Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

OPTIMAL IMPORTANCE SAMPLING FOR SIMULATION OF LÉVY PROCESSES

Guangxin Jiang

Department of Economics and Finance
City University of Hong Kong
Kowloon, Hong Kong

Michael C. Fu

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

Chenglong Xu

Department of Mathematics
Tongji University
Shanghai, 200092, China

ABSTRACT

This paper provides an efficient algorithm using Newton's method under sample average approximation (SAA) to solve the parametric optimization problem associated with the optimal importance sampling change of measure in simulating Lévy processes. Numerical experiments on variance gamma (VG), geometric Brownian motion (GBM), and normal inverse Gaussian (NIG) examples illustrate the computational advantages of the SAA-Newton algorithm over stochastic approximation (SA) based algorithms.

1 INTRODUCTION

Lévy processes are widely used to model financial underlying in derivatives pricing problems. Commonly used Lévy processes are the variance gamma (VG) process (Madan and Seneta 1990), the normal inverse Gaussian (NIG) process (Barndorff-Nielsen 1995), the generalized hyperbolic process (Eberlein and Keller 1995), the CGMY process (Carr et al. 2002), and the Meixner process (Schoutens and Teugels 1998). More complicated underlying models such as Lévy-driven stochastic volatility models have also been investigated in Barndorff-Nielsen and Shephard (2001) and Carr et al. (2003). For application of Greek estimation, see Fu (2007), Glasserman and Liu (2010).

In derivatives pricing, although analytical solutions are available for European-style options, most financial derivatives problems are solved by numerical methods, such as numerical solutions for PDEs (Cont and Voltchkova 2005), Fast-Fourier transition (FFT) (Carr and Madan 1999, Kwok et al. 2012), and Monte Carlo simulation (Glasserman 2003, Ribeiro and Webber 2006, Carr and Madan 2008). Although Monte Carlo simulation can solve high-dimensional problems, variance reduction techniques are often needed to improve the computational efficiency; see Glasserman (2003) for a survey. For Lévy processes models, variance reduction techniques used in derivatives pricing include importance sampling (Kawai 2008b), control variates (Dingeç and Hörmann 2012), stratified sampling (Kawai 2010), and their combination (Kawai 2008a).

In this paper, we focus on importance sampling. A commonly used approach is to formulate the importance sampling problem as a parametric optimization problem, and use stochastic approximation (SA) to solve this stochastic optimization problem. For more details, see Su and Fu (2002), Kushner and Yin (2003), Lapeyre and Lelong (2011). Kawai (2007) and Kawai (2008b) extended the approach to Lévy processes. SA is known to be sensitive to step sizes, which need to be chosen carefully. In this paper,

sample average approximation (SAA) (Shapiro 2003, Shapiro et al. 2009, Kim et al. 2014) is used to solve the resulting optimization problem. Jourdain and Lelong (2009) and Badouraly-Kassim et al. (2015) have used Newton's algorithm on SAA to find the optimal parameters in the Black-Scholes model and jump diffusion models, respectively. In our work, we consider more general Lévy process models.

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's method, and presents the resulting algorithms. Numerical examples are given in Section 4 for Asian options driven by normal inverse Gaussian (NIG) processes and geometric Brownian motion (GBM), and default probabilities of portfolios driven by variance gamma (VG) processes. Section 5 concludes.

2 LÉVY PROCESS MODELS AND MEASURE CHANGE

Lévy process models are commonly used to model the underlying assets. Let $\mathbf{X} = \{X_t, t \geq 0\}$ be a d-dimensional stochastic process defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, satisfying the following conditions: (i) $X_0 = 0$ a.s. (ii) $\mathbf{X}$ has independent and stationary increments. (iii) $\mathbf{X}$ is stochastically continuous, i.e., for all $a > 0$ and $s \geq 0$, $\lim_{t \to s} \mathbb{P}(||X_t - X_s|| > a) = 0$, where $||\cdot||$ is the Euclidean norm. Then, $\mathbf{X}$ is a Lévy process.

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 $F(X)$.

To improve computing efficiency, variance reduction techniques are routinely used with Monte Carlo simulation. In this paper, we focus on importance sampling, which attempts to give more weight to "important" outcomes, thereby increasing sampling efficiency. In the Black-Scholes model, the drift is changed (Glasserman et al. 1999, Su and Fu 2002). For Lévy processes, the Esscher measure change is commonly used (Kawai 2008b).

