Importance sampling for jump processes and applications to finance

Laetitia Badouraly Kassim* Jérôme Lelong* Imane Loumrhari*

March 3, 2020

Abstract

Adaptive importance sampling techniques are widely known for the Gaussian setting of Brownian driven diffusions. In this work, we want to extend them to jump processes. Our approach relies on a change of the jump intensity combined with the standard exponential tilting for the Brownian motion. The free parameters of our framework are optimized using sample average approximation techniques. We illustrate the efficiency of our method on the valuation of financial derivatives in several jump models.

Keywords: Importance sampling; sample average approximation; adaptive Monte Carlo methods.

Premia 22

1 Introduction

Lévy models have become quite popular in finance over the last decade. Vanilla options are easily and efficiently priced using the Fast Fourier Transform approach developed by Carr et al. (1999) but things become far more delicate for exotic options, for which Monte Carlo often reveals as the only possible approach from a numerical point of view. This becomes even more true when dealing with high dimensional products. In this work, we want to propose an adaptive Monte Carlo method based on importance sampling for computing the expectation of a function of a Lévy process. As explained by Kiessling and Tempone (2011), when resorting to Monte Carlo approaches, infinite activity Lévy processes are often approximated by finite activity processes, which can always be represented as the sum of a continuous diffusion (ie. driven by a Brownian motion) and a compound Poisson process. In this work, we will concentrate on such jump diffusions with a Brownian driven part and a jump part written as a compound Poisson process or possibly the sum of independent compound Poisson processes in the multidimensional case.

We consider a mixed Gaussian and Poisson framework in which we would like to settle an adaptive Monte Carlo method based on some importance sampling approach. Let $G = (G_1, \dots, G_d)$ be a standard normal random vector in $\mathbb{R}^d$ and $N^\mu = (N_1^{\mu_1}, \dots, N_p^{\mu_p})$ a vector of p independent Poisson

*Univ. Grenoble Alpes, Laboratoire Jean Kuntzmann, BP 53, 38041 Grenoble Cédex 9, FRANCE, e-mail: Laetitia.Badouraly-Kassim@ensimag.imag.fr, jerome.lelong@imag.fr, imane.loumrhari@ensimag.imag.fr.
This project was supported by the Finance for Energy Market Research Centre, www.fime-lab.org.

random variables with parameters $\mu = (\mu_1, \ldots, \mu_p)$. We assume that G and N^μ are independent. We focus on the computation of

$$\mathcal{E} = \mathbb{E}[f(G, N^\mu)] \tag{1.1}$$

where $f : \mathbb{R}^d \times \mathbb{N}^p \longrightarrow \mathbb{R}$ satisfies $\mathbb{E}[|f(G, N^\mu)|] < \infty$.

Lemma 1.1. *For any measurable function $h : \mathbb{R}^d \times \mathbb{N}^p \longrightarrow \mathbb{R}$ either nonnegative or such that $\mathbb{E}[|h(G, N^\mu)|] < \infty$, one has*

$$\forall\, \theta \in \mathbb{R}^d, \lambda \in (]0, +\infty[)^p, \qquad \mathbb{E}[h(G, N^\mu)] = \mathbb{E}\left[h(G + \theta, N^\lambda)\, \mathrm{e}^{-\theta \cdot G - \frac{|\theta|^2}{2}} \prod_{i=1}^{p} \mathrm{e}^{\lambda_i - \mu_i} \left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\lambda_i}} \right] \tag{1.2}$$

where N^λ is a vector of p independent Poisson random variables with parameters $\lambda = (\lambda_1, \ldots, \lambda_p)$.

The proof of this lemma relies on elementary variable changes. Lemma 1.1 enables us to introduce some extra degrees of freedom in the computation of $\mathcal{E}$. When the expectation $\mathcal{E}$ is computed using a Monte Carlo method, the Central Limit Theorem advises to use the representation of $f(G, N^\mu)$ with the smallest possible variable which is achieved by choosing the parameters (θ, λ) which minimize the variance of of $f(G + \theta, N^\lambda)\, \mathrm{e}^{-\theta \cdot G - \frac{|\theta|^2}{2}} \prod_{i=1}^{p} \mathrm{e}^{\lambda_i - \mu_i} \left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\lambda_i}}$. This raises several questions which are investigated in the paper. Does the variance of $f(G + \theta, N^\lambda)\, \mathrm{e}^{-\theta \cdot G - \frac{|\theta|^2}{2}} \prod_{i=1}^{p} \mathrm{e}^{\lambda_i - \mu_i} \left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\lambda_i}}$ admits a unique minimizer? If so, how can it be computed numerically and how to make the most of it in view of a further Monte Carlo computation?

These questions are quite natural in the context of Monte Carlo computations and have already been widely discussed in the pure Gaussian framework. The first applications to option pricing of some adaptive Monte Carlo methods based on importance sampling goes back to the papers of Arouna (Winter 2003/04, 2004). These papers were based on a change of mean for the Gaussian random normal vectors and the optimal parameter was searched for by using some stochastic approximation algorithm with random truncations. This approach was later further investigated by Lapeyre and Lelong (2011) who proposed a more general framework for settling adaptive Monte Carlo methods using stochastic approximation, which is know to be a little tricky to fine tune in practical applications. To circumvent the delicate behaviour of stochastic approximation, Jourdain and Lelong (2009) proposed to resort to sample average approximation instead, which basically relies on deterministic optimization techniques. An alternative to random truncations was studied by Lemaire and Pagès (2010) who managed to modify the initial problem in order ta apply the more standard Robbins Monro algorithm. Not only have they applied this to the Gaussian framework but they have considered a few examples of Levy processes relying on the Esscher transform to introduce a free parameter. The idea of using the Esscher transform was also extensively investigated by Kawai (2007, 2008a,b).

In this work, we want to understand how the jump intensity of a Lévy process can be modified to reduce the variance. First, we explain the parametric importance sampling transformation we use for the Gaussian and Poisson parts. Then, in Section 2, we prove that this transformation leads to a convex optimization problem and we study the properties of the optimal parameter estimator. Then, in Section 3, we explain how to use this estimator in a Monte Carlo method. We prove that this approach satisfies an adaptive strong law of large numbers and a central limit theorem with optimal limiting variance. Finally, in Section 4, we apply our methodology to option pricing with jump processes.

Notations.

- We define the space $\mathbb{R}_+^{*\,m} = \{x \in \mathbb{R}^m \ : \ \forall\, 1 \leq i \leq m,\ x_i > 0\}$.
- We define the space $\mathbb{N}^{*m} = \{x \in \mathbb{N}^m \ : \ \forall\, 1 \leq i \leq m,\ x_i > 0\}$.
- We encode any elements of $\mathbb{R}^m$ or $\mathbb{R}_+^{*\,m}$ as column vectors.
- If $x \in \mathbb{R}^m$, x^* is a row vector. We use the "*" notation to denote the transpose operator for vectors and matrices. Beware not to mix this notation with the subscript $_\star$, which is used to denote optimal solutions of optimization problems.
- If $x, y \in \mathbb{R}^m$, $x \cdot y$ denotes the scalar product of x and y and the associated norm is denoted by $|\cdot|$.
- If $x \in \mathbb{R}^m$, $\mathrm{diag}_m(x)$ is the matrix with diagonal elements given by the vector x and all extra diagonal elements equal to zero.
- The matrix I_m denotes the identity matrix in dimension m.
- If $x \in \mathbb{R}^m$, we defined $d_0(x) = \min_{1\leq i\leq m} |x_i|$ which is the distance between x and the set $\{y \in \mathbb{R}^m \ : \ \prod_{i=1}^m y_i = 0\}$.
- We say that a random vector X with values in $\mathbb{R}^m$ has Poisson distribution with parameter $\mu \in \mathbb{R}^m$ if the X_i are independent and have Poisson distribution with parameter μ_i.

2 Computing the optimal importance sampling parameters

2.1 Properties of the variance

Thanks to Lemma 1.1, the expectation $\mathcal{E}$ can be written

$$\mathcal{E} = \mathbb{E}\left[f(G+\theta, N^\lambda)\, \mathrm{e}^{-\theta\cdot G - \frac{|\theta|^2}{2}} \prod_{i=1}^p \mathrm{e}^{\lambda_i-\mu_i} \left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\lambda_i}} \right], \quad \forall\, \theta \in \mathbb{R}^d, \lambda \in \mathbb{R}_+^{*\,p}.$$

Note that for the particular choice of $\theta = 0$ and $\lambda = \mu$, we recover Equation (1.1).
The convergence rate of a Monte Carlo estimator of $\mathcal{E}$ based on this new representation is governed by the variance of $f(G+\theta, N^\lambda)\, \mathrm{e}^{-\theta\cdot G - \frac{|\theta|^2}{2}} \prod_{i=1}^p \mathrm{e}^{\lambda_i-\mu_i} \left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\lambda_i}}$ which can be written in the form $v(\theta,\lambda) - \mathcal{E}^2$ where

$$v(\theta,\lambda) = \mathbb{E}\left[f(G, N^\mu)^2\, \mathrm{e}^{-\theta\cdot G + \frac{|\theta|^2}{2}} \prod_{i=1}^p \mathrm{e}^{\lambda_i-\mu_i} \left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}} \right]. \tag{2.1}$$

This expression of v is easily obtained by applying Lemma 1.1 to the function $h(g,n) = f(g+\theta, n)^2\, \mathrm{e}^{-2\theta\cdot g - |\theta|^2} \prod_{i=1}^p \mathrm{e}^{2(\lambda_i-\mu_i)} \left(\frac{\mu_i}{\lambda_i}\right)^{2n_i}$. Applying the change of measure backward after computing the variance enables us to write the variance in a form which does not involve the parameters θ and λ in the arguments of the function f. This remark is of prime importance as it is the basement of the following key result stating the strong convexity of v.

Proposition 2.1. *Assume that*

(A1) *i.* $\exists (n_1, \dots, n_p) \in \mathbb{N}^{*p}$, *s.t.* $\mathbb{P}(|f(G, (n_1, \dots, n_p))| > 0) > 0$

ii. $\exists \gamma > 0, \quad \mathbb{E}\left[|f(G, N^\mu)|^{2+\gamma}\right] < \infty.$

Then, the function v is infinitely continuously differentiable, strongly convex and moreover the gradient vectors are given by

$$\nabla_\theta v(\theta, \lambda) = \mathbb{E}\left[(\theta - G)\, f(G, N^\mu)^2\, \mathrm{e}^{-\theta \cdot G + \frac{|\theta|^2}{2}} \prod_{i=1}^{p} \mathrm{e}^{\lambda_i - \mu_i} \left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right] \tag{2.2}$$

$$\nabla_\lambda v(\theta, \lambda) = \mathbb{E}\left[a(N^\mu, \lambda) f(G, N^\mu)^2\, \mathrm{e}^{-\theta \cdot G + \frac{|\theta|^2}{2}} \prod_{i=1}^{p} \mathrm{e}^{\lambda_i - \mu_i} \left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right] \tag{2.3}$$

where the vector $a(N^\mu, \lambda) = \left(1 - \frac{N_1^{\mu_1}}{\lambda_1}, \dots, 1 - \frac{N_p^{\mu_p}}{\lambda_p}\right)^*$. *The second derivatives are defined by*

$$\nabla^2_{\theta,\theta} v(\theta, \lambda) = \mathbb{E}\left[(I_d + (\theta - G)(\theta - G)^*)\, f(G, N^\mu)^2\, \mathrm{e}^{-\theta \cdot G + \frac{|\theta|^2}{2}} \prod_{i=1}^{p} \mathrm{e}^{\lambda_i - \mu_i} \left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right] \tag{2.4}$$

$$\nabla^2_{\theta,\lambda} v(\theta, \lambda) = \mathbb{E}\left[(\theta - G) a(N^\mu, \lambda)^* f(G, N^\mu)^2\, \mathrm{e}^{-\theta \cdot G + \frac{|\theta|^2}{2}} \prod_{i=1}^{p} \mathrm{e}^{\lambda_i - \mu_i} \left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right] \tag{2.5}$$

$$\nabla^2_{\lambda,\lambda} v(\theta, \lambda) = \mathbb{E}\left[(D + a(N^\mu, \lambda) a(N^\mu, \lambda)^*)\, f(G, N^\mu)^2\, \mathrm{e}^{-\theta \cdot G + \frac{|\theta|^2}{2}} \prod_{i=1}^{p} \mathrm{e}^{\lambda_i - \mu_i} \left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right] \tag{2.6}$$

where the diagonal matrix D is defined by $D = \mathrm{diag}_p\left(\frac{N_1^{\mu_1}}{\lambda_1^2}, \dots, \frac{N_p^{\mu_p}}{\lambda_p^2}\right)$.

