Adaptive Monte Carlo variance reduction for Lévy processes with two-time-scale stochastic approximation

REIICHIRO KAWAI

Abstract

We propose an approach to a two-fold optimal parameter search for a combined variance reduction technique of the control variates and the important sampling in a suitable pure-jump Lévy process framework. The parameter search procedure is based on the two-time-scale stochastic approximation algorithm with equilibrated control variates component and with quasi-static importance sampling one. We prove the almost sure convergence of the algorithm to a unique optimum. The parameter search algorithm is further embedded in adaptive Monte Carlo simulations in the case of the gamma distribution and process. Numerical examples of the CDO tranche pricing with the Gamma copula model and the intensity Gamma model are provided to illustrate the effectiveness of our method.

Keywords: Esscher transform, gamma distribution and process, Girsanov theorem, Monte Carlo simulation, infinitely divisible distribution, stochastic approximation, variance reduction.

2000 Mathematics Subject Classification: 65C05, 62L20, 93E35, 60G51.

1 Introduction

The class of Lévy processes has drawn a great amount of attention, for example, in the field of mathematical finance, where the Brownian motion has long been considered to be insufficient to describe a variety of asset price dynamics. Due to lower mathematical tractability of Lévy processes, it is evidently more often than in the Brownian motion case that one has to resort to Monte Carlo simulations for the computation of the expectation. Those facts will make the Monte Carlo simulation and its variance reduction in the Lévy process setting more and more important.

Published in *Methodology and Computing in Applied Probability* (2008) **10**(2) 199-223.

Email address: reiichiro_kawai@ybb.ne.jp. Postal address: Center for the Study of Finance and Insurance, Osaka University, Toyonaka, Osaka, 560-8531, Japan.

In this paper, we propose an adaptive Monte Carlo variance reduction procedure in a suitable class of Lévy processes. Our method is, on one hand, an extension of Kawai [16], where an optimal parameter search for the importance sampling is studied in a pure-jump Lévy process setting. On the other hand, our method can also be thought of as an extension of the same approach, but in the Brownian motion setting, of Kawai [17], that is, an adaptive Monte Carlo variance reduction for the combined control variates and importance sampling. To develop our method, we take the same route as in [17]: (i) we combine two variance reduction techniques, the control variates (CV) and the importance sampling (IS), and investigate the two-fold optimality of the combination, (ii) we apply the two-time-scale stochastic approximation algorithm in parameter search for the combination and prove almost sure convergence of the algorithm to a unique optimum, and (iii) we incorporate the parameter search procedure into adaptive Monte Carlo simulation.

There is a vast literature on optimal use of a single variance reduction technique, and some combinations have been investigated, for example, IS and the stratification in Glasserman *et al.* [9], IS and the temporal difference control variates in Randhawa and Juneja [21], while CV and IS in [17]. For the combination of CV and IS, as discussed in [17], it is unclear whether or not there exists global optimum, and thus by “simultaneous optimality”, we mean “optimality in IS parameter with CV always held optimal with respect to the IS parameter.” Later, we find it difficult to implement the opposite, that is, “optimality in CV parameter with IS always held optimal with respect to the CV parameter” in the Lévy process setting.

The optimal parameter search using stochastic approximation algorithm has been studied for the Gaussian setting in Su and Fu [24] and Arouna [2], and for a pure-jump Lévy process in [16]. A crucial difference from the Brownian motion case is that the parameter iterates for IS need to remain within a bounded domain due to the lack of the higher moments of the underlying Lévy measure. In order to further search two-fold parameters of CV and of IS simultaneously, as in [17], we apply the *two-time-scale* stochastic approximation algorithm, which is a stochastic recursive algorithm in which some of the components are updated using step-sizes that are very small compared to those of the remaining components and whose almost sure convergence is first rigorously proved in Borkar [6].

It would not be clever to perform the optimal parameter search algorithm first and then to re-run the Monte Carlo simulation in cooperation with the optimized variance reduction techniques, simply because this double implementation is evidently more time-consuming than the plain Monte Carlo simulation first performed. The idea of incorporating the parameter search with a stochastic approximation into adaptive Monte Carlo simulation is studied in Arouna [3] and is also applied to the two-time-scale version of [17]. Those methods make full use of a property of the Gaussian distribution that after the measure change by the Girsanov theorem, it again follows a Gaussian distribution with its mean shifted and with the same shape (variance). In general, the infinitely divisible distributions do not have such a convenient property, while an exception is the gamma

distribution, which has recently drawn a vast attention, for example, in credit derivatives modeling. For the adaptive Monte Carlo simulations, we restrict ourselves to the gamma distribution.

The rest of this paper is organized as follows. Section 2 recalls in brief the principle of the control variates and the importance sampling. Section 3 formulates combined control variates and importance sampling and discusses the optimality with the control variates component held fixed. Moreover, the almost sure convergence of a two-time-scale algorithm to a unique root of the gradients of the variance is proved. Then, in Section 4, the parameter search using the two-time-scale algorithm is embedded in adaptive Monte Carlo procedure in the case of the gamma distribution, and the strong law of large numbers for empirical mean and the central limit theorems for empirical variance are proven to hold. Section 5 presents and discusses results of numerical experiments of the CDO tranche pricing with the gamma copula model and the intensity gamma model. Finally, Section 6 concludes this study with some remarks.

2 Preliminaries

Let us begin with some notations which will be used throughout the text. $\mathbb{R}^d$ is the d -dimensional Euclidean space with the norm $\|\cdot\|$ and the inner product $\langle \cdot, \cdot \rangle$, $\mathbb{R}_0^d := \mathbb{R}^d \setminus \{0\}$. $\text{Leb}(\cdot)$ denotes the Lebesgue measure, $\mathbb{D}([0, \infty); \mathbb{R}^d)$ is the space of càdlàg functions from $[0, \infty)$ into $\mathbb{R}^d$ equipped with the Skorohod topology, and $\mathbb{P}|_{\mathcal{F}_t}$ is the restriction of a probability measure $\mathbb{P}$ to the σ -field $\mathcal{F}_t$. We denote by ∇_x the gradient, and by $\text{Hess}_x[\cdot]$ the Hessian matrix with respect to the variable x . By $\Pi_H(x)$, we indicate the projection of x onto the set H , that is, the closest point in H to x . We say that a stochastic process $\{X_t : t \geq 0\}$ in $\mathbb{R}^d$ is a Lévy process if it has independent and stationary increments, if it is continuous in probability, and if $X_0 = 0$, $\mathbb{P}$ -a.s. By the Lévy-Khintchine representation theorem, the characteristic function of its marginal distributions is uniquely given by

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

where $\gamma \in \mathbb{R}^d$, A is a symmetric nonnegative-definite $d \times d$ matrix, and ν is a Lévy measure on $\mathbb{R}_0^d$, that is, a measure satisfying $\int_{\mathbb{R}_0^d} (\|z\|^2 \wedge 1) \nu(dz) < +\infty$. When the above holds, then we say that the Lévy process $\{X_t : t \geq 0\}$ is generated by the triplet (γ, A, ν) . *In this paper, we restrict ourselves to pure-jump Lévy processes, that is, we set $A \equiv 0$ and $\{X_t : t \geq 0\}$ to be a Lévy process in $\mathbb{R}^d$ generated by $(\gamma, 0, \nu)$ throughout.* Moreover, we assume that all components of the marginal X_1 are non-degenerate, and that $\int_{\mathbb{R}_0^d} \|z\|^6 \nu(dz) < +\infty$. These guarantee that the variance-covariance matrix $\int_{\mathbb{R}_0^d} z z' \nu(dz)$ is well defined and is strictly positive definite. A function $f : \mathbb{R}^d \mapsto [0, \infty)$ is said to be submultiplicative if there exists a positive constant a such that $f(x+y) \leq a f(x) f(y)$ for $x, y \in \mathbb{R}^d$. Letting $c \in \mathbb{R}$, $\gamma \in \mathbb{R}^d$, and $b > 0$, if $f(x)$ is submultiplicative on $\mathbb{R}^d$, then $f(cx + \gamma)^b$ is submultiplicative, and the functions $\|x\| \vee 1$, $e^{\langle c, x \rangle}$ are submultiplicative, and a product of two

submultiplicative functions is submultiplicative. We recall an important moment property of Lévy processes, which will be used often in what follows.

Theorem 2.1. (Sato [22], Theorem 25.3) *Let f be a submultiplicative, locally bounded, measurable function on $\mathbb{R}^d$, and let $\{X_t : t \geq 0\}$ be a Lévy process in $\mathbb{R}^d$ with Lévy measure ν and without Gaussian component. Then, $\mathbb{E}[f(X_t)]$ is finite for every $t > 0$ if and only if $\int_{\|z\|>1} f(z)\nu(dz) < +\infty$.*

Define a set

$$\begin{aligned}\Lambda_1 &:= \left\{ \lambda \in \mathbb{R}^d : \mathbb{E}_{\mathbb{P}} \left[e^{\langle \lambda, X_1 \rangle} \right] < +\infty \right\} \\ &= \left\{ \lambda \in \mathbb{R}^d : \int_{\|z\|>1} e^{\langle \lambda, z \rangle} \nu(dz) < +\infty \right\},\end{aligned}$$

where the equality holds by Theorem 2.1. Moreover, the set Λ_1 is convex and contains the origin. (See Theorem 25.17 [22].) We will put in force the assumption $\text{Leb}(\Lambda_1) \in (0, +\infty)$, which is not as restrictive as it may look, in the sense that the tails of Lévy measure satisfying $\Lambda_1 = \mathbb{R}^d$ are too light to yield great practical benefits. (We come back to this point in Remark B.2.) For $\lambda \in \Lambda_1$, we denote by φ_t the cumulant generating function of the marginal distribution X_t under the probability measure $\mathbb{P}$, that is, $\varphi_t(\lambda) := \ln \mathbb{E}_{\mathbb{P}}[e^{\langle \lambda, X_t \rangle}]$. We also denote by $(\mathcal{F}_t)_{t \geq 0}$ the natural filtration of $\{X_t : t \geq 0\}$.

