

Algorithmes Stochastiques

Echantillonnage d'importance, changements de probabilités gaussiens

Corrigés mis en ligne au fur et à mesure à l'adresse <http://www.cmapx.polytechnique.fr/~benaych/M2AlgosStos/>

1. ECHANTILLONNAGE D'IMPORTANCE POUR DES V.A. DE BERNOULLI

Les v.a. sont définies sur $(\Omega, \mathcal{A}, \mathbb{P})$.

Qu. 1) Soient X_1, \dots, X_n des v.a. indépendantes de loi de Bernoulli $B(p_1), \dots, B(p_n)$; et Z_1, \dots, Z_n des v.a. indépendantes de lois de Bernoulli $B(q_1), \dots, B(q_n)$. Montrer que pour toute fonction $g : \{0, 1\}^n \rightarrow \mathbb{R}$ mesurable bornée, on a

$$\begin{aligned} \mathbb{E}[g(X_1, \dots, X_n)] &= \mathbb{E}\left[g(Z_1, \dots, Z_n) \prod_{i=1}^n \left(\mathbb{1}_{Z_i=0} \frac{1-p_i}{1-q_i} + \mathbb{1}_{Z_i=1} \frac{p_i}{q_i} \right)\right] \\ &= \left(\prod_{i=1}^n \frac{1-p_i}{1-q_i} \right) \mathbb{E}\left[g(Z_1, \dots, Z_n) \prod_{i=1}^n \left(\frac{p_i(1-q_i)}{q_i(1-p_i)} \right)^{Z_i}\right] \end{aligned}$$

Qu. 2) Soit $n \geq 1, p, q \in]0, 1[$; et S, S' deux v.a. respectivement de loi binomiale $\text{Binom}(n, p)$ et de loi binomiale $\text{Binom}(n, q)$. Dédurre de la question 1 que pour toute fonction $f : \{0, \dots, n\} \rightarrow \mathbb{R}$, on a

$$\mathbb{E}[f(S)] = \mathbb{E}\left[f(S') \left(\frac{p}{q}\right)^{S'} \left(\frac{1-p}{1-q}\right)^{n-S'}\right] = \left(\frac{1-p}{1-q}\right)^n \mathbb{E}\left[f(S') \left(\frac{p(1-q)}{q(1-p)}\right)^{S'}\right].$$

Qu. 3) Pour S de loi binomiale $\text{Binom}(n, p)$ et $x \in]0, p[$, estimer $\mathbb{P}(S \leq nx)$ des trois façons suivantes :
 — en utilisant la fonction de calcul exact de cette probabilité `scipy.stats.binom.cdf`,
 — par un Monte Carlo standard en simulant N copies de S avec la fonction `np.random.binomial`,
 — en utilisant la question 2 et en simulant N copies de S' avec la fonction `np.random.binomial`.
 On pourra choisir par exemple $n = 300, p = 0.25, x = 0.001, q = x$ et $N = 10^6$.

2. CHANGEMENT DE PROBABILITÉ DANS UN MODÈLE GAUSSIEN

Dans cette section, $X \sim \mathcal{N}(0, 1)$ sous \mathbb{P} . Nous cherchons à calculer $\mathbb{P}(X > a)$ puis $\mathbb{P}(|X| > a)$ pour des valeurs assez élevées de a de sorte que l'événement puisse être qualifié de rare.¹ Nous avons la relation, pour tout $a > 0$,

$$\frac{1}{a + 1/a} \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2}{2}} \leq \mathbb{P}(X > a) \leq \frac{1}{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2}{2}},$$

ce qui entraîne quand $a \rightarrow +\infty$,

$$\mathbb{P}(X > a) \sim \frac{1}{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{a^2}{2}}.$$

Ainsi, comme $e^{\frac{2.15^2}{2}} \approx 10$, $\mathbb{P}(X > a)$ et $\mathbb{P}(|X| > a)$ sont de l'ordre de $10^{-\left(\frac{a}{2.15}\right)^2}$ et des valeurs de a comprises dans l'intervalle $[5, 8]$ correspondent à notre champ d'investigation. Nous considérons ici un changement de probabilité pour améliorer le comportement de l'échantillonneur de Monte Carlo naïf

$$\mathbb{P}(X > a) \approx \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{X_k > a} \quad \text{où } \{X_k, k \geq 1\} \text{ i.i.d. } \mathcal{N}(0, 1).$$