Proof. Let us define the function $F : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}_+^{*p} \longrightarrow \mathbb{R}$ by

$$F(g, \theta, n, \lambda) = f(g, n)^2\, \mathrm{e}^{-\theta \cdot g + \frac{|\theta|^2}{2}} \prod_{i=1}^{p} \mathrm{e}^{\lambda_i - \mu_i} \left(\frac{\mu_i}{\lambda_i}\right)^{n_i}. \tag{2.7}$$

For any values of (g, n), the function $(\theta, \lambda) \longmapsto F(g, \theta, n, \lambda)$ is infinitely continuously differentiable. Since for all $0 < \underline{m} < M$,

$$\sup_{|(\theta,\lambda)| \le M, \underline{m} < d_0(\lambda)} |\partial_{\theta_j} F(G, \theta, N^\mu, \lambda)| \le \left(M + \mathrm{e}^{G_j} + \mathrm{e}^{-G_j}\right) f(G, N^\mu)^2\, \mathrm{e}^{M^2/2 + pM} \prod_{k=1}^{d} (\mathrm{e}^{MG_k} + \mathrm{e}^{-MG_k}) \prod_{i=1}^{p} \mathrm{e}^{-\mu_i} \left(\frac{\mu_i}{\underline{m}}\right)^{N_i^{\mu_i}} \tag{2.8}$$

where the right hand side is integrable because by Hölder's inequality and Assumption tegrable](*A*1-*ii)*, we have that for all $(\theta, \lambda) \in \mathbb{R}^d \times \mathbb{R}^p$, $\mathbb{E}(f(G, N^\mu)^2\, \mathrm{e}^{\theta \cdot G + \lambda \cdot N^\mu}) < \infty$. Hence, Lebesgue's theorem ensures that v is continuously differentiable w.r.t. θ and $\nabla_\theta v$ is given by Equation (2.2).

We proceed similarly for the derivative w.r.t. λ by using the following upper bound

$$\sup_{|(\theta,\lambda)| \le M, \underline{m} < d_0(\lambda)} |\partial_{\lambda_j} F(G, \theta, N^\mu, \lambda)| \le \left(1 + \mathrm{e}^{N_j^{\mu_j}/\underline{m}} + \mathrm{e}^{-N_j^{\mu_j}/m}\right) f(G, N^\mu)^2\, \mathrm{e}^{M^2/2 + pM} \prod_{k=1}^{d} (\mathrm{e}^{MG_k} + \mathrm{e}^{-MG_k}) \prod_{i=1}^{p} \mathrm{e}^{-\mu_i} \left(\frac{\mu_i}{\underline{m}}\right)^{N_i^{\mu_i}}. \tag{2.9}$$

High order differentiability properties are obtained by similar arguments and in particular the Hessian matrix writes with the help of the function F

$$\nabla^2 v(\theta,\lambda) = E\left[F(G,\theta,N^\mu,\lambda)\begin{pmatrix}(\theta-G)(\theta-G)^* & (\theta-G)a(N^\mu,\lambda)^*\\ a(N^\mu,\lambda)(\theta-G)^* & a(N^\mu,\lambda)a(N^\mu,\lambda)^*\end{pmatrix} + F(G,\theta,N^\mu,\lambda)\begin{pmatrix}I_d & 0\\ 0 & D\end{pmatrix}\right].$$

Note that

$$\begin{pmatrix}(\theta-G)(\theta-G)^* & (\theta-G)a(N^\mu,\lambda)^*\\ a(N^\mu,\lambda)(\theta-G)^* & a(N^\mu,\lambda)a(N^\mu,\lambda)^*\end{pmatrix} = \begin{pmatrix}\theta-G\\ a(N^\mu,\lambda)\end{pmatrix}\begin{pmatrix}\theta-G\\ a(N^\mu,\lambda)\end{pmatrix}^*.$$

Hence, the first part of the Hessian is a positive semi definite rank one matrix.

$$\mathbb{E}\left[F(G,\theta,N^\mu,\lambda)\begin{pmatrix}I_d & 0\\ 0 & D\end{pmatrix}\right] \geq \mathbb{E}[F(G,\theta,N^\mu,\lambda)\mathbf{1}_{\{N^\mu=(n_1,\dots,n_p)\}}]\,\mathrm{diag}\left(I_d,\frac{n_1}{\lambda_1^2},\dots,\frac{n_p}{\lambda_p^2}\right).$$

Moreover,

$$\begin{aligned}\mathbb{E}[F(G,\theta,N^\mu,\lambda)\mathbf{1}_{\{N^\mu=(n_1,\dots,n_p)\}}] &\geq \mathbb{E}\left[f(G,(n_1,\dots,n_p))^2\,\mathrm{e}^{-\theta\cdot G+\frac{|\theta|^2}{2}}\right]\prod_{i=1}^p \mathrm{e}^{n_i-2\mu_i}\left(\frac{\mu_i^2}{n_i}\right)^{n_i}\frac{1}{n_i!}\\ &\geq \mathbb{E}\left[f(G,(n_1,\dots,n_p))^2\,\mathrm{e}^{-\theta\cdot G}\right]\mathbb{E}\left[\mathrm{e}^{\theta\cdot G}\right]\prod_{i=1}^p \mathrm{e}^{n_i-2\mu_i}\left(\frac{\mu_i^2}{n_i}\right)^{n_i}\frac{1}{n_i!}\\ &\geq \left(\mathbb{E}\left[|f(G,(n_1,\dots,n_p))|\right]\right)^2\prod_{i=1}^p \mathrm{e}^{n_i-2\mu_i}\left(\frac{\mu_i^2}{n_i}\right)^{n_i}\frac{1}{n_i!}\end{aligned}$$

where the last line comes from the Cauchy Schwartz inequality. Thanks to Condition sitive](*A*1-*i)*, this lower bound is strictly positive. Hence, the Hessian matrix is uniformly bounded from below, which yields the strong convexity of v. ■

As a consequence, the function v admits a unique minimizer $(\theta_\star,\lambda_\star)$ defined by $\nabla_\theta v(\theta_\star,\lambda_\star) = \nabla_\lambda v(\theta_\star,\lambda_\star) = 0$. The characterization of $(\theta_\star,\lambda_\star)$ as the unique minimizer of a strongly convex function is very appealing but there is no hope to compute the gradient of v in a closed form, so we will need to resort to some kind of approximations before running the optimization step. Before studying the possible ways of approximating the optimal parameter, let us note that it is of dimension $d+p$ which can become very large in particular when the variables G and N^l come from the discretization of jump diffusion processes. In many situations, it is advisable to reduce the dimension of the space in which the optimization problem is solved.

Reducing the dimension of the optimization problem. Let $0 < d' \leq d$ and $0 < p' \leq p$ be the reduced dimensions. Instead of searching for the best importance sampling parameter (θ,λ) in the whole space $\mathbb{R}^d \times \mathbb{R}_+^{*\,p}$, we consider the subspace $\{(A\vartheta, B\lambda) : \vartheta \in \mathbb{R}^{d'},\ \lambda \in \mathbb{R}_+^{*\,p'}\}$ where $A \in \mathbb{R}^{d\times d'}$ is a matrix with rank $d' \leq d$ and $B \in \mathbb{R}_+^{*\,p\times p'}$ a matrix with rank $p' \leq p$. Note that since all the coefficients of B are non negative, for all $\vartheta \in \mathbb{R}_+^{*\,p'}$, $B\vartheta \in \mathbb{R}_+^{*\,p}$; actually, it is easily seen that the image of $\mathbb{R}_+^{*\,p'}$ through B is isomorphic to $\mathbb{R}_+^{*\,p'}$.
For such matrices A and B, we introduce the function $v^{A,B} : \mathbb{R}^{d'} \times \mathbb{R}_+^{*\,p'} \longmapsto \mathbb{R}$ defined by

$$v^{A,B}(\vartheta,\lambda) = v(A\vartheta, B\lambda) \tag{2.10}$$

The function $v^{A,B}$ inherits from the regularity and convexity properties of v. Hence, from Proposition 2.1, we know that $v^{A,B}$ is continuously infinitely differentiable and strongly convex. As a consequence, there exists a unique couple of minimizers $(\vartheta^{A,b}_\star, \lambda^{A,B}_\star)$ such that $v^{A,B}(\vartheta^{A,B}_\star, \lambda^{A,B}_\star) = \inf_{\vartheta\in\mathbb{R}^{d'}, \lambda\in\mathbb{R}^{*\,p'}_+} v^{A,B}(\vartheta, \lambda)$. We can also deduce the gradient vector of $v^{A,B}_n$

$$\nabla v^{A,B}(\vartheta, \lambda) = \begin{pmatrix} A^*\nabla_\theta(A\vartheta, B\lambda) \\ B^*\nabla_\lambda(A\vartheta, B\lambda) \end{pmatrix}$$

and its Hessian matrix

$$\nabla^2 v^{A,B}(\vartheta, \lambda) = \mathbb{E}\Bigg[F(G, A\vartheta, N^\mu, B\lambda) \begin{pmatrix} A^*(A\vartheta - G)(A\vartheta - G)^*A & A^*(A\vartheta - G)a(N^\mu, B\lambda)^*B \\ B^*a(N^\mu, B\lambda)(A\vartheta - G)^*A & B^*aa^*(N^\mu, B\lambda)B \end{pmatrix} + F(G, A\vartheta, N^\mu, B\lambda) \begin{pmatrix} A^*A & 0 \\ 0 & B^*DB \end{pmatrix} \Bigg]$$

where the function F is defined by Equation (2.7). For the particular choices $A = I_d$, $B = I_p$, $d = d'$ and $p = p'$, the functions v^{I_d, I_p} and v coincide.