Let $\varphi_t(\lambda) = \log \mathbb{E}_{\mathbb{P}} [\exp(\langle \lambda, X_t \rangle)]$ denote the cumulant generating function of X_t, 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

$$\left. \frac{d\mathbb{P}_\lambda}{d\mathbb{P}} \right|_{\mathscr{F}_t} = \frac{e^{\langle \lambda, X_t \rangle}}{\mathbb{E}_{\mathbb{P}} \left[e^{\langle \lambda, X_t \rangle} \right]} = e^{\langle \lambda, X_t \rangle - \varphi_t(\lambda)}, \tag{1}$$

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

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

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

$$\mathrm{Var}(F(X), \lambda) := \mathbb{E}_{\mathbb{P}_\lambda} \left[\left(\left. \frac{d\mathbb{P}}{d\mathbb{P}_\lambda} \right|_{\mathscr{F}_T} \right)^2 F(X)^2 \right] - V^2 = \mathbb{E}_{\mathbb{P}} \left[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} F(X)^2 \right] - V^2. \tag{2}$$

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

$$\lambda^* \in \arg\min_{\lambda \in C} \mathrm{Var}(F(X), \lambda). \tag{3}$$

Then we will provide an efficient algorithm to find λ^*. Note that solving problem (3) is equivalent to finding the parameters λ^* minimizing the second moment, i.e., finding

$$\lambda^* \in \arg\min_{\lambda \in C} \mathbb{E}_{\mathbb{P}} \left[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} F(X)^2 \right]. \tag{4}$$

3 THE SAA-NEWTON ALGORITHM AND IMPORTANCE SAMPLING

Under the SAA framework the optimization problem (4) is changed to a deterministic optimization problem. Denote

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

where

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

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

$$\min_{\lambda \in C} f_n(\lambda), \tag{5}$$

which can be solved by an iterative deterministic algorithm such as Newton's method. We call this approach the SAA-Newton algorithm. Generally, Newton's method is not globally convergent and may become computationally impractical in high dimensions due to the calculation of the Hessian matrix. However, in our setting, the SAA problem is convex (Jiang and Fu 2015), the problem dimension is low, and there is a natural starting point (parameter value 0), so Newton's method works well. For more details on the convergence properties of the algorithm, see Jiang and Fu (2015).

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

$$\begin{aligned}
\nabla f_n(\lambda) &= \frac{1}{n} \sum_{j=1}^{n} \left[\left(\nabla \varphi_T(\lambda) - X_T^j \right) e^{-\left\langle \lambda, X_T^j \right\rangle + \varphi_T(\lambda)} F\left(X^j\right)^2 \right], \\
Hess[f_n(\lambda)] &= \frac{1}{n} \sum_{j=1}^{n} \Bigg[\Bigg(Hess[\varphi_T(\lambda)] \\
&\quad + \left(\nabla \varphi_T(\lambda) - X_T^j \right) \left(\nabla \varphi_T(\lambda) - X_T^j \right)' \Bigg) e^{-\left\langle \lambda, X_t^j \right\rangle + \varphi_T(\lambda)} F\left(X^j\right)^2 \Bigg],
\end{aligned}$$

which are used to find the optimal parameter in the Esscher measure change in Algorithm 1.

Similarly as in Jiang and Fu (2015), after obtaining the estimated λ_n^*, we can using the measure change equation to find the new process parameters set Λ^*. Algorithm 2 carries out importance sampling using the Esscher measure change.

4 NUMERICAL EXAMPLES

In this section, we provide numerical examples: pricing Asian call options driven by NIG process and GBM, and calculating the default probabilities of a portfolio of European call options driven by VG process.

4.1 Asian options

In this subsection, we compare the Newton-SAA algorithm with the SA approaches in Kawai (2008b) and Su and Fu (2002) in finding the optimal parameter of the Esscher measure change, and also estimate the resulting variance reduction in option pricing.

Algorithm 1 SAA-Newton algorithm for finding 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's method 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 $\lambda_n^* = \lambda_{k+1}$.

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 $\left\{X_t^1, 0 \le t \le T\right\}, \left\{X_t^2, 0 \le t \le T\right\}, \ldots, \left\{X_t^N, 0 \le t \le T\right\}$ 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 $\hat{V} = 1/N \sum_{i=1}^{N} V^i$.

4.1.1 NIG process

Consider an Asian call option whose underlying asset S_t is driven by a NIG process, i.e.,

$$S_t = S_0 e^{at + X_t}, 0 \le t \le T,$$

where S_0 is the initial price of the asset, T is the maturity, and a is a constant. If M is the number of observation points, and the averaging begins at 0, the payoff is given by

$$F(X) = \left(\frac{1}{M} \sum_{i=1}^{M} S_{t_i} - K \right)^+ = \left(\frac{1}{M} \sum_{i=1}^{M} S_0 e^{at_i + X_{t_i}} - K \right)^+,$$

where $0 < t_1 < t_2 < \ldots < t_M = T$ and $\{X_{t_i}, i = 1, 2, \ldots\}$ is a discrete NIG process. For more details, see Barndorff-Nielsen (1995), and Schoutens (2003).

The characteristic function for the NIG process is given by

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

where $t = 1$ is the characteristic function of a NIG random variable. The process parameters under the original probability measure are (α, β, δ). According to Jiang and Fu (2015), the process parameters under the new probability measure are $(\alpha, \beta + \lambda^*, \delta)$, where λ^* is the optimal parameter.