In this paper, we are interested in evaluating the expectation

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

by Monte Carlo simulation, where F is an operator mapping $\mathbb{D}([0, T]; \mathbb{R}^d)$ into $\mathbb{R}$, and $F(X) := F(\{X_t : t \in [0, T]\}) \in L^4(\Omega, \mathcal{F}_T, \mathbb{P})$. We will henceforth assume that $\mathbb{P}(F(X) \neq 0) > 0$ and

$$\mathbb{E}_{\mathbb{P}}[F(X)^4 \|X_T\|^2] < +\infty. \quad (2.1)$$

2.1 Control variates variance reduction

The control variates method we will consider in this paper is of a linear type based upon the following equality,

$$C = \mathbb{E}_{\mathbb{P}}[F(X)] = \mathbb{E}_{\mathbb{P}}[F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle]. \quad (2.2)$$

Assume that $F(X)$ and X_T are correlated, that is,

$$\mathbb{E}_{\mathbb{P}}[(F(X) - C)(X_T - \mathbb{E}_{\mathbb{P}}[X_T])] \neq 0. \quad (2.3)$$

The condition (2.1) guarantees that the variance of the component inside the right hand side expectation in (2.2) is well defined and is given by

$$V_1(\theta) := \mathbb{E}_{\mathbb{P}} \left[(F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle)^2 \right] - C^2.$$

Thanks to the quadratic form of $V_1(\theta)$ in θ , it can easily be shown that V_1 is minimized uniquely at

$$\theta^* := \mathbb{E}_{\mathbb{P}} \left[(X_T - \mathbb{E}_{\mathbb{P}}[X_T])(X_T - \mathbb{E}_{\mathbb{P}}[X_T])' \right]^{-1} \mathbb{E}_{\mathbb{P}} [F(X)(X_T - \mathbb{E}_{\mathbb{P}}[X_T])], \quad (2.4)$$

where the matrix inverse above is well defined since for any $y \in \mathbb{R}_0^d$,

$$y' \mathbb{E}_{\mathbb{P}} \left[(X_T - \mathbb{E}_{\mathbb{P}}[X_T])(X_T - \mathbb{E}_{\mathbb{P}}[X_T])' \right] y = \mathbb{E}_{\mathbb{P}} \left[\langle y, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle^2 \right] > 0. \quad (2.5)$$

The resulting minimal variance is then given by

$$V_1(\theta^*) = V_1(0) - \langle \mathbb{E}_{\mathbb{P}} [F(X)(X_T - \mathbb{E}_{\mathbb{P}}[X_T])], \theta^* \rangle,$$

and the conditions (2.3) and (2.5) ensure that $V_1(\theta^*) < V_1(0)$.

2.2 Importance sampling variance reduction via Esscher transform

Fix $\lambda \in \Lambda_1$ and define the probability measure $\mathbb{Q}_\lambda$ via the Radon-Nikodym derivative, $\mathbb{P}$ -a.s.,

$$\frac{d\mathbb{Q}_\lambda}{d\mathbb{P}} \Big|_{\mathcal{F}_T} = \frac{e^{\langle \lambda, X_T \rangle}}{\mathbb{E}_{\mathbb{P}} [e^{\langle \lambda, X_T \rangle}]} = e^{\langle \lambda, X_T \rangle - \varphi_T(\lambda)}.$$

Then, the stochastic process $\{X_t : t \in [0, T]\}$ under $\mathbb{Q}_\lambda$ is again a Lévy process generated by $(\gamma_\lambda, 0, \nu_\lambda)$ where $\nu_\lambda(dz) = e^{\langle \lambda, z \rangle} \nu(dz)$, and $\gamma_\lambda = \gamma + \int_{\|z\| \leq 1} z(\nu_\lambda - \nu)(dz)$, and the probability measures $\mathbb{P}$ and $\mathbb{Q}_\lambda$ are mutually absolutely continuous. We also get for each $t \in [0, T]$, $\mathbb{E}_{\mathbb{Q}_\lambda} [e^{-\langle \lambda, X_t \rangle}] < +\infty$, and $(d\mathbb{P}/d\mathbb{Q}_\lambda)|_{\mathcal{F}_T} = ((d\mathbb{Q}_\lambda/d\mathbb{P})|_{\mathcal{F}_T})^{-1} = e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)}$, $\mathbb{Q}_\lambda$ -a.s. For details of the density transformation, see Section 33 [22].

The importance sampling method we will consider in this paper is based upon the equality

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[F(X)] &= \mathbb{E}_{\mathbb{Q}_\lambda} \left[\frac{d\mathbb{P}}{d\mathbb{Q}_\lambda} \Big|_{\mathcal{F}_T} F(X) \right] \\ &= \mathbb{E}_{\mathbb{Q}_\lambda} \left[\left(\frac{d\mathbb{Q}_\lambda}{d\mathbb{P}} \Big|_{\mathcal{F}_T} \right)^{-1} F(X) \right] \\ &= \mathbb{E}_{\mathbb{Q}_\lambda} \left[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} F(X) \right]. \end{aligned} \quad (2.6)$$

Define a set

$$\Lambda_2 := \left\{ \lambda \in \mathbb{R}^d : \int_{\|z\| > 1} \left(e^{\langle \lambda, z \rangle} \vee e^{-2\langle \lambda, z \rangle} \right) \nu(dz) < +\infty \right\} \subseteq \Lambda_1,$$

and suppose that $\text{Leb}(\Lambda_2) \neq 0$. We can easily show that the set Λ_2 is convex and contains the origin. For $\lambda \in \Lambda_2$, the variance of the random variable inside the expectation under the probability

measure $\mathbb{Q}_\lambda$ can also be written in terms of the expectation under $\mathbb{P}$ as

$$\begin{aligned} V_2(\lambda) &:= \mathbb{E}_{\mathbb{Q}_\lambda} \left[\left(\frac{d\mathbb{P}}{d\mathbb{Q}_\lambda} \Big|_{\mathcal{F}_T} \right)^2 F(X)^2 \right] - C^2 \\ &= \mathbb{E}_{\mathbb{P}} \left[\left(\frac{d\mathbb{Q}_\lambda}{d\mathbb{P}} \Big|_{\mathcal{F}_T} \right)^{-1} F(X)^2 \right] - C^2 \\ &= \mathbb{E}_{\mathbb{P}} \left[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} F(X)^2 \right] - C^2. \end{aligned}$$

The above expectation is well defined for $\lambda \in \Lambda_2$ since by the Cauchy-Schwartz inequality,

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

Also, define a convex set

$$\Lambda_3 := \Lambda_2 \cap \left\{ \lambda \in \mathbb{R}^d : \mathbb{E}_{\mathbb{P}} \left[\|X_T\|^k e^{-2\langle \lambda, X_T \rangle} F(X)^4 \right] < +\infty, k = 0, 2 \right\}.$$

Note that it contains the origin. Assume that $\text{Leb}(\Lambda_3) \neq 0$. In this setting, it is proved in [16] that a suitable stochastic approximation algorithm converges almost surely to the optimal parameter attaining $\min_{\lambda \in \Lambda_3} V_2(\lambda)$.

3 Main framework

In this section, we investigate the two-fold optimality in the combination of the control variates and the importance sampling, together with appropriate stochastic approximation algorithms for the optimal parameter search.

3.1 Combined control variates and importance sampling

The combination of the two variance reduction techniques is formulated as follows: For $\theta \in \mathbb{R}^d$ and $\lambda \in \Lambda_2$,

$$\begin{aligned} C = \mathbb{E}_{\mathbb{P}}[F(X)] &= \mathbb{E}_{\mathbb{P}}[F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle] \\ &= \mathbb{E}_{\mathbb{Q}_\lambda} \left[\frac{d\mathbb{P}}{d\mathbb{Q}_\lambda} \Big|_{\mathcal{F}_T} (F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle) \right] \\ &= \mathbb{E}_{\mathbb{Q}_\lambda} \left[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} (F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle) \right]. \end{aligned} \quad (3.1)$$

Again for $\theta \in \mathbb{R}^d$ and $\lambda \in \Lambda_2$, the variance of the component inside the expectation (3.1) can also be written in terms of the expectation under $\mathbb{P}$ as

$$\begin{aligned} V_3(\theta, \lambda) &:= \mathbb{E}_{\mathbb{Q}_\lambda} \left[\left(\frac{d\mathbb{P}}{d\mathbb{Q}_\lambda} \Big|_{\mathcal{F}_T} \right)^2 (F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle)^2 \right] - C^2 \\ &= \mathbb{E}_{\mathbb{P}} \left[\left(\frac{d\mathbb{Q}_\lambda}{d\mathbb{P}} \Big|_{\mathcal{F}_T} \right)^{-1} (F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle)^2 \right] - C^2 \\ &= \mathbb{E}_{\mathbb{P}} \left[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} (F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle)^2 \right] - C^2. \end{aligned}$$

It is easy to show that the above expectation is well defined using the Cauchy-Schwartz inequality. Clearly, $V_3(0, 0) = V_1(0) = V_2(0)$, $V_3(\theta, 0) = V_1(\theta)$, and $V_3(0, \lambda) = V_2(\lambda)$. Before proceeding our discussion, let us further define a convex set

$$\Lambda_4 := \left\{ \lambda \in \mathbb{R}^d : \int_{\|z\| > 1} \left(e^{\langle \lambda, z \rangle} \vee e^{-2\langle \lambda, z \rangle} \|z\|^6 \right) \nu(dz) < +\infty, \mathbb{E}_{\mathbb{P}} \left[e^{-2\langle \lambda, X_T \rangle} F(X)^4 \|X_T\|^2 \right] < +\infty \right\}.$$

Clearly, it contains the origin, and $\Lambda_4 \subseteq \Lambda_3$. We impose the assumption $\text{Leb}(\Lambda_4) \in (0, +\infty)$ in what follows.