The Esscher transform as a way to reduce the dimension. Consider a two dimensional process $(X_t)_{t\le T}$ of the form $X_t = (W_t, \tilde{N}^{\tilde{\mu}}_t)$ where W is a real Brownian motion and $\tilde{N}^{\tilde{\mu}}$ is a Poisson process with intensity $\tilde{\mu}$. The Esscher transform applied to X yields that for any nonnegative function h, we have the following equality $\forall\ \alpha \in \mathbb{R}, \tilde{\lambda} \in \mathbb{R}^*_+$,

$$\mathbb{E}[h((W_t, \tilde{N}^{\tilde{\mu}}_t),\ t \le T)] = \mathbb{E}\left[h((W_t + \alpha t, \tilde{N}^{\tilde{\lambda}},\ t \le T))\, \mathrm{e}^{-\alpha W_T - \frac{|\alpha|^2 T}{2}}\, \mathrm{e}^{T(\tilde{\lambda} - \tilde{\mu})} \left(\frac{\tilde{\mu}}{\tilde{\lambda}}\right)^{\tilde{N}^{\tilde{\lambda}}_T} \right]$$

Let $0 = t_0 < \cdots < t_p = T$ be a time grid of $[0, T]$. If we consider the vector G (resp. N^μ) as the increments of W (resp. $\tilde{N}^{\tilde{\mu}}$) on the grid, we can recover a particular form of Equation (1.2) with $A, B \in \mathbb{R}^p$ given by

$$A = \left(\sqrt{t_1},\ \sqrt{t_2 - t_1}, \ldots,\ \sqrt{t_p - t_{p-1}}\right)^*; \quad B = \left(t_1, t_2 - t_1, \ldots, t_p - t_{p-1}\right)^*.$$

2.2 Tracking the optimal importance sampling parameter

The optimal importance sampling parameter (θ^*, λ^*) can characterized as the unique zero of an expectation, which is the typical framework for applying stochastic approximation. In particular, we could use the algorithm introduced by Chen and Zhu (1986); we refer to Lelong (2008, 2011) for a study of the convergence and asymptotic behaviour of these algorithms. The use of stochastic approximation for devising adaptive importance sampling method was deeply investigated in a recent survey by Lapeyre and Lelong (2011) who highlighted the difficulties in making those algorithms practically converge.

In this work, we adopt a totally different point of view often called *sample average approximation*, which basically consists in first replacing expectations by sample averages and then using deterministic optimization techniques on these empirical means. This approach was studied in the Gaussian framework by Jourdain and Lelong (2009) and proved to be very efficient.

Let $(G^j)_{j\ge1}$ be a sequence of $d-$dimensional independent and identically distributed standard normal random variables. We also introduce $(N^{\mu,j})_{j\ge1}$ a sequence of $p-$ dimensional independent and

identically distributed random vector following the law of N^μ, ie. the components of the vectors are independent and Poisson distributed with parameter μ.
For $n \geq 1$, we introduce the sample average approximation of the function $v^{A,B}$ defined by

$$v_n^{A,B}(\vartheta,\lambda) = \frac{1}{n}\sum_{j=1}^{n} f(G^j, N^{\mu,j})^2\, \mathrm{e}^{-A\vartheta\cdot G^j + \frac{|A\vartheta|^2}{2}} \prod_{i=1}^{p} \mathrm{e}^{(B\lambda)_i-\mu_i}\left(\frac{\mu_i}{(B\lambda)_i}\right)^{N_i^{\mu_i,j}}. \tag{2.11}$$

For n large enough, $f(G^j, N^{\mu,j}) \neq 0$ for some index $j \in \{1,\dots,n\}$ and the approximation $v_n^{A,B}$ is also strongly convex and hence admits a unique minimizer $(\vartheta_n^{A,B}, \lambda_n^{A,B})$ defined by $v_n^{A,B}(\vartheta_n^{A,B}, \lambda_n^{A,B}) = \inf_{\vartheta\in\mathbb{R}^{d'},\lambda\in\mathbb{R}_+^{*\,p'}} v_n^{A,B}(\vartheta,\lambda)$.

Proposition 2.2. *Under Assumption p0](A1), the sequence of random functions $(v_n^{A,B})_n$ converges a.s. locally uniformly to the continuous function $v^{A,B}$.*

To prove this result, we use the uniform strong law of large numbers recalled hereafter, see for instance Rubinstein and Shapiro (1993, Lemma A1). This result is also a consequence of the strong law of large numbers in Banach spaces Ledoux and Talagrand (1991, Corollary 7.10, page 189).

Lemma 2.3. *Let $(X_i)_{i\geq 1}$ be a sequence of i.i.d. $\mathbb{R}^m$-valued random vectors, E an open set of $\mathbb{R}^d$ and $h : E\times\mathbb{R}^m \to \mathbb{R}$ be a measurable function. Assume that*

- *a.s., $\chi \in E \mapsto h(\chi, X_1)$ is continuous,*
- *for all compact sets K of $\mathbb{R}^d$ such that $K \subset E$, $\mathbb{E}\left(\sup_{\chi\in K}|h(\chi,X_1)|\right) < +\infty$.*

Then, a.s. the sequence of random functions $\chi \in K \mapsto \frac{1}{n}\sum_{i=1}^n h(\chi, X_i)$ converges locally uniformly to the continuous function $\chi \in E \mapsto \mathbb{E}(h(\chi, X_1))$.

Proof of Proposition 2.2. It is sufficient to prove the result for v_n and it will hold for $v_n^{A,B}$. Let $M > \underline{m} > 0$. For all (θ,λ) such that $|(\theta,\lambda)| \leq M$ and $d_0(\lambda) > \underline{m}$, we have

$$f(G,N^\mu)^2\,\mathrm{e}^{-\theta\cdot G+\frac{|\theta|^2}{2}}\prod_{i=1}^{p}\mathrm{e}^{\lambda_i-\mu_i}\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}} \leq f(G,N^\mu)^2\prod_{k=1}^{d}(\mathrm{e}^{-MG_k}+\mathrm{e}^{MG_k})\,\mathrm{e}^{\frac{M^2}{2}}\prod_{i=1}^{p}\mathrm{e}^{M-\mu_i}\left(\frac{\mu_i}{\underline{m}}\right)^{N_i^{\mu_i}}.$$

The r.h.s. is integrable by p0](A1) and Hölder's inequality; hence, we can apply Lemma 2.3. ∎

Proposition 2.4. *Under Assumption p0](A1), the pair $(\vartheta_n^{A,B}, \lambda_n^{A,B})$ converges a.s. to $(\vartheta_\star^{A,B}, \lambda_\star^{A,B})$ as $n \longrightarrow +\infty$. Moreover, if*

(A2) $\exists\delta > 0, \quad \mathbb{E}\left[|f(G,N^\mu)|^{4+\delta}\right] < \infty,$

$\sqrt{n}\left((\vartheta_n^{A,B},\lambda_n^{A,B}) - (\vartheta_\star^{A,B},\lambda_\star^{A,B})\right)$ converges in law to the normal distribution $\mathcal{N}_{d+p}(0,\Gamma)$ where

$$\Gamma = \left(\nabla^2 v^{A,B}(\vartheta_\star^{A,B},\lambda_\star^{A,B})\right)^{-1}\mathrm{Cov}(\nabla F(G, A\vartheta_\star^{A,B}, N^\mu, B\lambda_\star^{A,B}))\left(\nabla^2 v^{A,B}(\vartheta_\star^{A,B},\lambda_\star^{A,B})\right)^{-1}$$

with the function F defined by Equation (2.7) and its gradient computed w.r.t. the reduced parameters (ϑ,λ).

Condition tegrable2](A2) ensures that the covariance matrix $\mathrm{Cov}(\nabla F(G, A\vartheta_\star^{A,B}, N^\mu, B\lambda_\star^{A,B}))$ does exist. The non singularity of the matrix $\nabla^2 v^{A,B}(\vartheta_\star^{A,B},\lambda_\star^{A,B})$ is guaranteed by the strict convexity of v.

By combining Propositions 2.2 and 2.4, we can state the following result

Corollary 2.5. *Under Assumption p0](A1),* $v_n^{A,B}(\vartheta_n^{A,B},\lambda_n^{A,B})$ *converge a.s. to* $v^{A,B}(\vartheta_\star^{A,B},\lambda_\star^{A,B})$ *as* $n \longrightarrow +\infty$.

Proof of Proposition 2.4. Let $\varepsilon > 0$. We define a compact neighbourhood $\mathcal{V}_\varepsilon$ of $(\vartheta_\star,\lambda_\star)$

$$\mathcal{V}_\varepsilon \stackrel{\text{def}}{=} \left\{(\vartheta,\lambda) \in \mathbb{R}^d \times \mathbb{R}^p \ : \ |(\vartheta,\lambda) - (\vartheta_\star,\lambda_\star)| \le \varepsilon\right\}. \tag{2.12}$$

In the following, we assume that ε is small enough, so that $\mathcal{V}_\varepsilon$ is included in $\mathbb{R}^d \times \mathbb{R}_+^{*\,p}$.
By the strict convexity and the continuity of $v^{A,B}$,

$$\alpha \stackrel{\text{def}}{=} \inf_{(\vartheta,\lambda)\in\mathcal{V}_\varepsilon^c} v^{A,B}(\vartheta,\lambda) - v^{A,B}(\vartheta_\star^{A,B},\lambda_\star^{A,B}) > 0.$$

The local uniform convergence of $v_n^{A,B}$ to $v^{A,B}$ ensures that for some n_α sufficiently large,

$$\forall n \ge n_\alpha,\ \forall(\vartheta,\lambda) \in \mathcal{V}_\varepsilon,\ |v_n^{A,B}(\vartheta,\lambda) - v^{A,B}(\vartheta,\lambda)| \le \frac{\alpha}{3}. \tag{2.13}$$

For $n \ge n_\alpha$ and $(\vartheta,\lambda) \notin \mathcal{V}_\varepsilon$, we define $(\vartheta_\varepsilon^{A,B},\lambda_\varepsilon^{A,B}) \in \mathcal{V}_\varepsilon$ and writes as the convex combination of $(\vartheta_\star^{A,B},\lambda_\star^{A,B})$ and (ϑ,λ).

$$(\vartheta_\varepsilon^{A,B},\lambda_\varepsilon^{A,B}) \stackrel{\text{def}}{=} \left(\vartheta_\star^{A,B} + \varepsilon\frac{\vartheta-\vartheta_\star^{A,B}}{|(\vartheta-\vartheta_\star^{A,B},\lambda-\lambda_\star^{A,B})|}, \lambda_\star^{A,B} + \varepsilon\frac{\mu-\lambda_\star^{A,B}}{|(\vartheta-\vartheta_\star^{A,B},\lambda-\lambda_\star^{A,B})|}\right).$$

We deduce, using the convexity of $v_n^{A,B}$ for the first inequality and Equation (2.13) for the second one

$$\begin{aligned} v_n^{A,B}(\vartheta,\lambda) - v_n^{A,B}(\vartheta_\star^{A,B},\lambda_\star^{A,B}) &\ge \frac{|(\vartheta-\vartheta_\star^{A,B},\lambda-\lambda_\star^{A,B})|}{\varepsilon}\left[v_n^{A,B}(\vartheta_\varepsilon^{A,B},\lambda_\varepsilon^{A,B}) - v_n^{A,B}(\vartheta_\star^{A,B},\lambda_\star^{A,B})\right] \\ &\ge \left[v^{A,B}(\vartheta_\varepsilon^{A,B},\lambda_\varepsilon^{A,B}) - v^{A,B}(\vartheta_\star^{A,B},\lambda_\star^{A,B}) - \frac{2\alpha}{3}\right] \ge \frac{\alpha}{3}. \end{aligned}$$

The optimality of $(\vartheta_n^{A,G},\lambda_n^{A,B})$ yields that $v_n^{A,B}(\vartheta_n^{A,B},\lambda_n^{A,B}) \le v_n^{A,B}(\vartheta_\star^{A,B},\lambda_\star^{A,B})$. So, we conclude that $(\vartheta_n^{A,B},\lambda_n^{A,B}) \in \mathcal{V}_\varepsilon$ for $n \ge n_\alpha$. Therefore, $(\vartheta_n^{A,B},\lambda_n^{A,B})$ converges a.s. to $(\vartheta_\star^{A,B},\lambda_\star^{A,B})$.
We have seen in the proof of Proposition 2.1, that $\mathbb{E}\left[\sup_{|(\theta,\lambda)|\le M,\underline{m}<d_0(\lambda)} \nabla F(G,\theta,N^\mu,\lambda)\right] < \infty$, see Equation (2.9) and (2.8). Similarly, we can show that $\mathbb{E}\left[\sup_{|(\theta,\lambda)|\le M,\underline{m}<d_0(\lambda)} \nabla^2 F(G,\theta,N^\mu,\lambda)\right] < \infty$. The central limit theorem governing the convergence of the pair $(\vartheta_n^{A,B},\lambda_n^{A,B})$ to the pair $(\vartheta_\star^{A,B},\lambda_\star^{A,B})$ can be deduced from Rubinstein and Shapiro (1993, Theorem A2). ∎

3 Adaptive Monte Carlo

In this section, we assume to have at hand a sequence of optimal solutions $(\vartheta_n^{A,B},\lambda_n^{A,B})$ and want to devise an adaptive Monte Carlo taking advantage of the knowledge of these parameters through the use of Equation (1.2). In a previous work Jourdain and Lelong (2009) dedicated to the Gaussian framework, we had used the same samples for approximating v by v_n and after to build a Monte Carlo estimator of $\mathcal{E}$ involving θ_n. This was possible because a normal random vector X with mean vector θ naturally writes as $X = \theta + G$ where G is a standard normal random vector.
No such simple relation exists for the Poisson distribution to link a Poisson random variable with parameter μ to one with parameter λ. Hence, it is not worth trying to reuse, for the Monte Carlo estimator based on Equation (1.2), the same Poisson random samples as those involved in v_n. Then, we suggest the following two stages algorithm.