1. Bien entendu, ces deux probabilités sont liées par la relation $\mathbb{P}(|X| > a) = 2\mathbb{P}(X > a)$.

2.1. Changement de probabilité par décentrage. Facilement, on montre que pour toute fonction $f : \mathbb{R} \rightarrow \mathbb{R}_+$ mesurable

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) \exp(\theta x - \theta^2/2) \exp(-x^2/2) dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) \exp(-(x - \theta)^2/2) dx;$$

càd que

$$\mathbb{E}_{\mathcal{N}(0,1)}[f(X)e^{\theta X - \theta^2/2}] = \mathbb{E}_{\mathcal{N}(\theta,1)}[f(X)]$$

et, en appliquant à la fonction $f(x)e^{-(\theta x - \theta^2/2)}$,

$$\mathbb{E}_{X \sim \mathcal{N}(0,1)}[f(X)] = \mathbb{E}_{Y \sim \mathcal{N}(\theta,1)}[f(Y)e^{-\theta Y + \theta^2/2}].$$

On en déduit l'estimateur suivant :

$$(1) \quad \mathbb{P}(X > a) \approx \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{Y_i > a} e^{-\theta Y_i + \theta^2/2}$$

pour Y_1, \dots, Y_n des vaidd de loi $\mathcal{N}(\theta, 1)$.

Application à l'estimation de $\mathbb{P}(X > a)$.

Qu. 4) Simuler $n \gg 1$ gaussiennes indépendantes de loi $\mathcal{N}(0, 1)$ avec lesquelles on estimera $\mathbb{P}(X > a)$ d'abord par l'estimateur naïf de Monte Carlo, puis par l'estimateur (1) appliqué avec $\theta = a$. Pour chacun des estimateurs, utiliser le TCL pour construire un intervalle de confiance asymptotique de probabilité de couverture 95%. Comparer à la valeur exacte de $\mathbb{P}(X > a)$.

Qu. 5) La longueur de l'intervalle de confiance est une fonction croissante de la variance de l'estimateur : on veut donc choisir θ qui minimise cette variance. Nous allons montrer que le choix de θ fait en question 4 pourrait être amélioré si le critère d'optimalité souhaité est de minimiser la variance du nouvel estimateur.

Montrer que choisir θ qui minimise la variance de l'estimateur donné par (1), revient à choisir θ qui minimise

$$\theta \mapsto \mathbb{E} [\mathbb{1}_{X > a} \exp(0.5\theta^2 - \theta X)].$$

Montrer que cette fonction est convexe et qu'elle possède un unique minimum donné par la solution de²

$$\mathbb{E} [\mathbb{1}_{X > a} (\theta - X) \exp(-\theta X)] = 0.$$

En déduire que prendre $\theta = a$ est meilleur que prendre $\theta = 0$; et que la valeur de θ la plus efficace est en fait supérieure à a .

3. ALGORITHME ADAPTATIF DE MOYENNE OPTIMALE DANS LE CAS DE CHANGEMENT DE PROBABILITÉ GAUSSIEN

Dans cette section, sous \mathbb{P} , X est un vecteur gaussien de \mathbb{R}^d standard ($X \sim \mathcal{N}_d(0, I)$). $\langle a, b \rangle$ désigne le produit scalaire usuel sur \mathbb{R}^d et $\|\cdot\|$ la norme euclidienne associée. Pour un vecteur a , a' est la transposée. Par convention, les vecteurs sont des vecteurs colonnes.

On cherche à calculer $\mathbb{E}[f(X)]$ pour une fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable telle que pour tout $\theta \in \mathbb{R}^d$,

$$0 < \mathbb{E}[f^2(X) e^{-\langle \theta, X \rangle}] < \infty.$$

Pour ce faire, on peut implémenter un estimateur de Monte Carlo naïf ou exploiter la relation

$$(2) \quad \mathbb{E}[h(X)] = \mathbb{E}[h(X + \theta) e^{-\langle \theta, X \rangle - \|\theta\|^2/2}],$$

valable pour tout $\theta \in \mathbb{R}^d$, et toute fonction $h : \mathbb{R}^d \rightarrow \mathbb{R}$ mesurable positive. L'objectif de cette section est d'apprendre, par une procédure adaptative, la valeur optimale du paramètre θ .

