-entropy method for rare-event probability estimation, *INFORMS J. Comput.* 19 (2007) 381–394.
- [11] G. Jiang, M.C. Fu, C. Xu, Optimal importance sampling for simulation of Lévy processes, in: *Winter Simulation Conference*, 2015, (in press).
- [12] B. Jourdain, J. Lelong, Robust adaptive importance sampling for normal random vectors, *Ann. Appl. Probab.* 19 (2009) 1687–1718.
- [13] R. Kawai, An importance sampling method based on the density transformation of Lévy processes, *Monte Carlo Methods Appl.* 12 (2007) 171–186.
- [14] R. Kawai, Optimal importance sampling parameter search for Lévy processes via stochastic approximation, *SIAM J. Numer. Anal.* 47 (2008) 293–307.
- [15] R. Kawai, Asymptotically optimal allocation of stratified sampling with adaptive variance reduction by strata, *ACM Trans. Model. Comput. Simul.* 20 (2010) 1–17.
- [16] R. Kawai, Likelihood ratio gradient estimation for Meixner distribution and Lévy processes, *Comput. Statist.* 27 (2012) 739–755.
- [17] S. Kim, R. Pasupathy, S.G. Henderson, A guide to sample-average approximation, in: M.C. Fu (Ed.), *Handbook of Simulation Optimization*, Springer, New York, 2014, pp. 207–241. (Chapter 8).
- [18] B. Lapeyre, J. Lelong, A framework for adaptive Monte Carlo procedures, *Monte Carlo Methods Appl.* 17 (2011) 77–98.
- [19] A. Nemirovski, A. Juditsky, G. Lan, A. Shapiro, Robust stochastic approximation approach to stochastic programming, *SIAM J. Optim.* 19 (2009) 1574–1609.
- [20] W. Schoutens, *Lévy Processes in Finance: Pricing Financial Derivatives*, John Wiley Sons, Chichester, 2003.
- [21] A. Shapiro, Monte Carlo sampling methods, in: A. Ruszczyński, A. Shapiro (Eds.), *Handbooks in Operations Research and Management Science*, Vol. 10, Elsevier, New York, 2003, pp. 353–425. (Chapter 6).
- [22] A.N. Shiryaev, *Probability*, Springer, Berlin, 1995.
- [23] Y. Su, M.C. Fu, Optimal importance sampling in securities pricing, *J. Comput. Finance* 5 (2002) 27–50.