Algorithm 3.1.

First stage *Generate a sequence* $(G^j)_{j=1,\dots,m}$ *of i.i.d random vector following the standard normal distribution in* $\mathbb{R}^d$ *and a sequence* $(N^j = (N_1^j, \dots, N_p^j))_{j=1,\dots,m}$ *of i.i.d Poisson random vectors with parameter* μ.
Define

$$v_m^{A,B}(\vartheta, \lambda) = \frac{1}{m} \sum_{j=1}^{m} f(G^j, N^{\mu,j})^2 \, \mathrm{e}^{-A\vartheta \cdot G^j + \frac{|A\vartheta|^2}{2}} \prod_{i=1}^{p} \mathrm{e}^{B\lambda_i - \mu_i} \left(\frac{\mu_i}{B\lambda_i} \right)^{N_i^{\mu_i, j}}. \tag{3.1}$$

Compute

$$(\vartheta_m, \lambda_m) = \arg \min_{(\vartheta,\lambda) \in \mathbb{R}^d \times \mathbb{R}_+^{*\,p}} v_m^{A,B}(\vartheta, \lambda).$$

Second stage: *Generate a sequence* $(\bar{G}^j)_{j=1,\dots,n}$ *of i.i.d random vector following the standard normal distribution in* $\mathbb{R}^d$ *and a sequence* $(\bar{N}^j = (\bar{N}_1^j, \dots, \bar{N}_p^j))_{j=1,\dots,n}$ *of i.i.d Poisson random vectors with parameter* $B\lambda_m$. *Conditionally on* λ_m, *these two sequences are assumed to be independent of the sequences* $(G^j)_{j=1,\dots,m}$ *and* $(N^{\mu,j})_{j=1,\dots,m}$
Define

$$M_{n,m}^{A,B} = \frac{1}{n} \sum_{j=1}^{n} f(\bar{G}^j + A\vartheta_m, \bar{N}^j) \, \mathrm{e}^{-A\vartheta_m \cdot \bar{G}^j - \frac{|A\vartheta_m|^2}{2}} \prod_{i=1}^{p} \mathrm{e}^{(B\lambda_m)_i - \mu_i} \left(\frac{\mu_i}{(B\lambda_m)_i} \right)^{\bar{N}_i^j}. \tag{3.2}$$

3.1 Strong law of large numbers and central limit theorem

The conditional independence between the two stages combined with Lemma 1.1 immediately shows that for any fixed m, n, the estimator $M_{n,m}^{A,B}$ is unbiased, ie. $\mathbb{E}[M_{n,m}^{A,B}] = \mathcal{E}$. Conditionally on $(G_j, N_j)_{j=1,\dots,m}$, the terms involved in the sum of Equation (3.2) are i.i.d., hence the standard strong law of large numbers yields that $\lim_{n\to+\infty} M_{n,m}^{A,B} = \mathbb{E}[f(G, N^\mu)]$ a.s. by applying Lemma 1.1. Similarly, the central limit theorem applies and we can state the following result.

Proposition 3.2. *For any fixed* m, $M_{n,m}^{A,B}$ *converges a.s. to* $\mathbb{E}[f(G, N^\mu)]$ *as* n *goes to infinity and moreover* $\sqrt{n}(M_{n,m}^{A,B} - \mathbb{E}[f(G, N^\mu)]) \xrightarrow[n\to+\infty]{law} \mathcal{N}(0, v^{A,B}(\vartheta_m, \lambda_m))$.

This result is not fully satisfactory as from a practical point of view, we like to know the limiting of the estimator $M_{n,m(n)}^{A,B}$ where $m(n)$ is a function of n tending to infinity with n. To investigate the asymptotic behaviour when m and n tend to infinity together, it is convenient to rewrite $M_{n,m(n)}^{A,B}$ using an auxiliary sequence of random variables. We introduce a sequence $(\bar{U}_i^j)_{1\le i\le p, j\ge 1}$ of i.i.d. random variables following the uniform distribution on $[0, 1]$ and independent of all the other random variables used so far. If we define

$$\tilde{N}_i^j(\lambda) = \sum_{k=0}^{\infty} k \mathbf{1}_{\left\{P(\lambda_i;k) \le U_i^j < P(\lambda_i;k+1)\right\}} \quad \text{for all } 1 \le i \le p, \ 1 \le j$$

where $P(\lambda, \cdot)$ is the cumulative distribution function of the Poisson distribution with parameter λ, then $(\bar{N}^j)_{j=1,\dots,n} \overset{Law}{=} (\tilde{N}^j(\lambda_{m(n)}))_{j=1,\dots,n}$. Since for all $k \in \mathbb{N}$, the function $\lambda \in \mathbb{R}^* \longmapsto P(\lambda, k)$ is continuous and decreasing, we get that $\lim_{n\to\infty} \tilde{N}^j(\lambda_{m(n)}) = N^j(\lambda_\star)$ a.s. and for all $\lambda \le \lambda'$, $\tilde{N}^j(\lambda') < \tilde{N}^j(\lambda)$ where the ordering has to be understood component wise.

We define

$$\tilde{M}_n(\theta,\lambda) = \frac{1}{n}\sum_{j=1}^{n} f(\bar{G}^j+\theta, \tilde{N}^j(\lambda))\, \mathrm{e}^{-\theta\cdot\bar{G}^j-\frac{|\theta|^2}{2}} \prod_{i=1}^{p} \mathrm{e}^{\lambda_i-\mu_i}\left(\frac{\mu_i}{\lambda_i}\right)^{\tilde{N}_i^j(\lambda)}.$$

It is obvious that $M^{A,B}_{n,m(n)} \overset{Law}{=} \tilde{M}_n(A\vartheta_{m(n)}, B\lambda_{m(n)})$.

Theorem 3.3. *Let $m(n)$ be an increasing function of n tending to infinity. Then, under Assumptions p0](A1) and tegrable2](A2), $M^{A,B}_{n,m(n)}$ converges a.s. to $\mathbb{E}[f(G,N^\mu)]$ as n goes to infinity.*

It is actually sufficient to prove the result for A and B being identity matrices. For the sake of clear notations, when $A = I_d$ and $B = I_p$, we write $M_{n,m(n)}$ instead of $M^{A,B}_{n,m(n)}$.

Proof. We have already seen that $\mathbb{E}[M_{n,m}] = \mathcal{E}$. Thanks the independence of the samples used in the two stages of the algorithm, conditionally on $((G^j,N^j), j\geq 1)$, $M_{n,m}$ writes as a sum of i.i.d random variables. We introduce the $\sigma-$algebra $\mathcal{G} = \sigma((G^j,N^j), j\geq 1)$. We define for all $m, j \geq 1$

$$X_{m,j} = f(\bar{G}^j+\theta_m, \bar{N}^j)\, \mathrm{e}^{-\theta_m\cdot\bar{G}^j-\frac{|\theta_m|^2}{2}} \prod_{i=1}^{p} \mathrm{e}^{(\lambda_m)_i-\mu_i}\left(\frac{\mu_i}{(\lambda_m)_i}\right)^{\bar{N}_i^j}.$$

Note that conditionally on $\mathcal{G}$, the sequence $(X_{m,j})_{j\geq 1}$ is i.i.d. for any fixed $m \geq 1$.
For a fixed $\varepsilon > 0$, we recall the definition of $\mathcal{V}_\varepsilon$

$$\mathcal{V}_\varepsilon \overset{\text{def}}{=} \left\{(\theta,\lambda)\in\mathbb{R}^d\times\mathbb{R}^p \ : \ |(\theta,\lambda)-(\theta_\star,\lambda_\star)|\leq\varepsilon\right\}.$$

In the following, we assume that ε is small enough, so that $\mathcal{V}_\varepsilon$ is included in $\mathbb{R}^d\times\mathbb{R}_+^{*\,p}$. For all $m,n\geq 1$,

$$\begin{aligned}
\mathbb{E}\left[(M_{n,m}-\mathcal{E})^2\mathbf{1}_{\{(\theta_m,\lambda_m)\in\mathcal{V}_\varepsilon\}}\right] &= \mathbb{E}\left[\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}(X_{m,i}-\mathcal{E})\right)^2\Big|\mathcal{G}\right]\mathbf{1}_{\{(\theta_m,\lambda_m)\in\mathcal{V}_\varepsilon\}}\right]\\
&\leq \frac{1}{n}\,\mathbb{E}\left[\mathbb{E}\left[(X_{m,i}-\mathcal{E})^2\Big|\mathcal{G}\right]\mathbf{1}_{\{(\theta_m,\lambda_m)\in\mathcal{V}_\varepsilon\}}\right]\\
&\leq \frac{1}{n}\,\mathbb{E}\left[v(\theta_m,\lambda_m)\mathbf{1}_{\{(\theta_m,\lambda_m)\in\mathcal{V}_\varepsilon\}}\right]\\
&\leq \frac{1}{n}\left(\sup_{(\theta,\lambda)\in\mathcal{V}_\varepsilon} v(\theta,\lambda)-\mathcal{E}^2\right)\leq\frac{c}{n}.
\end{aligned} \tag{3.3}$$

We deduce from the Borell Cantelli Lemma that for any increasing function $\rho : \mathbb{N}\to\mathbb{N}$, $(M_{n^2,\rho(n)}-\mathcal{E})\mathbf{1}_{\{(\theta_{\rho(n)},\lambda_{\rho(n)})\in\mathcal{V}_\varepsilon\}}$ tends to 0 a.s.