We simulate the path of X_t using the independent increments and infinite divisibility of NIG random variables. Specifically, we generate an M-dimensional (d is replaced by M in the following) NIG random vector $X' = (X_1', X_2', \ldots, X_M')$, where X_i' is a NIG random variable with parameter $\delta' = \delta/M$, and let

$X_{t_i} = \sum_{j=1}^{i} X'_j$. The payoff is then given by

$$F(X') = \left(\frac{1}{M} \sum_{i=1}^{M} S_0 e^{at_i + \Sigma_{j=1}^{i} X'_j} - K \right)^+ . \tag{6}$$

Our objective is to find the optimal parameter $\lambda^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_M^*)$, where λ_i^* is the optimal parameter of X'_i in the measure change.

Based on Algorithms 1 and 2, computational results are presented. We consider process parameter values $\Lambda = (\alpha, \beta, \delta) = (2, 0.2, 0.8)$ as in Lemaire and Pages (2009). For the other parameters, let $a = 0$, interest rate $r = 0.02$, initial price $S_0 = 100$, strike price $K = 100$, the number of observation points $M = 5$, maturity $T = 1$, and the initial value in Newton's method $\lambda_0 = [0...0]_{1\times M}$. First, we compare SA with the SAA-Newton algorithm. In SA, the step size is ε/m. In Fig.1, the left and right panels show the convergences of the norms of the gradient against the number of iterations for different step sizes ε in SA and different sample sizes n in SAA, respectively. Fig.1 illustrates SA's well-known sensitivity to the choice of step size. On the other hand, for SAA larger sample size leads to better convergence rate, making it more straightforward for practitioners to apply. Although the SAA-Newton converges using far fewer iterations than SA, each iteration requires far more computation, so Table 1 compares total run time. The results indicate that the run time of the SAA-Newton algorithm is considerably less than SA for the same level of accuracy.

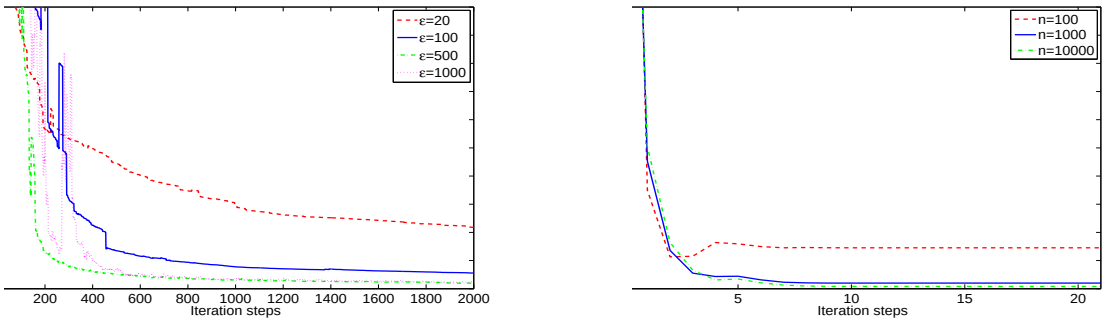

Figure 1: Norm of gradients for SA (left panel) and SAA-Newton (right panel)

Table 1: Run time and norm of gradient for SA and SAA-Newton

SAA-Newton			SA		
n	run time	norm of gradient	N	run time	norm of gradient
250	0.054	0.0022	5000	0.27	0.016
500	0.11	0.0012	10000	0.54	0.0042
1000	0.21	0.00097	20000	1.1	0.0014
2500	0.52	0.00062	50000	2.7	0.0010
5000	1.0	0.00051	100000	5.4	0.00081
10000	2.1	0.00039	200000	10.8	0.00079
25000	5.2	0.00035	500000	27.0	0.00052

Next, we consider the importance sampling performance for the following cases: strike price $K = 20, 100, 200$ and $\delta = 0.4, 0.8$. Figs. 2, 3, and 4 display the results in the form of box plots based on 100

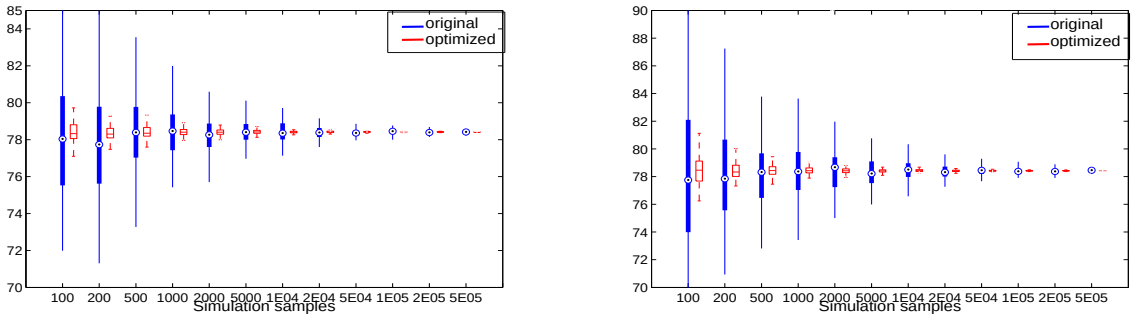

Figure 2: Convergence of prices for $K = 20$; $\delta = 0.4$ (left panel) and $\delta = 0.8$ (right panel)