3.2 Optimal parameter search with two-time-scale stochastic approximation algorithm

Prior to the construction of two-time-scale stochastic approximation algorithm, let us prepare some notations. Define $G_1 : \mathbb{R}^d \times \Lambda_4 \times \mathbb{D}([0, T]; \mathbb{R}^d) \mapsto \mathbb{R}^d$ and $G_2 : \mathbb{R}^d \times \Lambda_4 \times \mathbb{D}([0, T]; \mathbb{R}^d) \mapsto \mathbb{R}^d$ by

$$G_1(\theta, \lambda, X) := e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} (F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle)^2 (\nabla_\lambda \varphi_T(\lambda) - X_T),$$

and

$$G_2(\theta, \lambda, X) := -2e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} (F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle) (X_T - \mathbb{E}_{\mathbb{P}}[X_T]),$$

so that $\nabla_\lambda V_3(\theta, \lambda) = \mathbb{E}_{\mathbb{P}}[G_1(\theta, \lambda, X)]$ and $\nabla_\theta V_3(\theta, \lambda) = \mathbb{E}_{\mathbb{P}}[G_2(\theta, \lambda, X)]$. Moreover, define

$$\begin{aligned} h(\lambda) &:= \\ &\mathbb{E}_{\mathbb{P}} \left[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} (X_T - \mathbb{E}_{\mathbb{P}}[X_T]) (X_T - \mathbb{E}_{\mathbb{P}}[X_T])' \right]^{-1} \mathbb{E}_{\mathbb{P}} \left[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} F(X) (X_T - \mathbb{E}_{\mathbb{P}}[X_T]) \right], \end{aligned} \tag{3.2}$$

where the matrix inverse above is well defined due to its positive-definiteness, and for $\lambda \in \Lambda_4$, define

$$\begin{aligned} V_4(\lambda) &:= V_3(h(\lambda), \lambda) \\ &= V_3(0, \lambda) - \left\langle h(\lambda), \mathbb{E}_{\mathbb{P}} \left[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} F(X) (X_T - \mathbb{E}_{\mathbb{P}}[X_T]) \right] \right\rangle < V_3(0, \lambda), \end{aligned}$$

and define

$$\lambda^\dagger := \underset{\lambda \in H}{\operatorname{argmin}} V_4(\lambda),$$

where H is a compact subset of Λ_4 with $\{0\} \in H$. Also, let $\{H_n\}_{n \in \{0\} \cup \mathbb{N}}$ be an increasing sequence of compact sets in $\mathbb{R}^d$.

The two-time-scale stochastic approximation algorithm we consider is then formulated as follows:

$$\lambda_{n+1} = \Pi_H [\lambda_n - \varepsilon_n G_1(\theta_n, \lambda_n, X_{n+1})], \quad (3.3)$$

$$\theta_{n+1} = \Pi_{H_{\sigma(n)}} [\theta_n - \delta_n G_2(\theta_n, \lambda_n, X_{n+1})], \quad (3.4)$$

where $\sigma(n)$ is the number counter of projections up to the n -th step, and where $\{\varepsilon_n\}_{n \in \{0\} \cup \mathbb{N}}$ and $\{\delta_n\}_{n \in \{0\} \cup \mathbb{N}}$ are sequences of non-increasing positive constants satisfying

$$\sum_{n=0}^{+\infty} \varepsilon_n = +\infty, \quad \sum_{n=0}^{+\infty} \varepsilon_n^2 < +\infty, \quad \sum_{n=0}^{+\infty} \delta_n = +\infty, \quad \sum_{n=0}^{+\infty} \delta_n^2 < +\infty,$$

and

$$\lim_{n \uparrow +\infty} \frac{\varepsilon_n}{\delta_n} = 0, \quad (3.5)$$

that is, the iteration (3.3) moves much slower than the other (3.4). The following is the main result of this section.

Theorem 3.1. *The sequence $\{(\theta_n, \lambda_n)\}_{n \in \mathbb{N}}$ defined in the iteration (3.3)-(3.4) converges $\mathbb{P}$ -a.s. to $(h(\lambda^\dagger), \lambda^\dagger)$, with $\lim_{n \uparrow +\infty} \sigma(n) < +\infty$, $\mathbb{P}$ -a.s.*

This can be proved directly by the results of [6] together with the almost sure convergence of the two related *single-time-scale* stochastic approximation algorithms, appearing in the following. (For the proof of Theorem 3.1 and Proposition 3.2, see Appendix A.)

Proposition 3.2. *(i) Fix $\lambda \in \Lambda_4$. Then, $h(\lambda)$ uniquely solves the equation $\nabla_\theta V_3(\theta, \lambda) = 0$. Moreover, the sequence $\{\theta_n\}_{n \in \mathbb{N}}$ defined in the iteration*

$$\theta_{n+1} = \Pi_{H_{\sigma(n)}} [\theta_n - \delta_n G_2(\theta_n, \lambda, X_{n+1})] \quad (3.6)$$

converges $\mathbb{P}$ -a.s. to $h(\lambda)$, with $\lim_{n \uparrow +\infty} \sigma(n) < +\infty$, $\mathbb{P}$ -a.s.

(ii) The point $\lambda^\dagger$ is unique, and the sequence $\{\lambda_n\}_{n \in \mathbb{N}}$ defined in the iteration

$$\lambda_{n+1} = \Pi_H [\lambda_n - \varepsilon_n G_1(h(\lambda_n), \lambda_n, X_{n+1})] \quad (3.7)$$

converges $\mathbb{P}$ -a.s. to $\lambda^\dagger$.

The two-time-scale stochastic approximation algorithm (3.3)-(3.4) can be interpreted as follows. The fast θ sees λ as “quasi-static”, and thus the algorithm (3.4) has an effect similar to fixing λ and running the (single-time-scale) algorithm (3.6) for a long time. Meanwhile, in the algorithm (3.3), θ_n can be seen as a close approximation of $h(\lambda_n)$, that is, “equilibrated”, and thus almost same as running the (single-time-scale) algorithm (3.7).

Remark 3.3. It is easy to show that the matrix in (3.2) before the inverse operation is the variance-covariance matrix of the random vector $e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)}(X_T - \mathbb{E}_{\mathbb{P}}[X_T])$ under the probability measure $\mathbb{Q}_{\lambda}$. This is useful to ease the numerical computation of $h(\lambda)$. Apart from that, we can also derive

$$h(\lambda) = \mathbb{E}_{\mathbb{Q}_{-\lambda}} [(X_T - \mathbb{E}_{\mathbb{P}}[X_T])(X_T - \mathbb{E}_{\mathbb{P}}[X_T])']^{-1} \mathbb{E}_{\mathbb{Q}_{-\lambda}} [F(X)(X_T - \mathbb{E}_{\mathbb{P}}[X_T])],$$

and can infer that the vector $\mathbb{E}_{\mathbb{Q}_{-\lambda}} [F(X)(X_T - \mathbb{E}_{\mathbb{P}}[X_T])]$ vanishes as the parameter λ provides more effective importance sampling, since then $F(X)$ tends to be nonzero under the probability measure $\mathbb{Q}_{\lambda}$, while the expectation is taken on contrary under $\mathbb{Q}_{-\lambda}$. $\square$

Remark 3.4. Our theoretical developments lay exactly in the framework of the envelope theorem, which often appears in the microeconomics literature and which provides, in our case, the derivative formula $\nabla_{\lambda} V_3(h(\lambda), \lambda) = [\nabla_{\lambda} V_3(\theta, \lambda)]|_{\theta=h(\lambda)}$. For its details, see, for example, Bardi and Capuzzo-Dolcetta [4]. $\square$

4 Adaptive Monte Carlo variance reduction with gamma process

Recall that for a standard normal random variable W and for $\lambda \in \mathbb{R}$, the Girsanov theorem gives $\mathbb{Q}_{\lambda}(W \leq x) = \mathbb{P}(W + \lambda \leq x)$, $x \in \mathbb{R}$. Thus, W under $\mathbb{Q}_{\lambda}$ can even be generated as $W + \lambda$ under $\mathbb{P}$, and the importance sampling parameter search procedure can thus be embedded in adaptive Monte Carlo procedure, without changing the probability measure $\mathbb{Q}_{\lambda}$ at each step of a parameter search algorithm for λ . Meanwhile, general infinitely divisible distributions do not possess such a convenient feature, and so from a practical point of view, it is usually not clever to extend to the adaptive Monte Carlo procedure.

An exception is the *gamma distribution*, and its associated *gamma (Lévy) process*. Let us recall their definition and key properties. (For more details, see also Applebaum [1] and [22].) The gamma process is a (pure-jump) Lévy process in $[0, +\infty)$ whose Lévy measure is given in the form

$$\nu(dz) = a \frac{e^{-bz}}{z} dz, \quad z \in (0, +\infty),$$

where $a > 0$ and $b > 0$. For each (time) $t > 0$, its marginal distribution is characterized by the characteristic function

$$\mathbb{E}_{\mathbb{P}}[e^{iyX_t}] = \exp \left[t \int_{(0,+\infty)} (e^{iyz} - 1) \nu(dz) \right] = \left(1 - \frac{iy}{b} \right)^{-at}, \quad (4.1)$$

while the density function $f_{X_t}^{\mathbb{P}}$ of the marginal distribution at time t under $\mathbb{P}$ is given in closed form

$$f_{X_t}^{\mathbb{P}}(x) = \frac{b^{at}}{\Gamma(at)} x^{at-1} e^{-bx}, \quad x \in [0, +\infty).$$

Moreover, the first three (centered) polynomial moments are given by

$$\mathbb{E}_{\mathbb{P}}[X_t] = t \frac{a}{b}, \quad \text{Var}(X_t) = t \frac{a}{b^2}, \quad \text{and} \quad \mathbb{E}_{\mathbb{P}}[(X_t - \mathbb{E}_{\mathbb{P}}[X_t])^3] = t \frac{2a}{b^3},$$

The following is an important (exponential and polynomial) moment property of the distribution, which can be easily shown with the density function $f_{X_t}^{\mathbb{P}}$, or by Theorem 2.1.