To prove that $(M_{n,m(n)}-\mathcal{E})\mathbf{1}_{\{(\theta_{m(n)},\lambda_{m(n)})\in\mathcal{V}_\varepsilon\}}$ converges to zero a.s., we mimic the proof of the classical strong law of large numbers.
Let $n\in\mathbb{N}^*$, we define $k = \lfloor\sqrt{n}\rfloor$; then $k^2\leq n<(k+1)^2$.

$$\begin{aligned}
M_{n,m(n)}-\mathcal{E} &= \frac{1}{n}\sum_{i=1}^{k^2}(X_{m(n),i}-\mathcal{E}) + \frac{1}{n}\sum_{i=k^2+1}^{n}(X_{m(n),i}-\mathcal{E})\\
\left|M_{n,m(n)}-\mathcal{E}\right| &\leq \frac{1}{k^2}\left|\sum_{i=1}^{k^2}(X_{m(n),i}-\mathcal{E})\right| + \frac{1}{n}\left|\sum_{i=k^2+1}^{n}(X_{m(n),i}-\mathcal{E})\right|.
\end{aligned} \tag{3.4}$$

Using Equation (3.3),

$$\mathbb{E}\left[\left(\frac{1}{k^2}\left|\sum_{i=1}^{k^2}(X_{m(n),i}-\mathcal{E})\right|\right)^2 \mathbf{1}_{\{(\theta_m,\lambda_m)\in\mathcal{V}_\varepsilon\}}\right] \leq \frac{c}{k^2}.$$

Hence, we easily deduce from the Borrel Cantelli Lemma that $\frac{1}{k^2}\left|\sum_{i=1}^{k^2}(X_{m(n),i}-\mathcal{E})\right|\mathbf{1}_{\{(\theta_{m(n)},\lambda_{m(n)})\in\mathcal{V}_\varepsilon\}}$ tends to 0 a.s. when k goes to infinity, ie. when n goes to infinity. A similar computation as in Equation (3.3) leads to

$$\mathbb{E}\left[\left(\frac{1}{n}\left|\sum_{i=k^2+1}^{n}(X_{m(n),i}-\mathcal{E})\right|\right)^2 \mathbf{1}_{\{(\theta_{m(n)},\lambda_{m(n)})\in\mathcal{V}_\varepsilon\}}\right] \leq \frac{n-k^2}{n^2}\left(\sup_{(\theta,\lambda)\in K} v(\theta,\lambda)-\mathcal{E}^2\right) \leq \frac{c}{n^{3/2}}.$$

Hence, the Borel Cantelli Lemma yields that $\frac{1}{n}\left|\sum_{i=k^2+1}^{n}(X_{m(n),i}-\mathcal{E})\right|\mathbf{1}_{\{(\theta_{m(n)},\lambda_{m(n)})\in\mathcal{V}_\varepsilon\}} \to 0$ a.s. when n goes to infinity.
Eventually, we have proved that $(M_{n,m(n)}-\mathcal{E})\mathbf{1}_{\{(\theta_{m(n)},\lambda_{m(n)})\in\mathcal{V}_\varepsilon\}}$ converges to zero a.s. Since, $(\theta_{m(n)},\lambda_{m(n)})\to(\theta_\star,\lambda_\star)$ a.s., we deduce that $M_{n,m(n)}\to\mathcal{E}$ a.s. when n goes to infinity. ■

Theorem 3.4. *Let $m(n)$ be an integer valued function of n increasing to infinity with n and such that $m(n)\sim n^\beta$ for some $\beta>0$. Assume that*

(A3) *i. for all $k\in\mathbb{N}^p$, the function $g\in\mathbb{R}^d\longmapsto f(g,k)$ is continuous;*

ii. there exists a compact neighbourhood $\mathcal{V}$ of $(\vartheta_\star,\lambda_\star)$ included in $\mathbb{R}^{d'}\times\mathbb{R}_+^{*\,p'}$ and $\eta>0$ such that $\mathbb{E}\left[\sup_{(\vartheta,\lambda)\in\mathcal{V}}|f(\bar{G}+A\vartheta,\tilde{N}^1(B\lambda))|^{2(1+\eta)}\right]<\infty$.

Then, under Assumptions p0](A1) and tegrable2](A2),

$$\sqrt{n}(\tilde{M}_n(A\vartheta_{m(n)},B\lambda_{m(n)})-\mathbb{E}[f(G,N^\mu)]) \xrightarrow[n\to+\infty]{law} \mathcal{N}(0,v^{A,B}(\vartheta_\star,\lambda_\star)).$$

Proof. It is actually sufficient to prove the result for A and B being identity matrices.

$$\sqrt{n}(\tilde{M}_n(\theta_{m(n)},\lambda_{m(n)})-\mathcal{E}) = \sqrt{n}(\tilde{M}_n(\theta_\star,\lambda_\star)-\mathcal{E}) + \sqrt{n}(\tilde{M}_n(\theta_{m(n)},\lambda_{m(n)})-\tilde{M}_n(\theta_\star,\lambda_\star))$$

From the standard central limit theorem, $\sqrt{n}(\tilde{M}_n(\theta_\star,\lambda_\star)-\mathcal{E})\xrightarrow[n\to+\infty]{law}\mathcal{N}(0,v(\theta_\star,\lambda_\star))$. Therefore, it is sufficient to prove that $\sqrt{n}(\tilde{M}_n(\theta_{m(n)},\lambda_{m(n)})-\tilde{M}_n(\theta_\star,\lambda_\star))\xrightarrow[n\to+\infty]{Pr}0$. Let $\varepsilon>0$ and $0<\alpha<\beta/2$.

$$\begin{aligned}\mathbb{P}\left(\sqrt{n}\left|\tilde{M}_n(\theta_{m(n)},\lambda_{m(n)})-\tilde{M}_n(\theta_\star,\lambda_\star)\right|>\varepsilon\right) &\leq \mathbb{P}(n^\alpha\left|(\theta_{m(n)},\lambda_{m(n)})-(\theta_\star,\lambda_\star)\right|>1)\\ &+\frac{n}{\varepsilon^2}\mathbb{E}\left[\left|\tilde{M}_n(\theta_{m(n)},\lambda_{m(n)})-\tilde{M}_n(\theta_\star,\lambda_\star)\right|^2\mathbf{1}_{\{|(\theta_{m(n)},\lambda_{m(n)})-(\theta_\star,\lambda_\star)|\leq n^{-\alpha}\}}\right].\end{aligned}$$

Note that $n^\alpha\sim m(n)^{\alpha/\beta}$ with $\alpha/\beta<1/2$, hence we deduce from Proposition 2.4, that $\mathbb{P}(n^\alpha\left|(\theta_{m(n)},\lambda_{m(n)})-(\theta_\star,\lambda_\star)\right|>1)\longrightarrow 0$. We define

$$Q(\theta,\lambda)=\mathrm{e}^{-\theta\cdot\bar{G}^1-\frac{|\theta|^2}{2}}\prod_{i=1}^{p}\mathrm{e}^{\lambda_i-\mu_i}\left(\frac{\mu_i}{\lambda_i}\right)^{\tilde{N}_i^1(\lambda)}.$$

Conditionally on $(\theta_{m(n)}, \lambda_{m(n)})$, $\tilde{M}_n(\theta_{m(n)}, \lambda_{m(n)})$ writes as a sum of i.i.d random variables.

$$n\mathbb{E}\left[\left|\tilde{M}_n(\theta_{m(n)}, \lambda_{m(n)}) - \tilde{M}_n(\theta_\star, \lambda_\star)\right|^2 \mathbf{1}_{\{|(\theta_{m(n)}, \lambda_{m(n)})-(\theta_\star, \lambda_\star)|\leq n^{-\alpha}\}}\right] =$$
$$\mathbb{E}\Bigg[\left|f(\bar{G}^1 + \theta_\star, \tilde{N}^1(\lambda_\star))Q(\theta_\star, \lambda_\star) - f(\bar{G}^1 + \theta_{m(n)}, \tilde{N}^1(\lambda_{m(n)}))Q(\theta_{m(n)}, \lambda_{m(n)})\right|^2$$
$$\mathbf{1}_{\{|(\theta_{m(n)}, \lambda_{m(n)})-(\theta_\star, \lambda_\star)|\leq n^{-\alpha}\}}\Bigg]. \tag{3.5}$$

Thanks to the convergence of $\tilde{N}^1(\lambda_{m(n)})$, $Q(\theta_{m(n)}, \lambda_{m(n)})$ converges a.s. to $Q(\theta_\star, \lambda_\star)$ when n goes to infinity. Since for n large enough, $N^1(\lambda_{m(n)}) = N^1(\lambda_\star)$, the continuity of f with respect to its first argument enables to prove that $f(\bar{G}^1 + \theta_{m(n)}, \tilde{N}^1(\lambda_{m(n)}))$ converges a.s. to $f(\bar{G}^1 + \theta_\star, \tilde{N}^1(\lambda_\star))$. Hence, the absolute value inside the expectation tends to zero a.s. We need to bound the term inside the expectation by an integrable random variable to apply the bounded convergence theorem which yields the result.

$$\left|f(\bar{G}^1 + \theta_\star, \tilde{N}^1(\lambda_\star))Q(\theta_\star, \lambda_\star) - f(\bar{G}^1 + \theta_{m(n)}, \tilde{N}^1(\lambda_{m(n)}))Q(\theta_{m(n)}, \lambda_{m(n)})\right|^2$$
$$\mathbf{1}_{\{|(\theta_{m(n)}, \lambda_{m(n)})-(\theta_\star, \lambda_\star)|\leq n^{-\alpha}\}} \leq 2 \sup_{|(\theta,\lambda)-(\theta_\star, \lambda_\star)|\leq n^{-\alpha}} \left|f(\bar{G}^1 + \theta, \tilde{N}^1(\lambda))\right|^2 Q^2(\theta, \lambda).$$

For n large enough, $\{|(\theta, \lambda) - (\theta_\star, \lambda_\star)| \leq n^{-\alpha}\} \subset \mathcal{V}$. Moreover, there exist $\underline{m} > 0$ and $M > 0$ such that $\mathcal{V} \subset \{|\theta| \leq M, |\lambda| \leq M \text{ and } d_0(\lambda) \geq \underline{m}\}$. Hence,

$$\sup_{(\theta,\lambda)-(\theta_\star, \lambda_\star)|\leq n^{-\alpha}} \left|f(\bar{G}^1 + \theta, \tilde{N}^1(\lambda))\right| Q(\theta, \lambda)$$
$$\leq \sup_{(\theta,\lambda)\in\mathcal{V}} \left|f(\bar{G}^1 + \theta, \tilde{N}^1(\lambda))\right| \mathrm{e}^{pM} \prod_{i=1}^{d} (\mathrm{e}^{-MG_i^1} + \mathrm{e}^{MG_i^1}) \prod_{i=1}^{p} \left(\left(\frac{\mu_i}{\underline{m}}\right)^{\tilde{N}_i^1(\underline{m})} + \left(\frac{\mu_i}{\underline{m}}\right)^{\tilde{N}_i^1(M)}\right)$$
$$\leq \sum_{\substack{\sigma\in\{-M,M\}^d \\ \nu\in\{\underline{m},M\}^p}} \sup_{(\theta,\lambda)\in\mathcal{V}} \left|f(\bar{G}^1 + \theta, \tilde{N}^1(\lambda))\right| \mathrm{e}^{pM}\, \mathrm{e}^{\sigma\cdot G^1} \prod_{i=1}^{p} \left(\frac{\mu_i}{\underline{m}}\right)^{\tilde{N}_i^1(\nu)}.$$

