Robust Adaptive Importance Sampling for Normal Random Vectors

Corvisier Jean-Christophe, OREISTEIN Pierre

January 10, 2019

Introduction

- **Article**

Benjamin Jourdain et Jérôme Lelong.
Robust Adaptive Importance Sampling for Normal Random Vectors, 2008.

- **Contexte**

Calcul de $\mathbb{E}[f(G)]$ avec $G = (G_1, \ldots, G_d)$ un vecteur normal de dimension d et $f : \mathbb{R}^d \longrightarrow \mathbb{R}$ une fonction mesurable et intégrable pour G.

- **Motivations**

Utile en mathématiques financières: correspond au le calcul du prix d'une option européenne dans le modèle de Black-Scholes avec f le payoff.

- **Proposition du papier**

Proposition d'une méthode de Monte-Carlo adaptatif robuste, efficace et automatique.

Hypothèses & Premiers Résultats

- **Hypothèses :**
  - La variable aléatoire non nulle: $\mathbb{P}(f(G) \neq 0) > 0$
  - La variable aléatoire est légèrement plus que de carré intégrable:

$$\forall\, \theta \in \mathbb{R}^d,\ \mathbb{E}[f^2(G) e^{-\theta \cdot G}] < +\infty \tag{1}$$

Hypothèses & Premiers Résultats

- **Hypothèses :**
  - La variable aléatoire non nulle: $\mathbb{P}(f(G) \neq 0) > 0$
  - La variable aléatoire est légèrement plus que de carré intégrable:

$$\forall\ \theta \in \mathbb{R}^d,\ \mathbb{E}[f^2(G)e^{-\theta \cdot G}] < +\infty \tag{1}$$

- **Transformation :**
  - Pour toute fonction intégrable, on a:

$$\forall\ \theta \in \mathbb{R}^d,\ \mathbb{E}[h(G)] = \mathbb{E}\left[h(G+\theta)e^{-\theta \cdot G - \frac{|\theta|^2}{2}}\right] \tag{2}$$

Hypothèses & Premiers Résultats

- **Hypothèses :**
  - La variable aléatoire non nulle: $\mathbb{P}(f(G) \neq 0) > 0$
  - La variable aléatoire est légèrement plus que de carré intégrable:

$$\forall\ \theta \in \mathbb{R}^d,\ \mathbb{E}[f^2(G)e^{-\theta \cdot G}] < +\infty \tag{1}$$

- **Transformation :**
  - Pour toute fonction intégrable, on a:

$$\forall\ \theta \in \mathbb{R}^d,\ \mathbb{E}[h(G)] = \mathbb{E}\left[h(G+\theta)e^{-\theta \cdot G - \frac{|\theta|^2}{2}}\right] \tag{2}$$

- **Premiers Résultats :**
  - $\mathbb{E}[f(G)] = \mathbb{E}\left[f(G+\theta)e^{-\theta \cdot G - \frac{|\theta|^2}{2}}\right] \Longrightarrow$ Nouvel estimateur

Hypothèses & Premiers Résultats

- **Hypothèses :**
  - La variable aléatoire non nulle: $\mathbb{P}(f(G) \neq 0) > 0$
  - La variable aléatoire est légèrement plus que de carré intégrable:

$$\forall\, \theta \in \mathbb{R}^d,\ \mathbb{E}[f^2(G)e^{-\theta\cdot G}] < +\infty \tag{1}$$

- **Transformation :**
  - Pour toute fonction intégrable, on a:

$$\forall\, \theta \in \mathbb{R}^d,\ \mathbb{E}[h(G)] = \mathbb{E}\left[h(G+\theta)e^{-\theta\cdot G - \frac{|\theta|^2}{2}}\right] \tag{2}$$

- **Premiers Résultats :**
  - $\mathbb{E}[f(G)] = \mathbb{E}\left[f(G+\theta)e^{-\theta\cdot G - \frac{|\theta|^2}{2}}\right] \Longrightarrow$ Nouvel estimateur
  - $\mathrm{Var}[f(G+\theta)e^{-\theta\cdot G + \frac{|\theta|^2}{2}}] = \underbrace{\mathbb{E}\left[f^2(G)e^{-\theta\cdot G - \frac{|\theta|^2}{2}}\right]}_{=\,\nu^f(\theta)} - \mathbb{E}[f(G)]^2$

Estimateurs

- **Estimateurs :**
  - $(G_i)_{i \leq 1}$ une séquence i.i.d. de gaussiennes centrées de dimension d.
  - Nouvel estimateur non biaisé et convergent de $\mathbb{E}[f(G)]$:

$$M_n(\theta, f) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} f(G_i + \theta) e^{-\theta \cdot G_i - \frac{|\theta|^2}{2}}$$

  - Variance de l'estimateur:

$$n \operatorname{Var}[M_n(\theta, f)] = \nu^f(\theta) \; - \; \mathbb{E}[f(G)]^2$$