2. on admettra que le modèle est régulier i.e. que l'on permuter dérivée et intégration.

► *Algorithme de Lelong et Jourdain* [2]. On déduit de (2) que pour tout $\theta \in \mathbb{R}^d$, (i) la v.a.

$$(3) \quad M_n(\theta) := \frac{1}{n} \sum_{k=1}^n f(X_k + \theta) e^{-\langle \theta, X_k \rangle - \|\theta\|^2/2}, \quad \{X_k, k \geq 0\} \text{ v.a. i.i.d. de loi } \mathcal{N}(0, 1)$$

est un estimateur sans biais de l'objectif $\mathbb{E}[f(X)]$; et (ii) sa variance est égale à $n^{-1} \left(\mathcal{V}(\theta) - (\mathbb{E}[f(X)])^2 \right)$, où

$$\mathcal{V}(\theta) := \mathbb{E}[f^2(X) e^{-\langle \theta, X \rangle + \|\theta\|^2/2}].$$

On souhaite donc appliquer l'estimateur (3) avec $\theta = \theta_*$ où θ_* minimise $\mathcal{V}(\theta)$. Cependant, en général, lorsque $\mathbb{E}[f(X)]$ est inconnu, il en est de même pour $\mathcal{V}(\theta)$ et donc θ_* est défini comme un minimiseur d'une fonction incalculable. Or, sous des conditions d'intégrabilité sur f , la fonction \mathcal{V} est de classe C^2 , strictement convexe et possédant un unique minimum θ_* ; son gradient G et son Hessien H sont donnés par

$$\begin{aligned} G(\theta) &:= \mathbb{E} \left[f^2(X) (\theta - X) e^{-\langle \theta, X \rangle + \|\theta\|^2/2} \right] \in \mathbb{R}^{d \times 1}, \\ H(\theta) &:= \mathbb{E} \left[f^2(X) (I + (\theta - X)(\theta - X)') e^{-\langle \theta, X \rangle + \|\theta\|^2/2} \right] \in \mathbb{R}^{d \times d}; \end{aligned}$$

I désigne la matrice identité de taille $d \times d$. Afin d'approcher θ_* , on va utiliser une version bruitée de la *méthode de Newton* : pour n fixé, θ_* sera approché par θ_n défini comme la limite lorsque k tend vers l'infini, de la suite $\{t_k, k \geq 0\}$ satisfaisant à

$$(4) \quad t_{k+1} = t_k - (H_n(t_k))^{-1} G_n(t_k)$$

où $H_n(t), G_n(t)$ sont les approximations Monte Carlo de $H(t)$ et $G(t)$ calculées à partir des mêmes n réalisations indépendantes X_1, \dots, X_n de v.a. $\mathcal{N}_d(0, I)$:

$$(5) \quad G_n(\theta) := \frac{1}{n} \sum_{k=1}^n f^2(X_k) (\theta - X_k) e^{-\langle \theta, X_k \rangle + \|\theta\|^2/2},$$

$$(6) \quad H_n(\theta) := \frac{1}{n} \sum_{k=1}^n f^2(X_k) [I + (\theta - X_k)(\theta - X_k)'] e^{-\langle \theta, X_k \rangle + \|\theta\|^2/2}.$$

En pratique, on se fixe un seuil ε et on itère la relation (4) tant que $\|G_n(t_k)\| > \varepsilon$. On obtient alors θ_n ; on prendra pour estimateur de $\mathbb{E}[f(X)]$ la quantité $M_n(\theta_n)$ calculée à partir des mêmes tirages X_1, \dots, X_n que ceux utilisés pour le calcul de G_n et H_n .