Lemma 4.1. *Let $\{X_t : t \geq 0\}$ be a gamma process in $\mathbb{R}$ having the characteristic function (4.1). Then, for each $t > 0$, the expectation $\mathbb{E}_{\mathbb{P}}[X_t^\alpha e^{\beta X_t}]$ is finite if and only if $\alpha > -a$ and $\beta < b$.*

We will henceforth assume that there exists a positive constant c such that $\mathbb{P}(|F(X)| \leq c) = 1$. This almost-sure boundedness condition is in general very restrictive while not quite so in credit derivatives pricing, which will be seen later.

We extend this to a multivariate setting. Let $\{X_t : t \in [0, T]\}$ be a Lévy process in $\mathbb{R}^d$, of which components are independent, but not necessarily identically distributed, gamma processes, that is, its Lévy measure is supported on the positive axes of $\mathbb{R}^d$. Moreover, we characterize the gamma process in the k -th dimension of $\{X_t : t \in [0, T]\}$ by the parameter (a_k, b_k) for (4.1). With the help of Lemma 4.1, we can easily derive

$$\begin{aligned} \Lambda_1 &= \left\{ \lambda \in \mathbb{R}^d : \lambda_k \in (-\infty, +b_k), \text{ for every } k \right\}, \\ \Lambda_2 &= \left\{ \lambda \in \mathbb{R}^d : \lambda_k \in (-b_k, +b_k), \text{ for every } k \right\}, \end{aligned}$$

and

$$\Lambda_3 = \Lambda_4 = \left\{ \lambda \in \mathbb{R}^d : \lambda_k \in \left(-\frac{1}{2}b_k, +b_k \right), \text{ for every } k \right\}.$$

For $\lambda \in \Lambda_1$, we have

$$\varphi_T(\lambda) := T \sum_{k=1}^d a_k \ln \frac{b_k}{b_k - \lambda_k},$$

and

$$\nabla_{\lambda} \varphi_T(\lambda) = T \left[\frac{a_1}{b_1 - \lambda_1}, \dots, \frac{a_d}{b_d - \lambda_d} \right]'$$

Moreover, we can define the probability measure $\mathbb{Q}_\lambda$ via the Esscher transform and the stochastic process $\{X_t : t \in [0, T]\}$ under $\mathbb{Q}_\lambda$ is again a gamma process with the Lévy measure

$$\nu_\lambda(dz) = e^{\lambda z} \nu(dz) = a \frac{e^{-(b-\lambda)z}}{z} dz,$$

the characteristic function

$$\mathbb{E}_{\mathbb{Q}_\lambda} [e^{iyX_t}] = \exp \left[t \int_{(0, +\infty)} (e^{iyz} - 1) \nu_\lambda(dz) \right] = \left(1 - \frac{iy}{b-\lambda} \right)^{-at},$$

and the density function

$$f_{X_t}^{\mathbb{Q}_\lambda}(x) = \frac{e^{\lambda x}}{\mathbb{E}_{\mathbb{P}}[e^{\lambda X_t}]} f_{X_t}^{\mathbb{P}}(x) = \frac{(b-\lambda)^{at}}{\Gamma(at)} x^{at-1} e^{-(b-\lambda)x}, \quad x \in [0, +\infty).$$

The following is a key property for our purpose.

Lemma 4.2. Fix $\lambda \in \Lambda_1$. For $x \in [0, +\infty)$ and for $t \in [0, T]$, $\mathbb{Q}_\lambda(X_t \leq x) = \mathbb{P}\left(\frac{b}{b-\lambda}X_t \leq x\right)$. Moreover, it holds

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}_\lambda} \left[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} (F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle) \right] \\ = \mathbb{E}_{\mathbb{P}} \left[e^{-\langle \lambda, \frac{b}{b-\lambda}X_T \rangle + \varphi_T(\lambda)} \left(F\left(\frac{b}{b-\lambda}X\right) - \left\langle \theta, \frac{b}{b-\lambda}X_T - \mathbb{E}_{\mathbb{P}}[X_T] \right\rangle \right) \right]. \end{aligned}$$

Proof. The former claim is obvious by the structure of the density functions $f_{X_t}^{\mathbb{P}}$ and $f_{X_t}^{\mathbb{Q}_\lambda}$, while the latter is a direct application of the former. $\square$

We are now in a position to formulate our adaptive Monte Carlo procedure. Define $Y : \mathbb{R}^d \times \Lambda_4 \times \mathbb{D}([0, T]; \mathbb{R}^d) \mapsto \mathbb{R}$ by

$$Y(\theta, \lambda, X) := e^{-\langle \lambda, \frac{b}{b-\lambda}X_T \rangle + \varphi_T(\lambda)} \left[F\left(\frac{b}{b-\lambda}X\right) - \left\langle \theta, \frac{b}{b-\lambda}X_T - \mathbb{E}_{\mathbb{P}}[X_T] \right\rangle \right],$$

with its empirical mean and empirical variance defined respectively by

$$EM_n := \frac{1}{n} \sum_{k=1}^n Y(\theta_{k-1}, \lambda_{k-1}, X_k),$$

and

$$EV_n^2 := \frac{1}{n} \sum_{k=1}^n Y^2(\theta_{k-1}, \lambda_{k-1}, X_k) - EM_n^2.$$

The following verifies these formulation.

Theorem 4.3. *If the iterations (3.3)-(3.4) are performed with $H \subseteq \Lambda_4$, then*

$$\mathbb{P} \left[\lim_{n \uparrow +\infty} EM_n = \mathbb{E}_{\mathbb{P}}[F(X)] \right] = 1,$$

and as $n \uparrow +\infty$,

$$\sqrt{n}(EM_n - \mathbb{E}_{\mathbb{P}}[F(X)]) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, V_4(\lambda^\dagger) \right),$$

under the probability measure $\mathbb{P}$.

Proof. See Appendix A. □

Moreover, set

$$\Lambda_5 := \left\{ \lambda \in \mathbb{R}^d : \lambda_k \in \left(-\frac{1}{3}b_k, +b_k \right), \text{ for every } k \right\}.$$

The following provides useful criteria for the performance analysis of our method.

Proposition 4.4. *If the iterations (3.3)-(3.4) are performed with $H \subseteq \Lambda_5$, then*

$$\mathbb{P} \left[\lim_{n \uparrow +\infty} EV_n^2 = V_4(\lambda^\dagger) \right] = 1,$$

and as $n \uparrow +\infty$,

$$\sqrt{n} \frac{EM_n - \mathbb{E}_{\mathbb{P}}[F(X)]}{EV_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1),$$

under the probability measure $\mathbb{P}$.

Proof. See Appendix A. □

5 Numerical illustrations: credit derivatives pricing

Monte Carlo simulation for credit derivatives pricing is a challenging problem in the sense that (i) the default is usually a very rare event, that is, the pricing problem is basically concerned with the probability tails, and (ii) credit derivative have a complex structure with the default dependence among consisting credit entities, the digital indicator, the maximum, and the minimum, that is, the closed-form pricing is basically swept out of scope. See, for example, Schönbucher [23] for general details of credit derivatives pricing models.

In this paper, we restrict ourselves to the collateralized debt obligation (CDO) tranches, which are among most important credit derivatives structures. Some numerical methods for the CDO pricing with the *Gaussian copula* model of Li [18], which is the *de facto* market standard model, have been proposed to get around the use of Monte Carlo simulations by putting some simplification assumptions on the structure in force, for example, every credit entity has a common recovery rate, the payoff is insensitive to the order credit entities default in, and so on. (See, for example,

Gregory and Laurent [10] and Hull and White [11].) It however seems that without such simplification assumptions, we need to resort to Monte Carlo simulations.

In our numerical experiments, we put many assumptions only for simplicity in arguments (never for necessity);

- (i) every accrual period is constant Δ ,
- (ii) the continuously compounded risk-free rate is constant r ,
- (iii) all credit entities have a common recovery rate $R \in (0, 1)$, and
- (iv) CDS spreads are common for every entity and are constant for every maturity.

These assumptions are very unrealistic in practice, but do not affect any fundamentals of Monte Carlo pricing procedure.

Now, let $0 \equiv t_0 < t_1 < \dots < t_N \equiv T$ be time horizon such that the accrual period $\Delta = t_n - t_{n-1}$ is common for all n . In this paper, we consider the most liquidly traded quarter-annual 5-year maturity, that is, $T = 5.25$, $\Delta = 0.25$, and $N = 21$. Let $p_{k,n}$ be the survival probability for entity k at maturity t_n . Then, the CDS par spread $s_{k,n}$ of entity k for maturity t_n is computed (with discrete approximation) via

$$s_{k,n} = \frac{\sum_{l=1}^n e^{-rt_l} \Delta_k (p_{k,l-1} + p_{k,l}) / 2}{(1-R) \sum_{l=1}^n e^{-rt_l} (p_{k,l-1} - p_{k,l})}, \quad n = 1, \dots, N.$$

With the simplifications $\Delta_n \equiv \Delta > 0$ and $s_{k,n} \equiv s > 0$, it is easy to derive recursively that for every credit entity k ,

$$p_{k,n} = \left(\frac{1 - R - s\Delta/2}{1 - R + s\Delta/2} \right)^n, \quad n = 0, 1, \dots, N.$$

We consider a basket of 100 credit entities. Let L_n be the cumulative percentage loss of the credit basket at maturity t_n , and let $L_n(A, B)$ be its cumulative percentage loss “to the A - B % tranche” at maturity t_n , that is,

$$L_n(A, B) := ((1-R)L_n - A)_+ - ((1-R)L_n - B)_+ = ((1-R)L_n \wedge B - A)_+. \quad (5.1)$$

We defer the precise formulation of L_n 's, since it is quite different for different models. Now, the protection leg for the A - B % tranche is given by

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left[\sum_{n=1}^N e^{-rt_n} (L_n(A, B) - L_{n-1}(A, B)) \right] &= \sum_{n=1}^N e^{-rt_n} \mathbb{E}_{\mathbb{P}} [L_n(A, B)] - \sum_{n=1}^N e^{-rt_n} \mathbb{E}_{\mathbb{P}} [L_{n-1}(A, B)] \\ &=: C_1 - C_2, \end{aligned}$$

multiplied by the tranche notional, while the premium leg for the A - B % tranche is

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left[\sum_{n=1}^N e^{-rt_n} \Delta ((B - A) - L_n(A, B)) \right] &= \sum_{n=1}^N e^{-rt_n} \Delta (B - A) - \Delta \sum_{n=1}^N e^{-rt_n} \mathbb{E}_{\mathbb{P}} [L_n(A, B)] \\ &= \sum_{n=1}^N e^{-rt_n} \Delta (B - A) - \Delta C_1, \end{aligned}$$

multiplied by the tranche notional and the coupon paid on the tranche. Hence, the CDO tranche pricing problem reduces to the computation of the above two quantities C_1 and C_2 . As can be seen, the structures of C_1 and C_2 are fundamentally very similar, while we will consider both in our numerical experiments for better illustration.