Estimateurs

- **Estimateurs :**
  - $(G_i)_{i\leq 1}$ une séquence i.i.d. de gaussiennes centrées de dimension d.
  - Nouvel estimateur non biaisé et convergent de $\mathbb{E}[f(G)]$:

$$M_n(\theta, f) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} f(G_i+\theta)e^{-\theta\cdot G_i - \frac{|\theta|^2}{2}}$$

  - Variance de l'estimateur:

$$n\ \mathrm{Var}[M_n(\theta, f)] = \nu^f(\theta)\ -\ \mathbb{E}[f(G)]^2$$

- **Objectif :**
  - On cherche θ_*^f qui minimise la variance. *Ie*, trouver θ_*^f solution de :

$$\nabla_\theta \nu^f(\theta)\ =\ \mathbb{E}\left[(\theta - G)f(G+\theta)e^{-\theta\cdot G + \frac{|\theta|^2}{2}}\right]\ =\ 0$$

État de l'art

- **État de l'art :**
  - *[Arouna, 2004]* utilise l'algorithme de Robbins-Monro tronqué pour approximer θ_*^f.

    Problèmes : Difficulté dans le choix de la séquence de sous espaces compacts pour les variables aléatoires tronquées.

État de l'art

- **État de l'art :**
  - *[Arouna, 2004]* utilise l'algorithme de Robbins-Monro tronqué pour approximer θ_*^f.

    Problèmes : Difficulté dans le choix de la séquence de sous espaces compacts pour les variables aléatoires tronquées.

  - *[Lemaire and Pagès, 2008]* utilisation l'algorithme de Robbins-Monro pour caractériser θ_*^f comme solution de $\nabla_\theta \nu^f(\theta) = e^{|\theta|^2}\mathbb{E}\left[(2\theta - G)f^2(G-\theta)\right] = 0.$

    Problèmes : f doit satisfaire une hypothèse de croissance exponentielle à l'inifini.

    $\Longrightarrow$ Dans les deux cas nécessité d'optimiser la séquence des gains de l'algorithme de Robbins-Monro

État de l'art

- **État de l'art :**
  - *[Arouna, 2004]* utilise l'algorithme de Robbins-Monro tronqué pour approximer θ_*^f.

    Problèmes : Difficulté dans le choix de la séquence de sous espaces compacts pour les variables aléatoires tronquées.

  - *[Lemaire and Pagès, 2008]* utilisation l'algorithme de Robbins-Monro pour caractériser θ_*^f comme solution de $\nabla_\theta \nu^f(\theta) \; = \; e^{|\theta|^2}\mathbb{E}\left[(2\theta - G)f^2(G-\theta)\right] \; = \; 0.$

    Problèmes : f doit satisfaire une hypothèse de croissance exponentielle à l'inifini.

    $\Longrightarrow$ Dans les deux cas nécessité d'optimiser la séquence des gains de l'algorithme de Robbins-Monro

  - *[Kim and Henderson, 2007]* En deux étapes: 1) Newton sur la variance empirique pour estimer θ_*^f. 2) Estime $\mathbb{E}[f(G)]$ avec de nouveaux samples.

    Problème : Utilise des samples de G différents aux deux étapes.

Méthode

- **Estimation de θ_*^f:**

Approximation par θ_n^f, via procédure de Newton, minimisant :

$$\nu_n^f(\theta) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} f^2(G_i)\ e^{-\theta\cdot G_i+\frac{|\theta|^2}{2}}$$

Méthode

- **Estimation de θ_*^f:**

Approximation par θ_n^f, via procédure de Newton, minimisant :

$$\nu_n^f(\theta) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} f^2(G_i)\ e^{-\theta\cdot G_i+\frac{|\theta|^2}{2}}$$

- **Estimation de $\mathbb{E}[f(G)]$:**

Approximation par $M_n(\theta_n^f, f)$

$\Longrightarrow$ Utilisation des mêmes samples G_i pour ces deux approximations

Méthode

- **Estimation de θ_*^f:**

Approximation par θ_n^f, via procédure de Newton, minimisant :

$$\nu_n^f(\theta) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} f^2(G_i)\ e^{-\theta \cdot G_i + \frac{|\theta|^2}{2}}$$

- **Estimation de $\mathbb{E}[f(G)]$:**

Approximation par $M_n(\theta_n^f, f)$

$\Longrightarrow$ Utilisation des mêmes samples G_i pour ces deux approximations

- **Principales Contributions :**
  - Convergence p.s. de θ_n^f vers θ_*^f et de $\nu_n^f(\theta_n^f)$ vers $\nu^f(\theta_*^f)$
  - Convergence p.s. de $M_n(\theta_n^f, f)$ vers $\mathbb{E}[f(G)]$

Quelques détails sur les preuves

- **Preuve de la convergence de θ_n^f :**

1. Preuve que ν^f est $\mathcal{C}^\infty$ (via TCD)
2. Preuve que ν^f est fortement convexe. (Hypothèse 1)
3. Lemme *[Ledoux and Talagrand, 1991]* sur la convergence de la moyenne empirique de fonctions mesurables et intégrables.