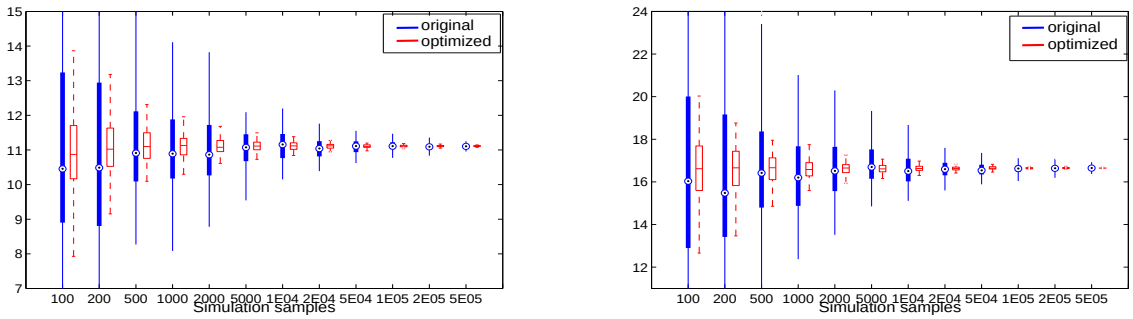

Figure 3: Convergence of prices for $K = 100$; $\delta = 0.4$ (left panel) and $\delta = 0.8$ (right panel)

macro-replications. The results indicate that the optimal importance sampling reduces variance significantly, and Table 2 displays the variance reduction ratios defined as the variance of classical MC divided by the variance of MC with importance sampling. The results indicate significantly greater variance reduction for both out-of-the-money and in-the-money options than at-the-money options.

Table 2: Variance reduction ratios

	$\alpha = 2, \beta = 0.2, \delta = 0.8$			$\alpha = 2, \beta = 0.2, \delta = 0.4$		
N	$K = 20$	$K = 100$	$K = 200$	$K = 20$	$K = 100$	$K = 200$
500	60.9	14.1	68.2	52.8	11.9	72.5
1000	73.9	14.3	79.3	91.1	11.5	70.8
5000	57.7	14.3	64.8	67.5	14.1	70.1
10000	83.7	17.8	65.4	143.1	13.9	73.8
50000	78.1	14.2	55.2	56.2	14.3	101.2
100000	72.3	18.7	52.6	88.5	18.7	89.9

4.1.2 Geometric Brownian motion

Different from the SA approach in Kawai (2008b), Su and Fu (2002) proposed an SA approach based on estimated gradients via IPA. Consider an Asian call option whose underlying assets are driven by GBM, i.e.,

$$S_t = S_0 e^{(r-\frac{1}{2}\sigma^2)t+\sigma W_t},$$

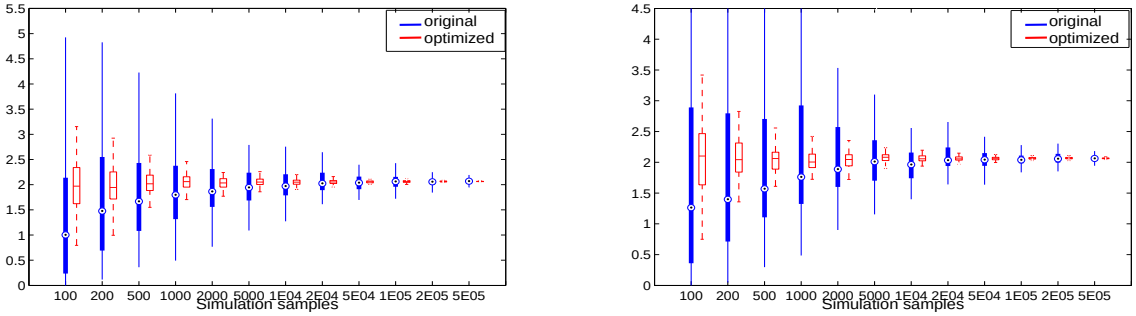

Figure 4: Convergence of prices for $K = 200$; $\delta = 0.4$ (left panel) and $\delta = 0.8$ (right panel)

where r is the interest rate, σ is volatility, S_0 is the initial price of the asset, and $\{W_t\}$ is standard Brownian motion under the original measure. Su and Fu (2002) change the drift in Brownian motion, which is a special case of the Esscher measure change. They assume a new Brownian motion

$$\tilde{W}_t = W_t - \lambda t,$$

with the Radon-Nikodym derivative given by

$$\frac{d\mathbb{P}_\lambda}{d\mathbb{P}} = e^{\lambda W_T - \frac{1}{2}\lambda^2 T}.$$

Note that, unlike the last example, we do not change the process parameter (the drift in Brownian motion) at each observation point, i.e., we use the same λ, which is a scalar, to obtain the new process parameter, at each observation point.

We consider the same setting in Su and Fu (2002): $S_0 = 50$, $K = 50$, $\sigma^2 = 0.2$, $r = 0.05$, $T = 1$, number of observation points $M = 255$, which is daily average and the averaging begins at 0. Let $\tilde{n}$ denote the number of replications per iteration in SA. Fig.5 plots SAA-Newton and SA for $\tilde{n} = n = 50$ and 500.