5.1 Gamma copula model

Many distributions have been proposed to replace the normal distribution of the Gaussian copula model to describe the heavy probability tails and the stronger tail dependence, for example, the student t -distribution in O’Kane and Schloegl [20], the double t -distribution in [11], and the normal inverse Gaussian distribution in Kalemánova, Schmid and Werner [13]. In this paper, we choose the gamma distribution and name it the *Gamma copula*.

Let $\{X_t : t \geq 0\}$ be a gamma process in $\mathbb{R}$ with parameter (a, b) in (4.1) and $\{Y_t^k : t \geq 0\}_{k \in \mathbb{N}}$ be a sequence of iid copies of $\{X_t : t \geq 0\}$. Fix $\rho \in (0, 1)$. Suppose that each component of the following d -dimensional random vector

$$\begin{bmatrix} Z^1 \\ \vdots \\ Z^d \end{bmatrix} := X_\rho + \begin{bmatrix} Y_{1-\rho}^1 \\ \vdots \\ Y_{1-\rho}^d \end{bmatrix}$$

corresponds to d credit entities. The formulation of this type, first proposed in Baxter [5], intends two nice features;

- (i) The distribution is completely closed under convolution (and does not have to be the gamma distribution), that is, for every k , the law of the random variable Z^k is identical to that of X_1 , and
- (ii) The correlation among credit entities is controlled by the single parameter ρ , that is, for $k_1 \neq k_2$, $\text{Corr}(Z^{k_1}, Z^{k_2}) = \rho$.

Remark 5.1. Notice that the random vector $(Z^1, \dots, Z^d)'$ is infinitely divisible *componentwise* as mentioned in (i) just above, while *it is not infinitely divisible random vector* since its characteristic function does not follow the Lévy-Khintchine formula. (The Gaussian copula is an exception.) Hence, the term “*copula*” in this example should not be confused with the Lévy copulas which are intended to describe the dependence structure of a multivariate Lévy measure. (See, for example, Kallsen and Tankov [15] for Lévy copulas.) $\square$

Define now the threshold function $z : [0, +1] \mapsto [0, +\infty)$ by

$$z(p) := \inf \left\{ c > 0 : \int_0^c f_{X_1}^{\mathbb{P}}(x) dx > p \right\}.$$

In the Gamma copula model, we say that the entity k has been in default at maturity t_n if

$$Z^k \in (z(p_{k,n}), +\infty).$$

We can also say that the entity k falls into default between $[t_{n-1}, t_n)$ if $Z^k \in (z(p_{k,n}), z(p_{k,n-1}))$. Hence, the cumulative percentage loss at maturity t_n is written as

$$L_n := 100\% \times \frac{1}{d} \sum_{k=1}^d \mathbb{1} \left(Z^k \in (z(p_{k,n}), +\infty) \right).$$

For simplicity and illustration purpose, in numerical experiments for this model, we will attempt variance reduction through only the global latent random variable X_ρ and without any consideration on the idiosyncratic components $Y_{1-\rho}^1, \dots, Y_{1-\rho}^d$. More precisely, the variance reduction for C_1 is based on

$$C_1 = \mathbb{E}_{\mathbb{P}} \left[e^{-\lambda X_\rho + \varphi_\rho(\lambda)} \left(\sum_{n=1}^N e^{-rt_n} L_n(A, B) - \theta (X_\rho - \mathbb{E}_{\mathbb{P}}[X_\rho]) \right) \right],$$

and that for C_2 can be written in a similar manner.

We report in Table 1 the numerical results for the Gamma copula model for $\rho = 0.10, 0.30$ and 0.50 where the parameters of the gamma distribution are fixed at $(a, b) = (2, 1)$. We generate $N(=5e+4)$ iid replications for plain Monte Carlo simulations and compute estimates for C_1 or C_2 (price) and for the variance $V_3(0, 0)$. We perform optimal parameter search, independently for C_1 and C_2 , and compute estimates for the variance $V_3(\theta_N, \lambda_N)$, using only the final point (θ_N, λ_N) of our two-time-scale stochastic approximation algorithms with

$$\delta_n = \frac{\alpha_\delta}{(n+1)^{7/11}}, \quad \varepsilon_n = \frac{\alpha_\varepsilon}{(n+1)^{10/11}}.$$

Moreover, we conduct adaptive Monte Carlo simulations along with the parameter search algorithms, and compute estimates for EV_N^2 . With those estimates, we compute two variance ratios and a normalized standard deviation as follows;

$$(\text{vratio1}) := \frac{V_3(0, 0)}{V_3(\theta_N, \lambda_N)}, \quad (\text{vratio2}) := \frac{V_3(0, 0)}{EV_N^2}, \quad (\text{nstdev}) := \frac{\sqrt{V_3(0, 0)}}{(\text{price})}.$$

To show the effectiveness of our method, we provide in Figure 1 some numerical illustrations in the case $\rho = 0.30$ for 12-22% tranche.

Concerning the total computation time, we need for one replication to generate $d+1(=101)$ independent gamma random variables and compare those with thresholds for $N(=21)$ maturities, while our stochastic approximation algorithms deal with only 1 gamma random variable. Although depending on the way of coding, we have observed that the addition of the stochastic approximation does not require very big computational cost. It should thus be safe to say that our adaptive Monte Carlo method succeeds in improving the rate of convergence.

5.2 Intensity Gamma model

The intensity Gamma model of Joshi and Stacey [12] is the one describing the survival probability as a stochastic process as follows. Let $\{X_t : t \geq 0\}$ and $\{Y_t^k : t \geq 0\}_{k \in \mathbb{N}}$ be as defined for the Gamma copula model of the last subsection. With those, fix $\rho \in [0, +1]$, and define a d -dimensional stochastic process $\{Z_t : t \geq 0\}$ by

$$\begin{bmatrix} Z_t^1 \\ \vdots \\ Z_t^d \end{bmatrix} := X_{\rho t} + \begin{bmatrix} Y_{(1-\rho)t}^1 \\ \vdots \\ Y_{(1-\rho)t}^d \end{bmatrix}.$$

Clearly, each component of $\{Z_t : t \geq 0\}$ is a Lévy process since the sum of two independent Lévy process is a Lévy process, and its marginals have a gamma distribution since for every k and for each $t > 0$, the laws of Z_t^k and X_t are identical. Thus, by a theorem of Skorokhod (see Theorem 15.17 of Kallenberg [14]), it holds that each component of $\{Z_t : t \geq 0\}$ is a gamma process whose finite dimensional distributions are identical to those of $\{X_t : t \geq 0\}$. (In a similar argument to Remark 5.1, the stochastic process $\{Z_t : t \geq 0\}$ is *not a multivariate Lévy process*.) Then, the survival probability for each entity at time t is modeled as

$$\begin{bmatrix} P_t^1 \\ \vdots \\ P_t^d \end{bmatrix} = \begin{bmatrix} e^{-\int_0^t c_1(s) dZ_s^1} \\ \vdots \\ e^{-\int_0^t c_d(s) dZ_s^d} \end{bmatrix},$$

where $\{c_k(t) : t \geq 0\}_{k \in \mathbb{N}}$ is a sequence of non-negative bounded functions on $[0, +\infty)$. In most practical cases, it is enough to set it to be piecewise constant as $c_k(t) \equiv c_{k,l} > 0$ for $t \in [t_{l-1}, t_l)$, and then we get

$$\mathbb{E}_{\mathbb{P}}[P_{t_n}^k] = \mathbb{E}_{\mathbb{P}}[e^{-\sum_{l=1}^n c_{k,l}(Z_{t_l}^k - Z_{t_{l-1}}^k)}] = \prod_{l=1}^n \mathbb{E}_{\mathbb{P}}[e^{-c_{k,l} X_1}]^{\Delta} = \prod_{l=1}^n \left(1 + \frac{c_{k,l}}{b}\right)^{-a\Delta}.$$

Solving the equation $\mathbb{E}_{\mathbb{P}}[P_{t_n}^k] = p_{k,n}$ for $c_{k,l}$, we get for every l ,

$$c_{k,l} \equiv b \left[\left(\frac{1 - R - s_k \Delta / 2}{1 - R + s_k \Delta / 2} \right)^{-\frac{1}{a\Delta}} - 1 \right].$$

In this setting, we say that the entity k has been in default at maturity t_n if

$$U^k \in [P_{t_n}^k, 1),$$

where $\{U^k\}_{k \in \mathbb{N}}$ is a sequence of iid uniform random variables on $(0, 1)$. Note that using the well known fact that the distribution of $-\ln U_1$ is exponential with unit parameter, the above criterion can also be

$$\sum_{l=1}^n c_{k,l} (Z_{t_l}^k - Z_{t_{l-1}}^k) \in [V_k, +\infty),$$

where $\{V^k\}_{k=\mathbb{N}}$ is a sequence of iid exponential random variables with unit parameter. Therefore, the cumulative percentage loss at maturity t_n is written as

$$L_n := 100\% \times \frac{1}{d} \sum_{k=1}^d \mathbb{1} \left(P_{t_n}^k \leq U^k \right).$$

For simplicity again, the variance reduction is modeled through the terminal value $X_{\rho T}$ of the global stochastic process only. More precisely, the variance reduction for C_1 is based on

$$C_1 = \mathbb{E}_{\mathbb{P}} \left[e^{-\lambda X_{\rho T} + \phi_{\rho T}(\lambda)} \left(\sum_{n=1}^N e^{-rt_n} L_n(A, B) - \theta \left(X_{\rho T} - \mathbb{E}_{\mathbb{P}} [X_{\rho T}] \right) \right) \right],$$

and that for C_2 can be written in a similar manner.