Then, using Hölder's inequality we get

$$\mathbb{E}\left[\sum_{\substack{\sigma\in\{-M,M\}^d \\ \nu\in\{\underline{m},M\}^p}} \sup_{(\theta,\lambda)\in\mathcal{V}} \left|f(\bar{G}^1 + \theta, \tilde{N}^1(\lambda))\right|^2 \left(\mathrm{e}^{pM}\, \mathrm{e}^{\sigma\cdot G^1} \prod_{i=1}^{p} \left(\frac{\mu_i}{\underline{m}}\right)^{\tilde{N}_i^1(\nu)}\right)^2\right]$$
$$\leq \sum_{\substack{\sigma\in\{-M,M\}^d \\ \nu\in\{\underline{m},M\}^p}} \mathbb{E}\left[\sup_{(\theta,\lambda)\in\mathcal{V}} \left|f(\bar{G}^1 + \theta, \tilde{N}^1(\lambda))\right|^{2(1+\eta)}\right]^{\frac{1}{1+\eta}} \mathbb{E}\left[\left(\mathrm{e}^{pM}\, \mathrm{e}^{\sigma\cdot G^1} \prod_{i=1}^{p} \left(\frac{\mu_i}{\underline{m}}\right)^{\tilde{N}_i^1(\nu)}\right)^{2+\frac{2}{\eta}}\right]^{\frac{\eta}{1+\eta}}.$$

Since we have assumed that $\mathbb{E}\left[\sup_{(\theta,\lambda)\in\mathcal{V}} \left|f(\bar{G}^1 + \theta, \tilde{N}^1(\lambda))\right|^{2(1+\eta)}\right] < \infty$, the random variables $\left|f(\bar{G}^1 + \theta_\star, \tilde{N}^1(\lambda_\star))Q(\theta_\star, \lambda_\star) - f(\bar{G}^1 + \theta_{m(n)}, \tilde{N}^1(\lambda_{m(n)}))Q(\theta_{m(n)}, \lambda_{m(n)})\right|^2 \mathbf{1}_{\{|(\theta_{m(n)}, \lambda_{m(n)})-(\theta_\star, \lambda_\star)|\leq n^{-\alpha}\}}$ are uniformly bounded w.r.t n by an integrable random variable. Hence, the left hand side of Equation (3.5) tends to zero which achieves to prove that $\sqrt{n}(\tilde{M}_n(\theta_{m(n)}, \lambda_{m(n)}) - \tilde{M}_n(\theta_\star, \lambda_\star)) \xrightarrow[n\to+\infty]{Pr} 0$. ■

3.2 Practical implementation

The difficult part of Algorithm 3.1 is the numerical computation of the minimizing pair (θ_m, λ_m). The efficiency of the optimization algorithm depends very much on the magnitude of the smallest eigenvalue of $\nabla^2 v$. From the end of the proof of Proposition 2.1, we can deduce that the smallest eigenvalue of $\nabla^2 v$ is larger than

$$\mathbb{E}\left[F(G,\theta,N^\mu,\lambda)\mathbf{1}_{\{N^\mu=(n_1,\dots,n_p)\}}\right]\min\left(1,\frac{n_1}{\lambda_1^2},\dots,\frac{n_p}{\lambda_p^2}\right).$$

This lower bound depends on the function f whereas we would rather find a uniform lower bound. Hence, we advice to rewrite ∇v as

$$\begin{aligned}\nabla v(\theta,\lambda) =& \mathbb{E}\left[\begin{pmatrix}\theta\\ 1_p\end{pmatrix} f(G,N^\mu)^2\,\mathrm{e}^{-\theta\cdot G+\frac{|\theta|^2}{2}}\prod_{i=1}^p \mathrm{e}^{\lambda_i-\mu_i}\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]\\ &-\mathbb{E}\left[\begin{pmatrix}G\\ \frac{N^\mu}{\lambda}\end{pmatrix} f(G,N^\mu)^2\,\mathrm{e}^{-\theta\cdot G+\frac{|\theta|^2}{2}}\prod_{i=1}^p \mathrm{e}^{\lambda_i-\mu_i}\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]\end{aligned}$$

where $\frac{N^\mu}{\lambda} = \left(\frac{N_1^{\mu_1}}{\lambda_1},\dots,\frac{N_p^{\mu_p}}{\lambda_p}\right)$. Hence, $(\theta^\star,\lambda^\star)$ can be seen as the root of

$$\nabla u(\theta,\lambda) = \begin{pmatrix}\theta\\ 1_p\end{pmatrix} - \frac{\mathbb{E}\left[\begin{pmatrix}G\\ \frac{N^\mu}{\lambda}\end{pmatrix} f(G,N^\mu)^2\,\mathrm{e}^{-\theta\cdot G}\prod_{i=1}^p\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]}{\mathbb{E}\left[f(G,N^\mu)^2\,\mathrm{e}^{-\theta\cdot G}\prod_{i=1}^p\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]}$$

with $u(\theta,\lambda) = \frac{|\theta|^2}{2}+\sum_{i=1}^p\lambda_i+\log\mathbb{E}\left[f(G,N^\mu)^2\,\mathrm{e}^{-\theta\cdot G}\prod_{i=1}^p\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]$. The Hessian matrix of u is given by

$$\begin{aligned}\nabla^2 u(\theta,\lambda) =& \begin{pmatrix} I_d & 0\\ 0 & \frac{\mathbb{E}\left[Df(G,N^\mu)^2\,\mathrm{e}^{-\theta\cdot G+}\prod_{i=1}^p\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]}{\mathbb{E}\left[f(G,N^\mu)^2\,\mathrm{e}^{-\theta\cdot G}\prod_{i=1}^p\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]}\end{pmatrix}\\ &+\frac{\mathbb{E}\left[\begin{pmatrix}G\\ \frac{N^\mu}{\lambda}\end{pmatrix}\begin{pmatrix}G\\ \frac{N^\mu}{\lambda}\end{pmatrix}^* f(G,N^\mu)^2\,\mathrm{e}^{-\theta\cdot G}\prod_{i=1}^p\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]}{\mathbb{E}\left[f(G,N^\mu)^2\,\mathrm{e}^{-\theta\cdot G}\prod_{i=1}^p\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]}\\ &-\frac{\mathbb{E}\left[\begin{pmatrix}G\\ \frac{N^\mu}{\lambda}\end{pmatrix} f(G,N^\mu)^2\,\mathrm{e}^{-\theta\cdot G}\prod_{i=1}^p\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]\mathbb{E}\left[\begin{pmatrix}G\\ \frac{N^\mu}{\lambda}\end{pmatrix} f(G,N^\mu)^2\,\mathrm{e}^{-\theta\cdot G}\prod_{i=1}^p\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]^*}{\mathbb{E}\left[f(G,N^\mu)^2\,\mathrm{e}^{-\theta\cdot G}\prod_{i=1}^p\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]^2}\end{aligned}$$

where we recall that the diagonal matrix D is defined by $D = \mathrm{diag}_p\left(\frac{N_1^{\mu_1}}{\lambda_1^2},\dots,\frac{N_p^{\mu_p}}{\lambda_p^2}\right)$. The Cauchy Schwartz inequality yields that the last two terms in the expression of $\nabla^2 u$ form a positive semi definite

matrix. The first part of the Hessian is a positive definite matrix with smallest eigenvalue larger than

$$\min\left(1, \frac{1}{\lambda_j^2}\frac{\mathbb{E}\left[N_i^{\mu_i} f(G,N^\mu)^2\, \mathrm{e}^{-\theta\cdot G}\prod_{i=1}^p\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]}{\mathbb{E}\left[f(G,N^\mu)^2\, \mathrm{e}^{-\theta\cdot G}\prod_{i=1}^p\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]}\right)$$
$$= \min\left(1, \frac{\mu_j}{\lambda_j^3}\frac{\mathbb{E}\left[f(G,N^\mu+e_j)^2\, \mathrm{e}^{-\theta\cdot G}\prod_{i=1}^p\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]}{\mathbb{E}\left[f(G,N^\mu)^2\, \mathrm{e}^{-\theta\cdot G}\prod_{i=1}^p\left(\frac{\mu_i}{\lambda_i}\right)^{N_i^{\mu_i}}\right]}\right)$$

where the equality comes from Stein's formula for Poisson random variables and e_j denotes the $j-th$ element of the canonical basis. When the function f is increasing with respect to each component of its second argument, then we come up with the following lower bound independent of the function f

$$\min\left(1, \frac{\mu_j}{\lambda_j^3}\right).$$

Our numerical experiments advocate the use of u instead of v to speed up the computation of $(\theta^\star, \lambda^\star)$. Using this new expression, we implement Algorithm 1 to construct an approximation x_n^k of (θ_n, λ_n). Since u_n is strongly convex, for any fixed n, x_n^k converges to (θ_n, λ_n) when k goes to infinity. The direction of descent d_n^k at step k should be computed as the solution of a linear system. There is no point in computing the inverse of $\nabla^2 u_n(x_n^k)$, which would be computationally much more expensive.
Remarks on the implementation : From a practical point of view, ε should be chosen reasonably small $\varepsilon \approx 10^{-6}$. This algorithm converges very quickly and, in most cases, less than 5 iterations are enough to get a very accurate estimate of (θ_n, λ_n), actually within the $\varepsilon-$error. Since the points at which the function f is evaluated remain constant through the iterations of Newton's algorithm, the values $f^2(G^j, N^j)$ for $j = 1, \ldots, n$ should be precomputed before starting the optimization algorithm which considerably speeds up the whole process. The Hessian matrix of our problem is easily tractable so there is no point in using Quasi-Newton's methods.

Algorithm 1 Projected Newton's algorithm

Choose an initial value $x_n^0 \in \mathbb{R}^{d+p}$.
$k = 1$
while $\left|\nabla u_n(x_n^k)\right| > \varepsilon$ **do**
  1. Compute d_n^k such that $(\nabla^2 u_n(x_n^k))d_n^k = -\nabla u_n(x_n^k)$
  2. $x_n^{k+1/2} = x_n^k + d_n^k$
  for $i = 1 : d + p$ **do**
    if $x_n^{k+1/2}(i) > 0$ **then**
      $x_n^{k+1}(i) = x_n^{k+1/2}(i)$
    else
      $x_n^{k+1}(i) = \frac{x_n^k(i)}{2}$
    end if
  end for
  3. $k = k + 1$
end while

4 Application to jump processes in finance

We will apply our methodology to two different classes of jump processes: jump diffusion processes and stochastic volatility processes with jumps, in this latter case the volatility itself may jump also.
We consider a filtered probability space $(\Omega, \mathcal{A}, (\mathcal{F}_t)_{0\leq t\leq T}, \mathbb{P})$ with a finite time horizon $T > 0$ and I financial assets. We define on this space a Brownian motion W with values in $\mathbb{R}^I$ and $I+1$ independent Poisson processes $(N^1, \dots, N^{I+1})$ with constant intensities $\mu^1, \dots, \mu^{I+1}$. We also consider $(I+1)$ independent sequences $(Y_j^i)_{j\geq 1}$ for $i = 1 \dots I+1$ of i.i.d. real valued random variables with common law denoted Y in the following. The Poisson processes, the Brownian motions and the sequences $(Y_j^i)_j$ are supposed to be independent of each other. Actually, we are interested in considering the compound Poisson process associated to the Poisson process N^i and to the jump sequences Y^i for $i = 1, \dots, I+1$.