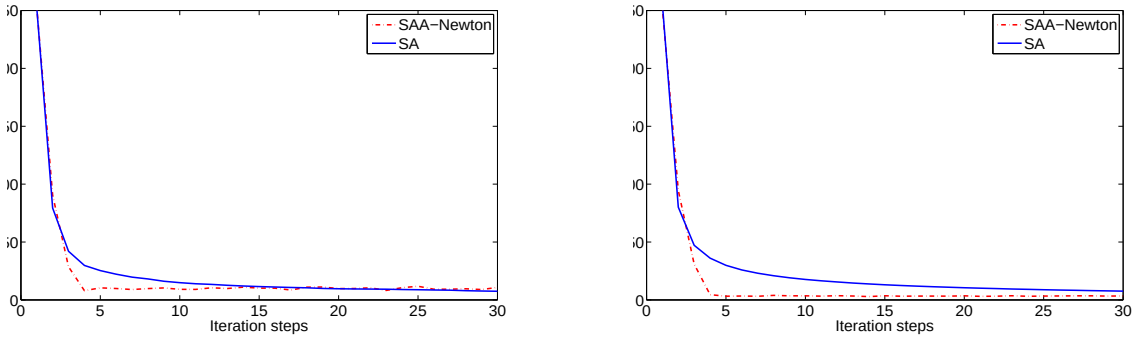

Figure 5: Norm of gradients; $\tilde{n} = n = 50$ (left panel) and $\tilde{n} = n = 500$ (right panel)

The results again indicate that SAA-Newton converges faster than SA. Table 3 shows the run times for SAA-Newton with iteration steps $N = 15$ and SA with sample size $\tilde{n} = 50$ per iteration. Although SAA-Newton requires more computation per iteration than SA, due to inverting a Hessian matrix, the overall run time of SAA-Newton is still less than SA for the same level of accuracy. Next, we consider the importance sampling performance. Similarly as in Su and Fu (2002), consider partial average Asian option pricing with the averaging beginning 60 days before the option maturity date. Let $r = 0.05$, $\sigma = 0.2$, $S_0 = 100$, $n = 50, 500$, and $K = 100, 120$, with other parameter values remaining the same. Fig. 6 displays

Table 3: Run time, estimate drift and norm of gradient for SAA-Newton and SA

	SAA-Newton				SA		
n	run time	estimate drift	$\|\nabla f\|$	N	run time	estimate drift	$\|\nabla f\|$
50	0.08	0.488	11.5	10	0.04	0.417	16.2
500	0.13	0.494	2.8	100	0.48	0.475	4.0
5000	0.61	0.493	1.2	1000	4.58	0.489	1.2
50000	5.81	0.492	0.65	10000	45.5	0.492	0.82
500000	79.8	0.492	0.71	100000	458.4	0.493	0.742

the results in the form of box plots based on 100 macro-replications. The box plots depict the median, quartiles, and whiskers without the extreme points.

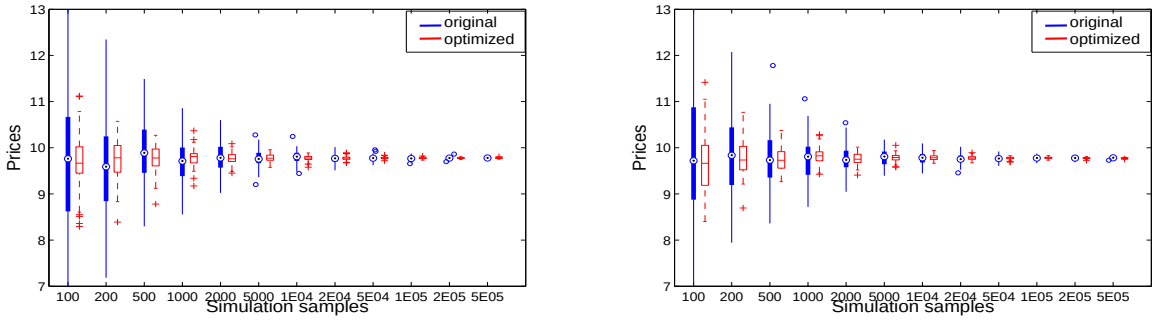

Figure 6: Convergence of prices for $K = 100$; $n = 50$ (left panel) and $n = 500$ (right panel)

4.1.3 Comparison of two schemes for importance sampling

For the NIG example, we changed the process parameter at every observation point, so the importance sampling parameter is a vector, denoted here by IS-M. For the GBM example, we kept the drift the same, i.e., the optimal parameter is a scalar, denoted by IS-1.