From a viewpoint of total computation time, we need for one replication to generate $(d + 1) \times N (= 101 \times 21)$ independent gamma random variables and compare those with thresholds for $N (= 21)$ maturities, while our stochastic approximation algorithms deals with 1-dimensional random variable $X_{\rho T}$. In this sense, compared to the case of the gamma copula, our stochastic approximation algorithms require very little additional computational cost. It should thus be safe to say that the variance ratios indicate a good improvement.

6 Concluding remarks

In this paper, we have developed and analyzed an application of the two-time-scale stochastic approximation algorithm in the optimal parameter search for the combined control variates and importance sampling in the Lévy process framework. We have proved that the algorithm converges almost surely to a unique root of the gradient of the variance in the sense of “equilibrated control variates and quasi-static importance sampling,” and the almost sure convergence guarantees the incorporation of the algorithms into the adaptive Monte Carlo variance reduction procedure. We have illustrated the effectiveness of our method through numerical experiments of the CDO tranche pricing with the gamma copula model and the intensity gamma model.

Let us close this study with some remarks. First, the parameter search with two-time-scale stochastic approximation itself works with a large class of Lévy processes, while the adaptive Monte Carlo method does not, since it relies greatly on a scaling property intrinsic to the gamma distribution described in Lemma 4.2. To our knowledge, there seems no other (non-Gaussian) infinitely divisible distribution for our purpose possessing such a scaling property. Regardless of the restriction to the gamma distribution and process, our adaptive Monte Carlo method still has considerable potential to yield great benefits in major applications. A good example is the queueing theory with Poisson arrivals or/and exponential service times, since the exponential distribution is in the class of the gamma distribution. Moreover, some option pricing application with the

variance gamma process of Madan *et al.* [19] is also within our scope since it can be expressed as the difference of two independent increasing gamma process.

Second, for numerical illustrations, we have only considered the CDO pricing, while our method does not rely on any intrinsic structures of the CDO tranche. For further example in the credit derivatives, more complex (and recently more popular) products, such as CDO-squared and CPDO, should also be very interesting objects for the use of our method. We expect that our method is useful as well in a wide variety of applications, other than the financial engineering.

Finally, the reader might have been wondered whether the opposite scheme of the two-time-scale stochastic approximation algorithm work as well, that is, with *equilibrated importance sampling and quasi-static control variates*. The answer would be “partially possible in theory”, while “perhaps negative in practice.” We discuss this point in Appendix B.

Acknowledgments

The author thanks an anonymous referee for detailed comments and various suggestions, which helped improve this paper. He is also grateful to Tatsuya Toshida and Masataka Okumura, both at Daiwa Securities SMBC Co.Ltd., for important discussions on credit derivatives.

A Proofs

Proof of Proposition 3.2 (i). It follows immediately from the quadratic form of $V_3(\theta, \lambda)$ in θ that $V_3(\theta, \lambda)$ is strictly convex in θ , and that $h(\lambda)$ is the unique solution of $\nabla_\theta V_3(\theta, \lambda) = 0$.

For the almost sure convergence of the sequence $\{\theta_n\}_{n \in \mathbb{N}}$ to $h(\lambda)$, following the results of Chen and Zhu [7] and Delyon [8], it suffices to show that for each $m \in \mathbb{N}$,

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

and this can be verified by $\mathbb{E}_{\mathbb{P}}[e^{-2\langle \lambda, X_T \rangle} F(X)^2 \|X_T\|^2] < +\infty$, and $\mathbb{E}_{\mathbb{P}}[e^{-2\langle \lambda, X_T \rangle} \langle \theta, X_T \rangle^2 \|X_T\|^2 \mathbb{1}(\|\theta\| \leq m)] < +\infty$. The first condition is straightforward by the definition of Λ_4 , while for the second,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}} \left[e^{-2\langle \lambda, X_T \rangle} \langle \theta, X_T \rangle^2 \|X_T\|^2 \mathbb{1}(\|\theta\| \leq m) \right] &\leq \mathbb{E}_{\mathbb{P}} \left[e^{-2\langle \lambda, X_T \rangle} \|\theta\|^2 \|X_T\|^4 \mathbb{1}(\|\theta\| \leq m) \right] \\ &\leq m^2 \mathbb{E}_{\mathbb{P}} \left[e^{-2\langle \lambda, X_T \rangle} \|X_T\|^4 \right] < +\infty, \end{aligned}$$

again by Theorem 2.1 and the definition of Λ_4 . □

Proof of Proposition 3.2 (ii). We first show that for $\lambda \in \Lambda_4$, the matrix $\nabla_\lambda [h(\lambda)']$ is well defined. For this, it suffices to show that the operation ∇_λ can be performed on both each element of the matrix $\mathbb{E}_{\mathbb{P}}[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} (X_T - \mathbb{E}_{\mathbb{P}}[X_T])(X_T - \mathbb{E}_{\mathbb{P}}[X_T])']$ and the transposed vector

$\mathbb{E}_{\mathbb{P}}[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} F(X)(X_T - \mathbb{E}_{\mathbb{P}}[X_T])']$. The former holds since the gradient for element (j, k) is well defined as

$$\mathbb{E}_{\mathbb{P}} \left[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} \left(X_T^j - \mathbb{E}_{\mathbb{P}}[X_T^j] \right) \left(X_T^k - \mathbb{E}_{\mathbb{P}}[X_T^k] \right) (\nabla_{\lambda} \varphi_T(\lambda) - X_T) \right],$$

while the latter holds true since the matrix

$$\mathbb{E}_{\mathbb{P}} \left[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} F(X) (\nabla_{\lambda} \varphi_T(\lambda) - X_T) (X_T - \mathbb{E}_{\mathbb{P}}[X_T])' \right]$$

is well defined. Hence, using the envelope theorem (or directly the chain rule of the gradient), we get

$$\nabla_{\lambda} V_4(\lambda) = \nabla_{\lambda} V_3(h(\lambda), \lambda) = [\nabla_{\lambda} V_3(\theta, \lambda)]|_{\theta=h(\lambda)}, \quad (\text{A.1})$$

where the right hand side is well defined since $\lambda \in \Lambda_4$. Moreover, using the chain rule, we have

$$\begin{aligned} \text{Hess}_{\lambda} [V_4(\lambda)] &= \nabla'_{\lambda} ([\nabla_{\lambda} V_3(\theta, \lambda)]|_{\theta=h(\lambda)}) \\ &= [\text{Hess}_{\lambda} V_3(\theta, \lambda) + \nabla'_{\theta} \nabla_{\lambda} V_3(\theta, \lambda)]|_{\theta=h(\lambda)} \\ &= [\text{Hess}_{\lambda} V_3(\theta, \lambda)]|_{\theta=h(\lambda)}, \end{aligned}$$

where the last equality follows from $\nabla'_{\theta} \nabla_{\lambda} (V_3(\theta, \lambda)) = (\nabla'_{\lambda} \nabla_{\theta} (V_3(\theta, \lambda)))'$ and from the definition of $h(\lambda)$. Finally, for any $y \in \mathbb{R}_0^d$,

$$\begin{aligned} &y' \text{Hess}_{\lambda} [V_4(\lambda)] y \\ &= \mathbb{E}_{\mathbb{P}} \left[(y' \text{Hess}_{\lambda} [\varphi_T(\lambda)] y + \langle y, \nabla_{\lambda} \varphi_T(\lambda) - X_T \rangle^2) e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} (F(X) - \langle h(\lambda), X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle)^2 \right], \end{aligned}$$

and this is strictly positive since $\text{Hess}_{\lambda} [\varphi_T(\lambda)]$ reduces to the variance-covariance matrix of the random vector X_T under the probability measure $\mathbb{Q}_{\lambda}$, which is positive-definite by definition. Hence, the function V_4 is strictly convex in λ on Λ_4 , and thus the uniqueness of $\lambda^{\dagger}$ follows. The argument so far ensures that the iteration (3.7) is a gradient-based stochastic approximation algorithm for a strictly convex function. Hence, as in the proof of Proposition 3.2 (i), it remains to show that $\sup_{n \in \mathbb{N}} \mathbb{E}_{\mathbb{P}}[\|G_2(h(\lambda_n), \lambda_n, X_{n+1})\|^2] < +\infty$. Since for $\lambda \in H$, $|\varphi_T(\lambda)| < +\infty$, $\|h(\lambda)\| < +\infty$, and $\|\nabla_{\lambda} \varphi(\lambda)\| < +\infty$, we get

$$\begin{aligned} &\sup_{n \in \mathbb{N}} \mathbb{E}_{\mathbb{P}} [\|G_2(h(\lambda_n), \lambda_n, X_{n+1})\|^2] \\ &\leq \sup_{\lambda \in H} \mathbb{E}_{\mathbb{P}} \left[e^{-2\langle \lambda, X_T \rangle + 2\varphi_T(\lambda)} (F(X) - \langle h(\lambda), X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle)^4 \|\nabla_{\lambda} \varphi(\lambda) - X_T\|^2 \right] < +\infty, \end{aligned}$$

where the last finiteness follows again from $\lambda \in H \subseteq \Lambda_4$. $\square$