Quelques détails sur les preuves

- **Preuve de la convergence de θ_n^f :**
  1. Preuve que ν^f est $\mathcal{C}^\infty$ (via TCD)
  2. Preuve que ν^f est fortement convexe. (Hypothèse 1)
  3. Lemme *[Ledoux and Talagrand, 1991]* sur la convergence de la moyenne empirique de fonctions mesurables et intégrables.

- **Preuve de la convergence de $M_n(\theta_n^f, f)$:**
  1. Convergence de θ_n^f vers θ^* (résultat précédent)
  2. Utilisation du Lemme *[Ledoux et Talagrand, 1991]* pour prouver la convergence local uniforme de $\theta \to M_n(f, \theta)$ vers $\mathbb{E}(f(G))$

Quelques détails sur les preuves

- **Preuve d'un TCL :**

1. Changement de variable : $\theta_n^f = \mathcal{A}\nu_n^f$ avec $\mathcal{A}$ une matrice de taille $dxd^{'}$ (et de rang $d^{'}$ avec $d^{'} \leq d$

2. Introduction d'un ensemble de fonctions

$$\mathcal{H}_\alpha = \Big\{g : \mathbb{R}^d \to \mathbb{R} \;\; s.t : \quad \exists \beta \in [0,2), \; \lambda \geq 0, \forall x, \; |g(x)| \leq \lambda * e^{|x|^\beta},$$
$$\forall x, y, \; |g(x) - g(y)| \leq \lambda e^{max(|x|^\beta, |y|^\beta)} |x-y|^\alpha \Big\}$$

3. On considère une décomposition $f = f_1 + f_2$ avec f_1 de classe $\mathcal{C}^1$ et f_2 dans l'ensemble $\mathcal{H}_\alpha$ avec $\alpha \in (\frac{\sqrt{{d'}^2 + 8d^2} - d^{'}}{4}, 1]$

4. Preuve de la convergence en probabilité des $\sqrt{n}(M_n(\theta_n^{f,\mathcal{A}}, f_i) - M_n(\theta_*^{f,\mathcal{A}}, f_i))$ vers 0.

5. Par théorème centrale limite, convergence en loi de $\sqrt{n}(M_n(\theta_n^{f,\mathcal{A}}, f_i) - \mathbb{E}(f(G))$ vers $\mathcal{N}_1(0, Var(f(G + \theta_*^{f,A})e^{-\theta_* . G - \frac{||\theta_*||^2}{2}}))$

Évaluation sur un modèle de Black-Scholes

- **Dynamique du modèle de Black-Scholes**

On considère la dynamique suivante :

$$dS_t^i = S_t^i(rdt + \sigma^i dW_t^i) \qquad S_0 = (S_0^1, \dots, S_0^d)$$

Où $W = (W^1, \dots, W^d)$ est un mouvement Brownien standard.

- **Matrice de Covariance entre les mouvements browniens**

Pour l'expérience on assume une matrice de covariance de la forme :

$$\langle W_t^i, W_t^j \rangle = \rho t \mathbb{1}_{\{i \neq j\}} + t \mathbb{1}_{\{i=j\}}$$

- **Fonction de payoff :**

$$f(G) = e^{-rT} \left(\sum_{i=1}^{d} \omega^i S_T^i(G) - K \right)_+$$

Résultats expérimentaux

- **Données :**

$d = 40, r = 0.05, T = 1, S_0^i = 50, \sigma^i = 0.2, \omega^i = 1/d, \forall i = 1, \ldots, d,$ et $N = 100.$

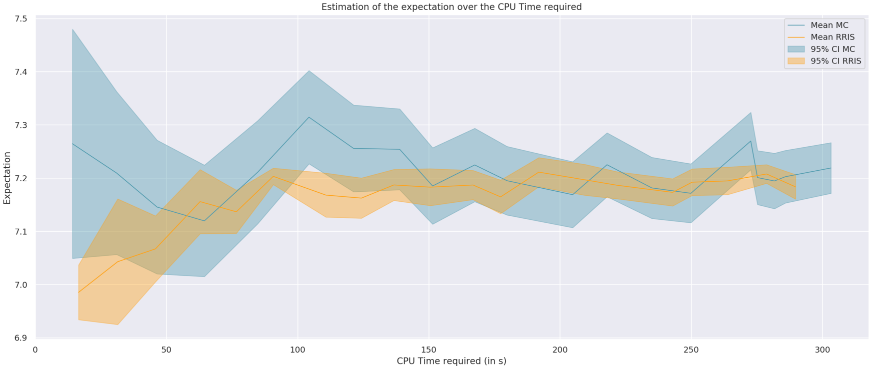

Figure: Approximation du prix et intervalle de confiance à 95% avec une méthode de Monte-Carlo classique (en bleu) et approximation via l'algorithme RRIS proposé (en orange)

Questions

Questions ...