Les résultats suivants, qui peuvent être démontrés sous hypothèses d'intégrabilité sur la fonction f , assurent que la méthode fonctionne (voir [2]) :

- (i) $M_n(\theta_n)$ tend presque sûrement vers $\mathbb{E}[f(X)]$ lorsque $n \rightarrow \infty$,
- (ii) si $\text{Var}(f(X)) > 0$, la suite de v.a.

$$\sqrt{\frac{n}{\mathcal{V}(\theta_n) - M_n(\theta_n)^2}} (M_n(\theta_n) - \mathbb{E}[f(X)])$$

converge en loi vers une loi gaussienne standard,

- (iii) $\sqrt{n}(\theta_n - \theta_*)$ est asymptotiquement gaussienne centrée (sa variance peut être exprimée, de façon relativement complexe, en fonction des paramètres).

► *Mise en oeuvre de la méthode*. On se limitera dans cette question au cas $d = 1$. On s'intéresse au cas particulier $f : x \mapsto (e^x - K)^+$, qui correspond au calcul du prix d'une option call en finance (pour $y \in \mathbb{R}$, on note $y^+ := \max(y, 0)$).

Il s'agit d'un exemple jouet puisque $\mathbb{E}[f(X)]$ a une expression explicite donnée par (cette formule n'est plus valable en dimension d supérieure à 1)

$$(7) \quad \mathbb{E}[f(X)] = e^{\frac{1}{2}} F(1 - \ln(K)) - K F(-\ln(K)), \quad \text{avec } F(x) := \mathbb{P}(X \leq x).$$

Qu. 6) Fixer $n_{\max} = 10^4$ et simuler $X_1, \dots, X_{n_{\max}}$ v.a. i.i.d. $\mathcal{N}(0, 1)$.

- (a) Pour différentes valeurs de $n \leq n_{\max}$, représenter la fonction

$$\mathcal{V}_n : \theta \mapsto \frac{1}{n} \sum_{k=1}^n f^2(X_k) \exp(-\langle \theta, X_k \rangle + \|\theta\|^2/2).$$

Tracer aussi la suite $\{t_k, 1 \leq k \leq \min\{\tau_\varepsilon, 100\}\}$, où τ_ε est le premier k tel que $\|G_n(t_k)\| \leq \varepsilon$. Comment cette suite converge-t-elle ?

On pourra prendre $K = 1$, se limiter à tracer \mathcal{V}_n sur l'intervalle $[-1, 4]$ et prendre $\varepsilon = 0.05$. Noter que lorsque $d = 1$,

$$\theta - \frac{G_n(\theta)}{H_n(\theta)} = \frac{\sum_{k=1}^n f^2(X_k) [\theta^3 + (1 - 2\theta^2)X_k + \theta X_k^2] e^{-\theta X_k}}{\sum_{k=1}^n f^2(X_k) [1 + (\theta - X_k)^2] e^{-\theta X_k}}.$$

- (b) Tracer sur un même graphique les suites $\{M_n(0), n \geq 0\}$ et $\{M_n(\theta_n), n \geq 0\}$ pour n variant de 100 à n_{\max} par sauts de 50 ou de 100 (de façon à accélérer l'algorithme).

Pour comparaison, tracer aussi la droite horizontale $y = \mathbb{E}[f(X)]$ où cette constante est calculée avec la formule (7)

- (c) Pour différentes valeurs de n , simuler un grand nombre de réalisations de $M_n(0)$ et de $M_n(\theta_n)$, estimer les variances de ces deux v.a., et tracer l'histogramme de ces réalisations, ainsi que la droite verticale $x = \mathbb{E}[f(X)]$.

Que nous disent ces histogrammes sur l'efficacité des deux méthodes ?

- Qu. 7) Reprendre la question précédente en dimension d supérieure, pour la fonction $f : x = (x_1, \dots, x_d) \mapsto \left(\frac{1}{d} \sum_{i=1}^d e^{x_i} - K\right)^+$, qui correspond au calcul du prix d'un call sur moyenne. On affichera le tracé des suites $\{M_n(\theta_n), n \geq 0\}$ et $\{M_n(0), n \geq 0\}$ ainsi que les histogrammes.

RÉFÉRENCES

- [1] J. Lelong and B. Jourdain. Robust Adaptive Importance Sampling for Normal Random Vectors. *Ann. Appl. Probab.*, 19(5), 2009.