Proof of Theorem 3.1. By Proposition 3.2 (i), for each $\lambda \in \Lambda_4$, the sequence $\{\theta_n\}_{n \in \mathbb{N}}$ in (3.6) converges $\mathbb{P}$ -a.s. to $h(\lambda)$. Moreover, by Proposition 3.2 (ii), the sequence $\{\lambda_n\}_{n \in \mathbb{N}}$ in (3.7) converges $\mathbb{P}$ -a.s. to $\lambda^{\dagger}$. The claim then holds by Theorem 1.1 [6] with the condition (3.5). $\square$

Proof of Theorem 4.3. Following Theorem 1-2 of Arouna [3], we will show that the expectation $\mathbb{E}_{\mathbb{P}}[Y^2(h(\lambda), \lambda, X)]$ is continuous at $\lambda^\dagger$, and that $\sup_{n \in \mathbb{N}} \mathbb{E}_{\mathbb{P}}[Y^2(\theta_{n-1}, \lambda_{n-1}, X_n)] < +\infty$. To this end, note first that for $\theta \in \mathbb{R}^d$ and for $\lambda \in \Lambda_4$,

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[Y^2(\theta, \lambda, X)] &= \mathbb{E}_{\mathbb{Q}_\lambda} \left[e^{-2\langle \lambda, X_T \rangle + 2\varphi_T(\lambda)} (F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle)^2 \right] \\ &= e^{\varphi_T(\lambda) + \varphi_T(-\lambda)} \mathbb{E}_{\mathbb{P}} \left[e^{-\langle \lambda, X_T \rangle + \varphi_T(\lambda)} (F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle)^2 \right] \\ &= e^{\varphi_T(\lambda) + \varphi_T(-\lambda)} (V_3(\theta, \lambda) + C^2). \end{aligned}$$

In view of (A.1), we get for $\lambda \in \Lambda_4$,

$$\begin{aligned} \nabla_\lambda (\mathbb{E}_{\mathbb{P}}[Y^2(h(\lambda), \lambda, X)])|_{\lambda=\lambda^\dagger} &= e^{\varphi_T(\lambda^\dagger) + \varphi_T(-\lambda^\dagger)} \\ &\times \left([\nabla_\lambda \varphi_T(\lambda) + \nabla_\lambda \varphi_T(-\lambda)]|_{\lambda=\lambda^\dagger} \left(V_3(h(\lambda^\dagger), \lambda^\dagger) + C^2 \right) + \mathbb{E}_{\mathbb{P}}[G_1(h(\lambda^\dagger), \lambda^\dagger, X)] \right), \end{aligned}$$

which is sufficient for the continuity at $\lambda = \lambda^\dagger$. For the finiteness of the supremum, observe that

$$\sup_{n \in \mathbb{N}} (\varphi_T(\lambda_n) + \varphi_T(-\lambda_n)) \leq \sup_{\lambda \in \Lambda_4} (\varphi_T(\lambda) + \varphi_T(-\lambda)) < +\infty.$$

Due to the definition of the sequence $\{H_n\}_{n \in \mathbb{N}}$ in the iteration (3.4) and $\lim_{n \uparrow +\infty} \sigma(n) < +\infty$, $\mathbb{P}$ -a.s. shown in Theorem 4.3, it is clear that $\sup_{n \in \mathbb{N}} V_3(\theta_n, \lambda_n) < +\infty$. $\square$

Proof of Proposition 4.4. Following Corollary 1 [3], it is sufficient to show that the expectation $\mathbb{E}_{\mathbb{P}}[Y^4(h(\lambda), \lambda, X)]$ is continuous at $\lambda^\dagger$, and that $\sup_{n \in \mathbb{N}} \mathbb{E}_{\mathbb{P}}[Y^4(\theta_{n-1}, \lambda_{n-1}, X_n)] < +\infty$. The proof is similar to that of Theorem 4.3. First observe that

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[Y^4(\theta, \lambda, X)] &= \mathbb{E}_{\mathbb{Q}_\lambda} \left[e^{-4\langle \lambda, X_T \rangle + 4\varphi_T(\lambda)} (F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle)^4 \right] \\ &= e^{4(\varphi_T(\lambda) + \varphi_T(-\lambda))} \mathbb{E}_{\mathbb{P}} \left[e^{-3\langle \lambda, X_T \rangle} (F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle)^4 \right] < +\infty, \end{aligned}$$

where the finiteness holds true since $\lambda \in \Lambda_5$. Moreover, the continuity of $h(\lambda)$ in λ on Λ_5 implies that of the expectation $\mathbb{E}_{\mathbb{P}}[Y^4(h(\lambda), \lambda, X)]$ at $\lambda = \lambda^\dagger$. As in the proof of Theorem 4.3, the finiteness of $\sup_{n \in \mathbb{N}} \mathbb{E}_{\mathbb{P}}[Y^4(\theta_{n-1}, \lambda_{n-1}, X_n)]$ follows from the definition of the sequence $\{H_n\}_{n \in \mathbb{N}}$ in the iteration (3.3)-(3.4) and $\lim_{n \uparrow +\infty} \sigma(n) < +\infty$, $\mathbb{P}$ -a.s. $\square$

B Equilibrated importance sampling and quasi-static control variates

In this section, we briefly explain, yet with full theoretical support, why it is doubtful that the opposite scheme “*equilibrated importance sampling and quasi-static control variates*” works. The corresponding two-time-scale stochastic approximation algorithm is as follows:

$$\lambda_{n+1} = \Pi_{H^\lambda} [\lambda_n - \varepsilon_n G_1(\theta_n, \lambda_n, X_{n+1})], \quad (\text{B.1})$$

$$\theta_{n+1} = \Pi_{H^\theta} [\theta_n - \delta_n G_2(\theta_n, \lambda_n, X_{n+1})], \quad (\text{B.2})$$

with

$$\lim_{n \uparrow +\infty} \frac{\delta_n}{\varepsilon_n} = 0, \quad (\text{B.3})$$

that is, the iteration for λ moves much faster than the one for θ . Here, H^λ is a compact subset of Λ_4 with $\{0\} \in H^\lambda$, while H^θ is a compact subset of $\Theta := \{\theta \in \mathbb{R}^d : [\nabla_\lambda V_3(\theta, \lambda)]|_{\lambda=g(\theta)} = 0\}$. Moreover, define a function $V_5 : \Theta \mapsto (0, +\infty)$ by $V_5(\theta) := V_3(\theta, g(\theta))$.

Proposition B.1. (i) Fix $\theta \in \mathbb{R}^d$. Then, the point $g(\theta) := \operatorname{argmin}_{\lambda \in H^\lambda} V_3(\theta, \lambda)$ is unique, and the sequence $\{\lambda_n\}_{n \in \mathbb{N}}$ defined in the iteration

$$\lambda_{n+1} = \Pi_{H^\lambda} [\lambda_n - \varepsilon_n G_1(\theta, \lambda_n, X_{n+1})]$$

converges $\mathbb{P}$ -a.s. to $g(\theta)$.

(ii) Assume that the set Θ is not empty and that the matrix $\nabla_\theta [g(\theta)']$ is well defined. Then, the point $\theta^\dagger := \operatorname{argmin}_{\theta \in H^\theta} V_5(\theta)$ is unique, and the sequence $\{\theta_n\}_{n \in \mathbb{N}}$ defined in the iteration

$$\theta_{n+1} = \Pi_{H^\theta} [\theta_n - \delta_n G_2(\theta_n, g(\theta_n), X_{n+1})]$$

converges $\mathbb{P}$ -a.s. to $\theta^\dagger$.

(iii) Under the same assumptions imposed in (ii), the sequence $\{(\theta_n, \lambda_n)\}_{n \in \mathbb{N}}$ defined in the iteration (B.1)-(B.2) converges $\mathbb{P}$ -a.s. to $(\theta^\dagger, g(\theta^\dagger))$.

Proof. (i) This part can be proved in a similar manner to the corresponding result in [16].

(ii) This part can be proved in a similar manner to Proposition 3.2. For each $\theta \in \Theta$, the envelope theorem yields

$$\nabla_\theta V_5(\theta) = [\nabla_\theta V_3(\theta, \lambda)]|_{\lambda=g(\theta)}, \quad (\text{B.4})$$

and moreover $\operatorname{Hess}_\theta [V_5(\theta)] = [\operatorname{Hess}_\theta V_3(\theta, \lambda)]|_{\lambda=g(\theta)}$, whose positive-definiteness implies the first assertion. As in the proof of Proposition 3.2 (ii), the second assertion follows from

$$\begin{aligned} & \sup_{n \in \mathbb{N}} \mathbb{E}_{\mathbb{P}} [\|G_2(\theta_n, g(\theta_n), X_{n+1})\|^2] \\ & \leq \sup_{\theta \in H^\theta} \mathbb{E}_{\mathbb{P}} \left[e^{-2\langle g(\theta), X_T \rangle + 2\varphi_T(g(\theta))} (F(X) - \langle \theta, X_T - \mathbb{E}_{\mathbb{P}}[X_T] \rangle)^2 \|X_T - \mathbb{E}_{\mathbb{P}}[X_T]\|^2 \right], \end{aligned}$$

which is finite since $g(\theta) \in \Lambda_4$.

(iii) The assertion holds by Theorem 1.1 [6] with the above results (i) and (ii), and with the condition (B.3). $\square$

Let us below discuss various reasons why the above results are unlikely to be useful in practice.

1. The function $g(\theta)$ does not have closed form solution, unlike $h(\lambda)$ as given in (3.2). It is thus not clear how to check whether the matrix $\nabla_\theta [g(\theta)']$ is well defined. Then, we do not know if the equation (B.4) holds true.