4.1 Jump diffusion processes

In this class of models, we assume that the log-prices evolve according to the following equation

$$X_t^i = \left(\beta^i - \frac{(\sigma^i)^2}{2}\right)t + \sigma^i L^i W_t + \sum_{j=1}^{N_t^i} Y_j^i + \sum_{j=1}^{N_t^{I+1}} Y_j^{I+1} \tag{4.1}$$

where $\beta = (\beta^i, \dots, \beta^I)^*$ is the drift vector and $\sigma = (\sigma^i, \dots, \sigma^I)^*$ the volatility vector. The row vectors L_i are such that the matrix $L = (L^1; \dots; L^I)$ verifies that $\Gamma = LL^*$ is a symmetric definite positive matrix with unit diagonal elements. The matrix Γ embeds the covariance structure of the continuous part of the model. We have also chosen to take into account in the model the possibility to have simultaneous jumps which explains the extra jump term $\sum_{j=1}^{N_t^{I+1}} Y_j^{I+1}$ common to all underlying assets. This common jump term embeds the systemic risk of the market.
From Equation 4.1, we deduce that the prices at time t $S_t^i = \mathrm{e}^{X_t^i}$ are defined by

$$S_t^i = S_0^i \exp\left\{\left(\beta^i - \frac{(\sigma^i)^2}{2}\right)t + \sigma^i L^i W_t\right\} \prod_{j=1}^{N_t^i} \mathrm{e}^{Y_j^i} \prod_{j=1}^{N_t^{I+1}} \mathrm{e}^{Y_j^{I+1}}$$

which corresponds for each asset to a one dimensional Merton model with intensity $\mu^i + \mu^{I+1}$ when the Y_j^i are normally distributed.
As, we assumed that $\mathbb{P}$ was the martingale measure associated to the risk free rate $r > 0$ supposed to be deterministic, the processes $(\mathrm{e}^{-rt} S_t)_t$ must be martingales under $\mathbb{P}$. This martingale condition imposes that for every $i = 1, \dots, I$,

$$\beta^i = r - (\mu^i \mathbb{E}[Y^i] + \mu^{I+1} \mathbb{E}[Y^{I+1}]).$$

In the following, β_i will always stand for this quantity.

Remark 4.1. *In the one dimensional case, ie. when $I = 1$, we only consider a single compound Poisson process as the systemic risk jump term becomes irrelevant. Hence, the log-price in dimension one will follow*

$$X_t = \left(\beta - \frac{\sigma^2}{2}\right)t + \sigma W_t + \sum_{j=1}^{N_t} Y_j.$$

For the sake of clearness, we will not treat the one dimensional case separately in the following, even though the practical one dimensional implementation relies on a single Poisson process. So, we will always consider that the Poisson process has values in $\mathbb{R}^{I+1}$.

In the numerical examples, we will need to discretize the multi dimensional price process on a time grid $0 = t_0 < t_1 < \cdots < t_J = T$. We will assume that this time grid is regular and given by $t_j = \frac{jT}{J}$, $j = 0, \ldots, J$. Just to fix our notations, we consider that the Brownian (resp. Poisson) increments are stored as a column vector with size $I \times J$ (resp. $(I+1) \times J$).

$$\begin{pmatrix} W_{t_1} \\ W_{t_2} \\ \vdots \\ W_{t_{J-1}} \\ W_{t_J} \end{pmatrix} = \begin{pmatrix} \sqrt{t_1} Id & 0 & 0 & \ldots & 0 \\ \sqrt{t_1} Id & \sqrt{t_2 - t_1} Id & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sqrt{t_{J-1} - t_{J-2}} Id & 0 \\ \sqrt{t_1} Id & \sqrt{t_2 - t_1} Id & \ldots & \sqrt{t_{J-1} - t_{J-2}} Id & \sqrt{t_J - t_{J-1}} Id \end{pmatrix} G,$$

where G is a normal random vector in $\mathbb{R}^{I \times J}$ and Id is the identity matrix in dimension $I \times I$. The Poisson process is discretized in a similar way.

The Merton jump diffusion model. The Merton model corresponds to the particular choice of a normal distribution for the variables (Y^i), $Y^i \sim \mathcal{N}(\alpha, \delta)$ where $\alpha \in \mathbb{R}$ and $\delta > 0$. In this framework, the jump sizes in the price follow a log normal distribution.

The Kou model. In the Kou model Kou (2002), the variables Y^i follow an asymmetric exponential distribution with density

$$p^i \mu_+^i \, \mathrm{e}^{-\mu_+^i x} \mathbf{1}_{\{x>0\}} + (1-p)^i \mu_-^i \, \mathrm{e}^{\mu_-^i x} \mathbf{1}_{\{x<0\}}$$

where $p^i \in [0,1]$ is the probability of a positive jump for the $i-th$ component and the variables $\mu_+^i > 0, \mu_-^i > 0$ govern the decay of each exponential part.

4.2 Stochastic volatility models with jumps

In this section, we consider the stochastic volatility type model developed by Barndorff-Nielsen and Shephard (2001b,a) in which the volatility process is a non Gaussian Ornstein Uhlenbeck driven by a compound Poisson process.
We consider that the log-prices satisfy for $i = 1, \ldots, I$

$$dX_t^i = (a^i - \sigma^i/2)dt + \sqrt{\sigma_{t^-}^i} dW_t^i + \psi^i dZ_{\kappa^i t}^i + \psi^{I+1} dZ_{\kappa^{I+1} t}^{I+1}$$

where $a \in \mathbb{R}^I$, $\psi \in \mathbb{R}^{I+1}$ has non-positive components which account for the positive leverage effect, Z is $(I+1)$-dimensional Lévy process defined by $Z_t^i = \sum_{k=1}^{N_t^i} Y_k^i$ for $i = 1, \ldots, I+1$ and the squared volatility process $(\sigma_t)_t$ is Lévy driven Ornstein Uhlenbeck

$$d\sigma_t^i = -(\kappa^i + \kappa^{I+1})\sigma_t^i dt + dZ_{\kappa^i t}^i + dZ_{\kappa^{I+1} t}^{I+1}.$$

For the squared volatility process to remain positive, we assume that the components of Z only jumps upward, which means that the random variables Y_j^i are non-negative.
More specifically, the jump sequence Y^i is i.i.d following the exponential distribution with parameter $\beta^i > 0$ for $i = 1, \ldots, I+1$. The drift vector a is chosen such that the discounted prices are martingales under $\mathbb{P}$. Hence, a straight computation shows that we need to set

$$a^i = r - \psi^i \frac{\kappa^i \mu^i}{\beta^i - \psi^i} - \psi^{I+1} \frac{\kappa^{I+1} \mu^{I+1}}{\beta^{I+1} - \psi^{I+1}}, \quad \text{for } i = 1, \ldots, I$$

to ensure the martingale property of $(e^{-rt} \exp X_t)_t$.
As in the section on jump diffusion models, the extra Poisson process giving raise to the term dZ^{I+1} in the dynamics of X and σ accounts for modelling a systemic risk. When Z^{I+1} jumps, all the volatilities and possibly all the assets (when there is a leverage effect) jump together. This parametrization of multi-dimensional stochastic volatility models with jumps corresponds to Section 5.3 of Barndorff-Nielsen and Stelzer (2013). Adding this extra jump process only makes sense in a multi-dimensional framework, hence we write the one-dimensional model using the previous equations but without the terms involving the index $I + 1$.
In the following, we compare the efficiencies of several different approaches based on the theoretical part of the paper in the context of option pricing with jumps. The problem always boils down to computing the expectation of a function of a jump diffusion process.

4.3 Several importance sampling approaches

To design an importance sampling Monte Carlo method, we can either play with the Brownian part — referred to hereafter as *Gaussian importance sampling* with an optimal variance denoted *VarG*, or with the Poisson part — referred to as *Poisson importance sampling* with an optimal variance denoted *VarP*, or with both at the same time. This last approach is named *Gaussian+Poisson importance sampling* and yields to an optimal variance denoted *VarGP*. The Gaussian importance sampling approach actually corresponds the methodology developed in Jourdain and Lelong (2009) but with independent sets of samples for the optimization part and the true Monte Carlo computation.

For each of the three methods, we consider two approaches.

Full importance sampling. The first approach consists in allowing to optimize the parameters per time steps, this means that $d = d' = I \times J$ and $p = p' = (I + 1) \times J$. In this setting, the matrices A and B are identity matrices. This approach is the more general in one framework, but the dimension of the optimization problem linked to the variance minimization with the square of the number of time steps, which yields some interest in trying to find a sub vector space with smaller dimension in which optimizing the parameters and which achieves a variance close the global minimum.

Reduced importance sampling. The idea of reducing the dimension of the problem is to search for the parameter (θ, λ) in the subspace $\{(A\vartheta, B\lambda) \ : \ \vartheta \in \mathbb{R}^{d'}, \ \lambda \in \mathbb{R}_+^{*\,p'}\}$ where $A \in \mathbb{R}^{d \times d'}$ is a matrix with rank $d' \leq d$ and $B \in \mathbb{R}_+^{*\,p \times p'}$ a matrix with rank $p' \leq p$.
We choose to restrict ourselves to adding a constant drift to the Brownian motion and keeping the Poisson intensity time independent. This corresponds to $d' = I$ and $p' = I + 1$

$$A_{(j-1)I+i,i} = \sqrt{t_j - t_{j-1}}, \quad B_{(j-1)(I+1)+k,k} = t_j - t_{j-1}$$

for $j = 1, \ldots, J$, $i = 1, \ldots, I$ and $k = 1, \ldots, I + 1$, all the other coefficients of A and B being zero.

4.4 Numerical experiments

We compare the different importance sampling approaches on four different financial derivatives: the first two examples are path-dependent single asset options while the last two examples are basket option with or without barrier monitoring. To compare the different strategies, we have decided to fix the number of samples for the Monte Carlo part, which implies that their accuracies only depend on their variances, which we will compare in different examples. To determine which method is best, it

is convenient to compute their efficiencies defined as the ratio of the variance divided by the CPU time.

In all the following examples, we use the same number of samples for the approximation of the optimal importance sampling parameters and for the Monte Carlo computation, ie. $m(n) = n$.

Asian option. We consider a discretely monitored Asian option with payoff

$$\left(\frac{1}{J}\sum_{i=1}^{J} S_{t_i} - K\right)_+ .$$

Our tests on one dimensional Asian options (see Tables 1 and 2) show that the *Poisson* and *Gaussian+Poisson* importance sampling methods perform generally better than the pure *Gaussian* importance sampling approach but they also require a longer computational time. When taking into account this extra computational times along with the variance reduction we notice that the *Poisson* and *Gaussian+Poisson* importance sampling methods yield the same efficiency for the Merton model (see Table 1). For the BNS model (see Table 2), the mixed *Gaussian+Poisson* importance sampling approach achieves a better variance reduction than the two other methods for a comparable computational time. By closely looking at the CPU times of the different strategies, it clearly appears that the reduced approach shows the better efficiency and should be used in practice.

	Strike	Price	Var	VarG	VarP	VarGP
Full	90	17.88	2639	2395	636	529
Reduced		17.88	2639	2640	839	752
Full	100	14.37	2750	2624	720	622
Reduced		14.37	2750	2624	552	470
Full	110	12.11	2327	2301	470	420
Reduced		12.11	2327	2301	676	585

Table 1: Discrete Asian option in dimension 1 in the Merton model with $S_0 = 100$, $r = 0.05$, $\sigma = 0.25$, $\mu = 1$, $\alpha = 0.5$, $\delta = 0.2$, $T = 1$, $J = 12$ and $n = 50000$. The CPU time for the crude Monte Carlo approach is 0.08. The CPU times for the full importance sampling approach are $(0.21, 0.28, 0.39)$ and for the reduced approach $(0.20, 0.21, 0.26)$.

	Strike	Price	Var	VarG	VarP	VarGP
Full	90	11.85	63	22.7	50	13.3
Reduced		11.85	63	28.7	52.7	22.1
Full	100	3.96	47	19	29.7	9.4
Reduced		3.96	47	22	33	14.7
Full	110	0.92	19	7.8	9	3.5
Reduced		0.92	19	10	11.1	5.56

Table 2: Discrete Asian option in dimension 1 in the BNS model with $S_0 = 100$, $r = 0.05$, $\lambda_0 = 0.01$, $\mu = 1$, $\kappa = 0.5474$, $\beta = 18.6$, $T = 1$, $J = 12$ and $n = 50000$. The CPU time for the crude Monte Carlo approach is 0.13. The CPU times for the full importance sampling approach are $(0.36, 0.52, 0.93)$ and for the reduced approach $(0.29, 0.29, 0.33)$.