We compare the variance reduction of these two schemes on Asian options driven by GBM, using SAA-Newton to find the optimal parameter. Let $S_0 = 100$, $K = 120$, $T = 1$, $M = 10$ averaged over entire period, iteration steps $m = 20$, with $N = 10000$ for the pricing. Changing n w.r.t. three different values of (r, σ), Table 4 is obtained. CMC stands for classical Monte Carlo, and VR-1 and VR-M stand for the variance reduction ratio for IS-1 and IS-M, respectively. IS-M has larger variance reduction ratios than the IS-1, but requires much longer run time than IS-1, which in this example outweighs the variance reduction. Also note that when $n = 1000$, the estimated optimal parameter for IS-1 is quite close to the true optimal parameter, so no further variance reduction is gained by increasing n. However, for IS-M, $n = 1000$ is not enough to solve the high-dimensional problem, so higher n is required to better estimate the optimal parameter value.

4.2 Default probability of portfolio

In this subsection, we estimate the default probability of a simple portfolio L containing M European call options whose underlying assets are driven by VG processes, i.e.,

$$S_t^i = S_0^i e^{a_i t + X_t^i}, 0 \le t \le T, i = 1, 2, \ldots, M, \tag{7}$$

Table 4: Comparison of IS-1 and IS-M

	CMC		IS-1			IS-M		
	prices	n	run time	prices	VR-1	run time	prices	VR-M
$r=0.15$	0.162	1000	0.01	0.165	9.6	0.18	0.160	16.0
$\sigma=0.1$		10000	0.06	0.159	10.5	1.79	0.161	35.3
$r=0.05$	0.712	1000	0.01	0.700	7.8	0.18	0.706	14.3
$\sigma=0.2$		10000	0.06	0.690	7.9	1.80	0.707	25.0
$r=0.05$	4.372	1000	0.01	4.313	5.6	0.18	4.362	12.4
$\sigma=0.4$		10000	0.06	4.353	5.8	1.82	4.368	16.7

where $\{X_t^i\}$ is a VG process and $\{S_0^1, S_0^2, \ldots, S_0^M\}$ are the initial prices of the M underlying assets, and all the assets have the same maturity T, and $\{a_i, i=1,2,\ldots,M\}$ are constants. The portfolio is given by

$$L = \sum_{i=1}^{M} V_i,$$

with

$$V_i = e^{-rT}\mathbb{E}[(S_T^i - K_i)^+] = e^{-rT}\mathbb{E}[(S_0 e^{a_i + X_T^i} - K_i)^+],$$

where K_i is strike price of the ith option. The process parameters of X_T in the original probability are (σ, ν, θ), according to Jiang and Fu (2015), the new process parameters for the Esscher measure change are

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

Suppose we are interested in the probability that the portfolio is below some level L_b, i.e.,

$$\Pr\{L \le L_b\} = \mathbb{E}[\mathbf{1}_{L\le L_b}].$$

We know if L_b is small, this is the probability of a rare event. Let the process parameter $(\sigma, \nu, \theta) = (0.1, 0.3, 0)$, the number of options $M = 5$, $a_i = 0, i = 1,2,\ldots,M$, $S_0 = [80\ 90\ 100\ 110\ 120]$ and the corresponding strike prices $K = [70\ 80\ 90\ 100\ 110]$. We consider sample sizes $n = 500, 5000, 50000$ and default levels $L_b = 10, 20$. Fig.7 displays the convergence of the norm of the gradient via SAA-Newton. Similar to the previous examples, higher n leads to smaller norms of gradients.

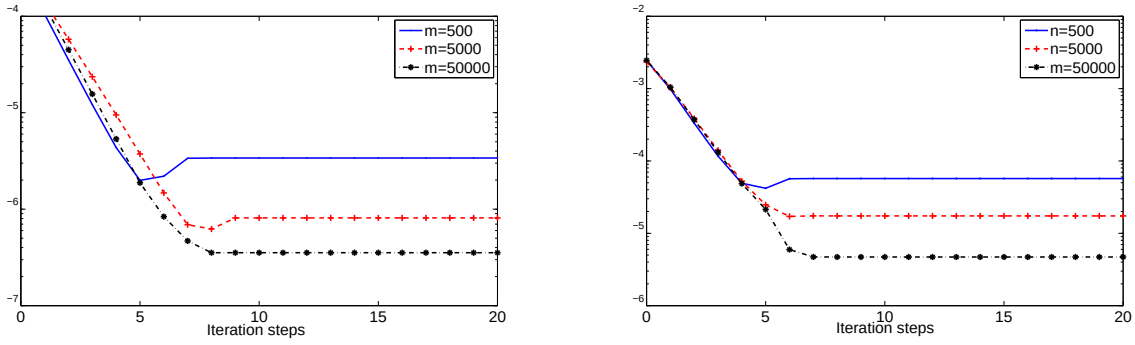

Figure 7: Norm of gradient; $L_b = 10$ (left panel) and $L_b = 20$ (right panel)

Next, we consider the importance sampling performance. Let $n = 5000$, and $L_b = 10, 20$, respectively, with other parameter values remaining unchanged. Fig. 8 displays the results in the form of box plots based

on 100 macro-replications. The cross marks in the box plots are the mean of the estimated probabilities of the 100 macro-replications. For $L_b = 10$, default is a rare event, and classical Monte Carlo returns an estimate of zero for small sample sizes ($N = 100, 200$), whereas once importance sampling is applied, reasonable estimates can be obtained even for small N.