2. For the same reason as discussed in the item 1, we do not know if the strict convexity of V_5 remains true on $\mathbb{R}^d \setminus \Theta$. Hence, we set the iteration (B.2) to be constrained for safety reason. (Recall that the sequence $\{\theta_n\}_{n \in \mathbb{N}}$ is originally not required to be bounded.)
3. It is not clear when $[\nabla_\lambda V_3(\theta, \lambda)]|_{\lambda=g(\theta)} = 0$ holds true, that is, we do not know what the set Θ is like. We even do not know whether or not the origin is contained in Θ .
4. In relation to the item 3, the optimal variance might actually end up worse, that is, $V_5(\theta^\dagger) > V_3(\theta^*, g(\theta^*))$.
5. Suppose $\theta_n \notin \Theta$. Then, the $(n+1)$ -st iteration of the algorithm (3.4) may be meaningless in the sense that the sequence might have missed its step out of the gradient-based domain. Worse, it is even not clear how to check whether $\theta_n \notin \Theta$.

Remark B.2. On contrary, if $\Lambda_4 = \mathbb{R}^d$, then under a mild condition on ν , we may prove the almost sure convergence to $\lambda^\dagger$ and $\theta^\dagger$ satisfying $[\nabla_\lambda V_4(\lambda)]|_{\lambda=\lambda^\dagger} = [\nabla_\theta V_5(\theta)]|_{\theta=\theta^\dagger} = 0$, but we do not go into that. The primal reason is that for $\Lambda_4 = \mathbb{R}^d$, the Lévy measure is required to decay extremely fast at infinity, for example, $\nu(dz) = (2\pi)^{-1/2} e^{-z^2/2} dz$, $z \in \mathbb{R}_0$, as considered in [16] for numerical illustrations, while in view of Theorem 2.1, the resulting distributions have also extremely light tails, which is not very appealing in practice. $\square$

References

- [1] Applebaum, D. (2004) *Lévy Processes and Stochastic Calculus*, Cambridge University Press, Cambridge.
- [2] Arouna, B. (2004) Robbins-Monro algorithms and variance reduction in finance, *Journal of Computational Finance*, **7**, 35-61.
- [3] Arouna, B (2004) Adaptive Monte Carlo method, a variance reduction technique, *Monte Carlo Methods and Applications*, **10**, 1-24.
- [4] Bardi, M., Capuzzo-Dolcetta, I. (1997) *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*, Birkhauser, Boston.
- [5] Baxter, M. (2006) Lévy simple structural models, available at <http://www.nomura.com/europe/services/derivatives/credit.shtml>
- [6] Borkar, V.S. (1997) Stochastic approximation with two time scales, *Systems & Control Letters*, **29**, 291-294.
- [7] Chen, K.-F., Zhu, Y. (1986) *Stochastic approximation procedure with randomly varying truncations*, Scientia Sinica Series.

- [8] Delyon, B. (1996) General results on the convergence of stochastic algorithms, *IEEE Transactions on Automatic Control*, **41**, 1245-1255.
- [9] Glasserman, P., Heidelberger, P., Shahabuddin, P. (1999) Asymptotic optimal importance sampling and stratification for pricing path-dependent options, *Mathematical Finance*, **9**, 117-152.
- [10] Gregory, J., Laurent, A. (2003) I will survive, *Risk*, **16**(6), 103-107.
- [11] Hull, J., White, A. (2004) Valuation of a CDO and an n -th to Default CDS Without Monte Carlo Simulation, *Journal of Derivatives*, **12**(2), 8-23.
- [12] Joshi, M., Stacey, A. (2006) Intensity gamma, *Risk*, **19**(7), 78-83.
- [13] Kalemánova, A., Schmid, B., Werner, R. (2007) The normal inverse Gaussian distribution for synthetic CDO pricing, *Journal of Derivatives*, **14**(3), 80-93.
- [14] Kallenberg, O. (2002) *Foundations of Modern Probability* (2nd edn). New York: Springer-Verlag.
- [15] Kallsen, J., Tankov, P. (2006) Characterization of dependence of multidimensional Lévy processes using Lévy copulas, *Journal of Multivariate Analysis*, **97**, 1551-1572.
- [16] Kawai, R. (2007) Optimal importance sampling via stochastic approximation for Lévy processes with exponential tilting, *submitted*.
- [17] Kawai, R. (2007) Adaptive Monte Carlo variance reduction with two-time-scale stochastic approximation, *Monte Carlo Methods and Applications*, **13**(3), to appear.
- [18] Li, D.X. (2000) On default correlation: a copula function approach, *Journal of Fixed Income*, **9**, 43-54.
- [19] Madan, D., Carr, P., Change, E. (1998) The variance gamma process and option pricing, *European Finance Review*, **2**, 79-105.
- [20] O’Kane, D., Schloegl, L. (2003) An analytical portfolio credit model with tail dependence, Quantitative Credit Research, Lehman Brothers.
- [21] Randhawa, R.S., Juneja, S. (2004) Combining importance sampling and temporal difference control variates to simulate Markov Chains, *ACM Transactions on Modeling and Computer Simulation*, **14**, 1-30.
- [22] Sato, K. (1999) *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, Cambridge.

- [23] Schönbucher, P. (2003) *Credit Derivatives Pricing Models*, Wiley, Chichester.
- [24] Su, Y., Fu, M.-C. (1998) Optimal importance sampling in securities pricing, *Journal of Computational Finance*, **5**, 27-50.

$\rho = 0.10$								
Tranche		price	variance	nstdev	α_δ	α_ε	vratio1	vratio2
6-9%	C_1	0.28	10.03	11.19	3e-1	3e-1	13.68	9.59
	C_2	0.25	8.53	11.60			14.31	14.08
9-12%	C_1	0.16	6.05	15.17	5e-1	2e-1	21.50	20.75
	C_2	0.14	5.16	15.70			22.62	21.92
12-22%	C_1	0.27	33.22	20.97	5e-1	2e-2	41.78	29.52
	C_2	0.25	28.37	21.59			44.07	30.08
$\rho = 0.30$								
Tranche		price	variance	nstdev	α_δ	α_ε	vratio1	vratio2
6-9%	C_1	0.73	27.58	7.17	1e-1	1e-1	13.02	12.65
	C_2	0.66	23.66	7.39			13.55	13.21
9-12%	C_1	0.47	18.19	9.06	2e-1	2e-2	18.15	17.96
	C_2	0.42	15.61	9.34			18.98	18.57
12-22%	C_1	0.88	112.54	12.00	2e-1	2e-2	30.31	28.99
	C_2	0.80	96.69	12.36			31.79	29.64
$\rho = 0.50$								
Tranche		price	variance	nstdev	α_δ	α_ε	vratio1	vratio2
6-9%	C_1	1.04	40.31	6.08	3e-2	2e-2	13.50	11.69
	C_2	0.94	34.71	6.25			13.99	12.07
9-12%	C_1	0.73	28.76	7.35	1e-1	2e-2	17.50	15.05
	C_2	0.66	24.77	7.56			18.19	15.57
12-22%	C_1	1.53	196.64	9.19	1e-1	2e-2	25.73	24.51
	C_2	1.38	169.13	9.44			26.80	25.28

Table 1: Numerical results for the Gamma copula model

$\{\theta_n\}_{n \in \mathbb{N}}$ and $\{\lambda_n\}_{n \in \mathbb{N}}$
 (---, -◇-) for C_1

(-◇-, -+-) are $\{\|\nabla_{\lambda} V_3(\theta_n, \lambda_n)\|\}_{n \in \mathbb{N}}$
 and $\{\|\nabla_{\theta} V_3(\theta_n, \lambda_n)\|\}_{n \in \mathbb{N}}$ for C_1

Convergence of plain Monte Carlo and adaptive Monte Carlo

Figure 1: Illustrations for the Gamma copula model for 12-22% tranche with $\rho = 0.30$. Resulting (vratio2) are 28.99 and 29.64.

$\rho = 0.10$								
Tranche		price	variance	nstdev	α_δ	α_ε	vratio1	vratio2
6-9%	C_1	0.19	6.53	13.74	2e-1	5e-2	4.55	4.44
	C_2	0.17	5.62	14.16			4.63	4.47
9-12%	C_1	0.083	3.24	21.68	2e0	3e-1	8.36	7.51
	C_2	0.075	2.80	22.21			8.51	7.74
12-22%	C_1	0.070	6.11	35.26	2e0	3e-1	19.99	17.81
	C_2	0.064	5.31	35.92			20.32	18.22

$\rho = 0.30$								
Tranche		price	variance	nstdev	α_δ	α_ε	vratio1	vratio2
6-9%	C_1	0.53	18.19	8.00	1e-1	3e-2	4.05	3.92
	C_2	0.48	15.58	8.26			4.08	3.96
9-12%	C_1	0.23	8.01	12.41	2e-0	2e-1	6.18	5.81
	C_2	0.20	6.87	12.81			6.19	5.99
12-22%	C_1	0.17	13.90	21.58	1e0	3e-1	13.51	13.09
	C_2	0.15	11.91	22.28			13.55	12.81

$\rho = 0.50$								
Tranche		price	variance	nstdev	α_δ	α_ε	vratio1	vratio2
6-9%	C_1	0.87	30.24	6.35	5e-2	3e-2	4.02	3.91
	C_2	0.78	26.01	6.54			4.05	3.93
9-12%	C_1	0.39	14.02	9.50	5e-1	2e-1	6.29	6.15
	C_2	0.36	12.05	9.78			6.28	6.09
12-22%	C_1	0.31	26.94	16.56	3e-1	3e-1	14.17	13.30
	C_2	0.28	23.15	17.08			14.14	13.73

Table 2: Numerical results for the intensity Gamma model

$\{\theta_n\}_{n \in \mathbb{N}}$ and $\{\lambda_n\}_{n \in \mathbb{N}}$
 $(-+-, -\diamond-)$ for C_1

$(-\diamond-, -+-)$ are $\{\|\nabla_{\lambda} V_3(\theta_n, \lambda_n)\|\}_{n \in \mathbb{N}}$
 and $\{\|\nabla_{\theta} V_3(\theta_n, \lambda_n)\|\}_{n \in \mathbb{N}}$ for C_1

Convergences of plain Monte Carlo and adaptive Monte Carlo

Figure 2: Illustrations for the intensity Gamma model for 12-22% tranche with $\rho = 0.50$. Resulting (vratio2) are 13.30 and 13.73.