Barrier option. We consider a discrete monitoring barrier option with payoff

$$(S_T - K)_+ \times \mathbf{1}_{\left\{\forall 1 \le j \le J,\, S_{t_j} < U\right\}}$$

where U is the upper barrier.

	Strike	Price	Var	VarG	VarP	VarGP
Full	90	17.88	2639	2395	636	529
Reduced		17.88	2639	2640	839	752
Full	100	14.37	2750	2624	720	622
Reduced		14.37	2750	2624	552	470
Full	110	12.11	2327	2301	470	420
Reduced		12.11	2327	2301	676	585

Table 3: Discrete barrier option in dimension 1 in the Merton model with $S_0 = 100$, $r = 0.05$, $\sigma = 0.2$, $\mu = 0.1$, $\alpha = 0$, $\delta = 0.1$, $T = 1$, $J = 12$, $U = 140$ and $n = 50000$. The CPU time for the crude Monte Carlo approach is 0.08. The CPU times for the full importance sampling approach are $(0.22, 0.26, 0.37)$ and for the reduced approach $(0.20, 0.20, 0.23)$.

Basket option. We consider a basket option on 10 assets with payoff

$$\left(\sum_{i=1}^{I} \omega^i S_T^i - K\right)_+$$

where the vector $\omega \in \mathbb{R}^I$ describes the weight of each asset in the basket.

Strike	Price	Var	VarG	VarP	VarGP
-10	10.61	112	85	66	48
0	3.66	85	66	33	25
10	1.17	111	52	12	10

Table 4: Basket option in dimension $I = 10$ in the Merton model with $S_0^i = 100$, $r = 0.05$, $\sigma^i = 0.2$, $\mu^i = 0.1$, $\alpha^i = 0.3$, $\delta^i = 0.2$, $\rho = 0.3$, $T = 1$, $\omega^i = \frac{1}{I}$ for $i = 1, \dots, I/2$, $\omega^i = -\frac{1}{I}$ for $i = I/2 + 1, \dots, I$ and $n = 50000$. The CPU time for the crude Monte Carlo approach is 0.06. The CPU times for the importance sampling approach are $(0.17, 0.20, 0.32)$.

The experiments on the one dimensional barrier option (see Table 3 lead to very similar conclusions regarding the efficiencies of the different approaches. Roughly speaking, the *Gaussian* approach does not bring any variance reduction but costs 2.5 times the CPU times of the crude Monte Carlo approach. The *Poisson* and *Gaussian+Poisson* importance sampling approaches do provide an impressive variance reductions for equivalent computational times at least in the reduced size approach. The improvement of the optimal variance obtained by the full size approaches does not look enough to counter balance the extra computational time. Actually, the reduced size approaches show far better efficiencies.
Since basket options are not path dependent derivatives, the full and reduced size approaches coincide and we do not distinguish between the two in Tables 4 and 5. In these tables, we can see that the *Gaussian+Poisson* approach provides better variance reductions that the pure *Poisson* approach, which in turn outperforms the pure *Gaussian* strategy. However, except for out of the money options, the gain

Strike	Price	Var	VarG	VarP	VarGP
-10	10.21	60	41	48	29
0	3.35	30	21	22	13
10	0.68	8.3	5.9	5.2	2.8

Table 5: Basket option in dimension $I = 10$ in the Merton model with $S_0^i = 100$, $r = 0.05$, $\sigma^i = 0.2$, $\mu^i = 1$, $\alpha^i = 0.1$, $\delta^i = 0.01$, $\rho = 0.3$, $T = 1$, $\omega^i = \frac{1}{I}$ for $i = 1, \dots, I/2$, $\omega^i = -\frac{1}{I}$ for $i = I/2 + 1, \dots, I$ and $n = 50000$. The CPU time for the crude Monte Carlo approach is 0.06. The CPU times for the importance sampling approach are $(0.17, 0.20, 0.32)$.

brought by the different importance sampling approaches do not compensate the extra computational time in order to keep up with the crude Monte Carlo strategy. This lack of efficiency mainly comes from the very simple form of the payoff which makes the crude Monte Carlo method very fast.

Multidimensional barrier option. We consider a discrete monitoring down and out barrier option on a basket of assets with payoff

$$\left(\sum_{i=1}^{I} \omega^i S_T^i - K \right)_+ \mathbf{1}_{\left\{ \forall 1 \le i \le I,\ \forall 1 \le j \le J,\ S_{t_j}^i > b^i \right\}}$$

where the vector $b \in \mathbb{R}^I$ is a lower barrier.

	Strike	Price	Var	VarG	VarP	VarGP
Full	0	0.59	7.00	4.03	4.36	1.98
Reduced		0.59	7.00	3.51	4.36	2.05
Full	-5	1.06	13.33	8.43	9.64	4.79
Reduced		1.06	13.33	8.56	9.81	5.42
Full	-10	1.64	24.26	16.57	18.39	10.57
Reduced		1.64	24.26	16.77	18.96	11.68

Table 6: Barrier option in dimension $I = 10$ in the Merton model with $S_0^i = 100$, $r = 0.05$, $\sigma^i = 0.2$, $\mu^i = 1$, $\alpha^i = 0.1$, $\delta^i = 0.01$, $b^i = 80$, $\rho = 0.3$, $T = 1$, $\omega^i = \frac{1}{I}$ for $i = 1, \dots, I/2$, $\omega^i = -\frac{1}{I}$ for $i = I/2 + 1, \dots, I$ and $J = 12$, $n = 50000$. The CPU time for the crude Monte Carlo approach is 0.76. The CPU times for the reduced importance sampling approach are $(1.42, 1.44, 1.52)$ and for the full importance sampling approach they are $(1.53, 2.41, 3.05)$.

Our last two examples deal multi-dimensional barrier options with discrete monitoring. The first striking result to notice when looking at Tables 6 and 7 concerns the huge CPU times of the full approaches which nonetheless do not significantly improve the variance reduction compared to the reduced size methods. This remark definitely advocates the use of reduced size approaches. The variance is always divided by a factor between 2 and 3, whereas the CPU time is only twice the one of the crude Monte Carlo approach. In the Merton model case (Table 6), the *Gaussian+Poisson* approach always provides the largest variance reduction for a computational time very close to the two other methods, meanwhile in the BNS model (Table 7) the *Poisson* and *Gaussian+Poisson* methods perform similarly. The efficiency of the pure *Poisson* approach comes from the particular form of the BNS model which includes jumps in the volatility process. These jumps seem to have a larger impact on the overall variance that the Brownian motion itself.

	Strike	Price	Var	VarG	VarP	VarGP
Full	100	2.97	36	37.8	16	16
Reduced		2.97	36	36.2	16	16
Full	90	12.52	36	36.4	14.5	14.5
Reduced		12.52	36	36	14.3	14.3
Full	110	1.64	12.1	13.3	6.1	5.5
Reduced		0.80	12.1	12	5.3	5.4

Table 7: Barrier option in dimension $I = 5$ in the BNS model with $S_0^i = 100$, $r = 0.05$, $\lambda^i = 0.01$, $\mu^i = 1$, $\kappa^i = 0.54$, $\beta^i = 18.6$, $b^i = 70$, $\rho = 0.2$, $T = 1$, $\omega^i = \frac{1}{I}$ for $i = 1, \ldots, I$ and $J = 12$, $n = 50000$. The CPU time for the crude Monte Carlo approach is 0.52. The CPU times for the reduced importance sampling approach are (1.06, 1.1, 1.17) and for the full importance sampling approach they are (2.1, 3.8, 10.5).

5 Conclusion

In this work, we have studied an importance sampling based Monte Carlo method for jump processes. The proposed algorithm splits into two parts. First, we compute the optimal change of measure using Newton's algorithm on a sample average approximation of the stochastic optimization problem. This method is very robust and does not require any fine tuning unlike stochastic approximation methods. Second, we use this estimator of the optimal measure change in an independent Monte Carlo method. We have established a several convergence results for this approach and in particular we have proved that it satisfies a Central Limit Theorem with optimal limiting variance. All the numerical examples we have investigated advocates the use of reduced size problems to significantly speed up the computations since the loss of variance reduction compared to the full approach remains negligible. This importance sampling approach proves to be all the more efficient as the jump and the diffusion parts are mixed up.

References

B. Arouna. Adaptative Monte Carlo method, a variance reduction technique. *Monte Carlo Methods Appl.*, 10(1):1–24, 2004. 2

B. Arouna. Robbins Monro algorithms and variance reduction in finance. *J. of Computational Finance*, 7(2), Winter 2003/04. 2

O. Barndorff-Nielsen and N. Shephard. Modelling by lévy processess for financial econometrics. In O. Barndorff-Nielsen, S. Resnick, and T. Mikosch, editors, *Lévy Processes*, pages 283–318. Birkhäuser Boston, 2001a. 16

O. E. Barndorff-Nielsen and N. Shephard. Non-gaussian ornstein–uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):167–241, 2001b. 16

O. E. Barndorff-Nielsen and R. Stelzer. The multivariate supou stochastic volatility model. *Mathematical Finance*, 23(2):275–296, 2013. doi: 10.1111/j.1467-9965.2011.00494.x. 17

P. Carr, D. B. Madan, and R. H. Smith. Option valuation using the fast fourier transform. *Journal of Computational Finance*, 2:61–73, 1999. 1

H. Chen and Y. Zhu. Stochastic approximation procedure with randomly varying truncations. *Scientia Sinica*, 29(9):914–926, 1986. 6

B. Jourdain and J. Lelong. Robust Adaptive Importance Sampling for Normal Random Vectors. *Ann. Appl. Probab.*, 19(5):1687–1718, 2009. 2, 6, 8, 17

R. Kawai. Adaptive Monte Carlo variance reduction with two-time-scale stochastic approximation. *Monte Carlo Methods Appl.*, 13(3):197–217, 2007. 2

R. Kawai. Adaptive Monte Carlo variance reduction for Lévy processes with two-time-scale stochastic approximation. *Methodol. Comput. Appl. Probab.*, 10(2):199–223, 2008a. 2

R. Kawai. Optimal importance sampling parameter search for Lévy processes via stochastic approximation. *SIAM J. Numer. Anal.*, 47(1):293–307, 2008b. 2

J. Kiessling and R. Tempone. Diffusion approximation of Lévy processes with a view towards finance. *Monte Carlo Methods Appl.*, 17(1):11–45, 2011. 1

S. G. Kou. A jump-diffusion model for option pricing. *Manage. Sci.*, 48(8):1086–1101, 2002. 16

B. Lapeyre and J. Lelong. A framework for adaptive Monte–Carlo procedures. *Monte Carlo Methods Appl.*, 17(1), 2011. 2, 6

M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991. 7

J. Lelong. Almost sure convergence of randomly truncated stochastic algorithms under verifiable conditions. *Statistics & Probability Letters*, 78(16), 2008. 6

J. Lelong. Asymptotic normality of randomly truncated stochastic algorithms. *ESAIM. Probability and Statistics*, 2011. 6

V. Lemaire and G. Pagès. Unconstrained Recursive Importance Sampling. *Ann. Appl. Probab.*, 20(3): 1029–1067, 2010. 2

R. Y. Rubinstein and A. Shapiro. *Discrete event systems*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Ltd., Chichester, 1993. Sensitivity analysis and stochastic optimization by the score function method. 7, 8