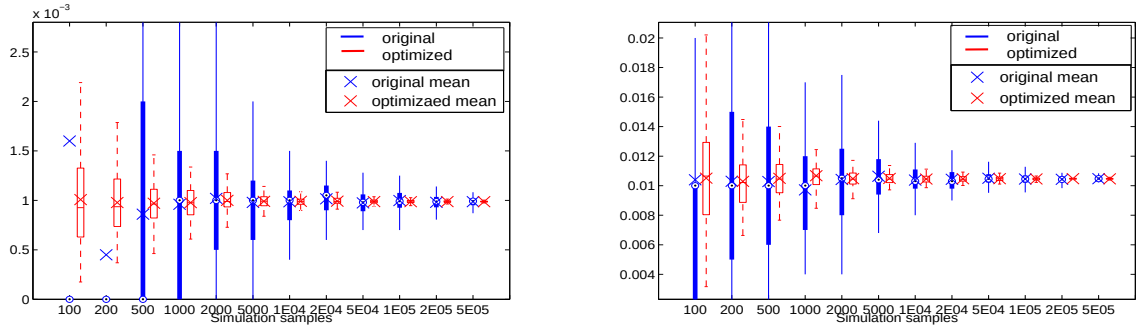

Figure 8: Convergence of probabilities; $L_b = 10$ (left panel) and $L_b = 20$ (right panel)

Tables 5 and 6 show the variance reduction, where P_c and P_{IS} denote the estimated probability using the classical Monte Carlo and the optimal importance sampling, respectively, with the 90% confidence half widths shown in parentheses. Clearly importance sampling works better for estimating rare event probabilities ($L_b = 10$), especially for small sample size.

Table 5: Estimated probabilities (half widths) and variance reduction ratios for $L_b = 10$

N	P_c	P_{IS}	VR
100	0.1600% (0.0667%)	0.1008% (0.0082%)	356.7
500	0.0860% (0.0219%)	0.0971% (0.0037%)	48.1
1000	0.0960% (0.0163%)	0.0977% (0.0026%)	42.5
5000	0.0980% (0.0074%)	0.0990% (0.0012%)	39.5
10000	0.0986% (0.0052%)	0.0990% (0.0008%)	39.5
50000	0.0975% (0.0023%)	0.0988% (0.0004%)	39.4
100000	0.0998% (0.0017%)	0.0987% (0.0003%)	40.2

Table 6: Estimated probabilities (half widths) and variance reduction ratios for $L_b = 20$

N	P_c	P_{IS}	VR
100	1.040% (0.1692%)	1.051% (0.0489%)	16.6
500	1.028% (0.0753%)	1.050% (0.0217%)	12.7
1000	0.972% (0.0518%)	1.068% (0.0155%)	11.5
5000	1.064% (0.0242%)	1.050% (0.0069%)	12.5
10000	1.038% (0.0169%)	1.045% (0.0048%)	12.2
50000	1.051% (0.0076%)	1.047% (0.0022%)	12.3
100000	1.045% (0.0054%)	1.046% (0.0015%)	12.3

5 CONCLUSION

In this paper, we first formulate the importance sampling problem as a parametric stochastic optimization problem, and then propose a new method, the SAA-Newton algorithm, to find the optimal importance sampling parameters based on the Esscher measure change for Lévy processes. Numerical experiments study the effectiveness of the method, and indicate that the Newton-SAA algorithm can find optimal parameters faster than SA-based algorithms.

ACKNOWLEDGEMENTS

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

Badouraly-Kassim, L., J. Lelong, and I. Loumrhari. 2015. "Importance sampling for jump processes and applications to finance,". *Journal of Computational Finance*. To appear.

Barndorff-Nielsen, O. E. 1995. "Normal inverse Gaussian distributions and the modeling of stock returns". Technical report, Department of Theoretical Statistics, Aarhus University.

Barndorff-Nielsen, O. E., and N. Shephard. 2001. "Non-Gaussian Ornstein-Uhlenbeck-based models and some of their applications in financial economics". *Journal of the Royal Statistical Society* 63:167–241.

Carr, P., H. Geman, D. Madan, and M. Yor. 2002. "Non-Gaussian Ornstein-Uhlenbeck-based models and some of their applications in financial economics". *Journal of Business* 75:305–332.

Carr, P., H. Geman, D. Madan, and M. Yor. 2003. "Stochastic volatility for Lévy processes". *Mathematical Finance* 13:345–382.

Carr, P., and D. B. Madan. 1999. "Option valuation using the fast Fourier transform". *Journal of Computational Finance* 2:61–73.

Carr, P., and D. B. Madan. 2008. "Representing the CGMY and Meixner Lévy processes as time changed Brownian motions". *Journal of Computational Finance* 2:27–47.

Cont, R., and E. Voltchkova. 2005. "A finite difference scheme for option pricing in jump diffusion and exponential Lévy models". *SIAM Journal on Numerical Analysis* 43:1596–1626.

Dingeç, K. D., and W. Hörmann. 2012. "A general control variate method for option pricing under Lévy processes". *European Journal of Operational Research* 221:368–377.

Eberlein, E., and U. Keller. 1995. "Hyperbolic distributions in finance". *Bernoulli* 1:281–299.

Fu, M. C. 2007. "Variance-Gamma and Monte Carlo". In *Advances in Mathematical Finance*, edited by M. Fu, R. A. Jarrow, J. Y. Yen, and R. J. Elliott, 21–34. Boston: Birkhäuser.

Glasserman, P. 2003. *Monte Carlo Methods in Financial Engineering*. New York: Springer.

Glasserman, P., P. Heidelberger, and P. Shahabuddin. 1999. "Asymptotic optimal importance sampling and stratification for pricing path-dependent options". *Mathematical Finance* 9:117–152.

Glasserman, P., and Z. J. Liu. 2010. "Sensitivity estimates from characteristic functions". *Operations Research* 58:1611–1623.

Jiang, G., and M. C. Fu. 2015. "On sample average approximation algorithms for determining the optimal importance sampling parameters in pricing financial derivatives on Lévy processes". *Working paper*.

Jourdain, B., and J. Lelong. 2009. "Robust adaptive importance sampling for normal random vectors". *Annals of Applied Probability* 19:1687–1718.

Kawai, R. 2007. "An importance sampling method based on the density transformation of Lévy processes". *Monte Carlo Methods and Applications* 12:171–186.

Kawai, R. 2008a. "Adaptive Monte Carlo variance reduction for Lévy processes with two time-scale stochastic approximation". *Methodology and Computing in Applied Probability* 10:199–223.

Kawai, R. 2008b. "Optimal importance sampling parameter search for Lévy processes via stochastic approximation". *SIAM Journal on Numerical Analysis* 47:293–307.

Kawai, R. 2010. "Asymptotically optimal allocation of stratified sampling with adaptive variance reduction by strata". *ACM Transactions on Modeling and Computer Simulation* 20:1–17.

Kim, S., R. Pasupathy, and S. G. Henderson. 2014. "A guide to sample-average approximation". In *Handbook of Simulation Optimization*, edited by M. Fu, Chapter 8, 207–241. New York: Springer.

Kushner, H., and G. G. Yin. 2003. *Stochastic Approximation and Recursive Algorithms and Applications*. New York: Springer.

Kwok, Y. K., K. S. Leung, and H. Y. Wong. 2012. "Efficient options pricing using the fast Fourier transform". In *Handbook of Computational Finance*, edited by J. C. Duan, W. K. Härdle, and J. E. Gentle, Chapter 21, 579–604. Berlin: Springer.

Lapeyre, B., and J. Lelong. 2011. "A framework for adaptive Monte Carlo procedures". *Monte Carlo Methods and Applications* 17:77–98.

Lemaire, V., and G. Pages. 2009. "Unconstrained recursive importance sampling". *The Annals of Applied Probability* 20:1029–1067.

Madan, D. B., and E. Seneta. 1990. "The VG model for share market returns". *Journal of Business* 63:511–524.

Ribeiro, C., and N. Webber. 2006. "Correcting for simulation bias in Monte Carlo methods to value exotic options in models driven by Lévy processes". *Applied Mathematical Finance* 13:333–352.

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

Schoutens, W., and J. L. Teugels. 1998. "Lévy processes, polynomials and martingales". *Communications in Statistics* 14:335–349.

Shapiro, A. 2003. "Monte Carlo sampling methods". In *Handbooks in Operations Research and Management Science*, edited by A. Ruszczynski and A. Shapiro, Volume 10, Chapter 6, 353–425. New York: Elsevier.

Shapiro, A., D. Dentcheva, and A. Ruszczynski. 2009. *Lectures on Stochastic Programming Modeling and Theory*. Philadelphia: SIAM-MPS.

Su, Y., and M. C. Fu. 2002. "Optimal importance sampling in securities pricing". *Journal of Computational Finance* 5:27–50.

AUTHOR BIOGRAPHIES

GUANGXIN JIANG is research assistant in the Department of Economics and Finance at City University of Hong Kong, Hong Kong. His research interests lie in simulation methodology, modeling, analysis, and optimization, with applications in financial engineering. His email address is guajiang@cityu.edu.hk.

MICHAEL C. FU is the Ralph J. Tyser Professor of Management Science in the Robert H. Smith School of Business, with a joint appointment in the Institute for Systems Research and affiliate appointment in the Department of Electrical and Computer Engineering, at the University of Maryland. His research interests include simulation modeling, analysis, and optimization, with applications to manufacturing, supply chain management, and financial engineering. He served as WSC 2011 Program Chair and is a Fellow of **INFORMS** and **IEEE**. His e-mail address is mfu@rhsmith.umd.edu.

CHENGLONG XU is professor of Department of Mathematics at Tongji University, China. His research interests include simulation methodology, numerical analysis, spectral method, partial differential equations, with applications in financial engineering. His e-mail address is clxu@tongji.edu.cn.