PROJET 8

Techniques de réduction de variance pour la méthode de Monte-Carlo

sujet proposé par G. Conforti

giovanni.conforti@polytechnique.edu

Dans ce projet on s'intéresse à trois méthodes pour accélérer la convergence de l'estimateur Monte Carlo classique: les variables antithétiques, l'échantillonage d'importance et les variables de contrôle.

1 Introduction

Soit $f : [0,1] \to \mathbb{R}$ une fonction de carré intégrable. On veut estimer $I = \int_0^1 f(x)\,dx$ par une méthode de Monte Carlo. Dans la suite, U désigne une variable aléatoire de loi uniforme sur $[0,1]$.

(S) Pour chaque fonction $f(x) = x^2, \exp(x), \cos^2(2\pi x)$, considérer l'estimateur de Monte-Carlo

$$J_n = \frac{1}{n}\sum_{i=1}^{n} f(U_i),$$

où $(U_i)_{i\geq 1}$ est un suite de variables aléatoires i.i.d. de loi uniforme sur $[0,1]$. Déterminer n t.q.

$$\mathbb{P}[|J_{2n} - I| \geq \varepsilon] \leq 0.95. \tag{8.1}$$

avec $\varepsilon = 0.1, 0.05$. Vérifier empiriquement la convergence de l'estimateur et la relation (8.1) en lançant l'algorithme de simulation un grand nombre de fois.

2 Variables antithétiques

(T) Montrer que

$$I = \mathbb{E}\left(\frac{1}{2}(f(U) + f(1-U))\right).$$

Montrer que l'estimateur

$$I_{2n} = \frac{1}{2n}\sum_{i=1}^{n}(f(U_i) + f(1-U_i))$$

converge presque sûrement vers I. Rappeler comment on peut obtenir une estimation de l'erreur $|I_{2n} - I|$.

(T) On veut comparer la variance de l'estimateur I_{2n} avec la variance de l'estimateur de Monte-Carlo classique

$$J_{2n} = \frac{1}{2n} \sum_{i=1}^{2n} f(U_i). \tag{8.2}$$

Montrer que $\mathrm{Var}(I_{2n}) \leq \mathrm{Var}(J_{2n})$ si et seulement si $\mathrm{Cov}(f(U), f(1-U)) \leq 0$.

(T) On suppose que la fonction f est monotone. Montrer que

$$\mathbb{E}[(f(U_1) - f(U_2))(f(1-U_1) - f(1-U_2))] \leq 0.$$

En déduire que dans ce cas, on a bien $\mathrm{Cov}(f(U), f(1-U)) \leq 0$. Conclure sur l'intérêt pratique de cette méthode.

(S) Pour $f(x) = x^2, \exp(x), \cos^2(2\pi x)$, comparer les estimateur J_{2n} et I_{2n} à n fixe. On pourra prendre $n = 10, 30, 50$. Quelles sont les fonctions pour lesquelles I_{2n} se comporte mieux que J_{2n}?

3 Échantillonage d'importance

On s'intéresse ici à une deuxième méthode pour améliorer la convergence de l'estimateur Monte-Carlo classique. Soit $(Y_i)_{i\geq 1}$ une suite de variables aléatoires i.i.d. à valeurs dans $[0,1]$ et de densité q, tel que $q(x) > 0$ pour tout $x \in [0,1]$. On note

$$S_n = \frac{1}{n} \sum_{i=1}^{n} (\frac{f}{q})(Y_i) \tag{8.3}$$

(T) Que vaut $\mathbb{E}[S_n]$? Donner une expression pour $\mathrm{Var}[S_n]$ en fonction de n et

$$v_1(q) = \int_0^1 \frac{f^2}{q}(x)dx \; - I^2.$$

Donner un intervalle de confiance asymptotique à 95% approchant I, en fonction de S_n, $v_1(q)$ et n.

(S) On considère pour $\lambda \in \mathbb{R}$ la densité

$$q_\lambda(x) = \lambda \frac{\exp(\lambda x) - 1}{\exp(\lambda) - 1} 1_{[0,1]}(x).$$

Pour les fonctions $f(x) = x^2$, $\exp(x)$, choisir la valeur optimale de λ pour l'estimateur (8.3) et implémenter la méthode d'échantillonage d'importance pour $n = 10,\ 30,\ 50$. Vérifier empiriquement la convergence.

(S) De nouveau pour les fonctions $f(x) = x^2$, $\exp(x)$, comparer empiriquement, à n fixé, l'estimateur S_{2n} avec l'estimateur I_{2n} (de la section précédente). Pour cela, on pourra considérer des copies i.i.d. $(I_{2n}^k)_{k\leq M}$ et $(S_{2n}^k)_{k\leq M}$ de I_{2n} et S_{2n} et comparer les variances empiriques

$$\frac{1}{M}\sum_{k=1}^{M} (I_{2n}^k)^2 - \Big(\frac{1}{M}\sum_{k=1}^{M} I_{2n}^k\Big)^2$$
$$\frac{1}{M}\sum_{k=1}^{M} (S_{2n}^k)^2 - \Big(\frac{1}{M}\sum_{k=1}^{M} S_{2n}^k\Big)^2$$

Quel estimateur semble le meilleur ?

4 Variables de contrôle

On considère une variable aléatoire réelle de carré intégrable Y telle que $\mathbb{E}[Y] = 0$ et on considère l'estimateur, pour un réel α à determiner,

$$K_n^\alpha = \frac{1}{n} \sum_{i=1}^{n} (f(U_i) - \alpha Y_i)$$

où $(U_i, Y_i)_{i\geq 1}$ désignent des réalisations indépendantes du couple (U, Y).

(T) Quel est le comportement de K_n^α dans la limite $n \to \infty$? Calculer $\mathbb{E}[K_n^\alpha]$ et $\mathrm{Var}(K_n^\alpha)$. Montrer que la variance est minimisée pour

$$\alpha^\star = \frac{\mathrm{Cov}(f(U), Y)}{\mathrm{Var}(Y)}.$$

(On rappelle que $\mathrm{Cov}(f(U), Y) = \mathbb{E}[f(U)Y] - \mathbb{E}[f(U)]\mathbb{E}[Y]$.).

(S) Pour $Y \overset{\text{loi}}{=} U - \frac{1}{2}$ et les fonctions $f(x) = x^2, \exp(x)$, calculer explicitement la constante $\alpha^\star$ et construire l'estimateur $K_n^{\alpha^\star}$. Vérifier empiriquement la convergence de l'estimateur $K_n^{\alpha^\star}$ et comparer avec la méthode des variables anthithétiques.

(S) En pratique, $\alpha^\star$ n'est pas connu. On se propose donc de l'estimer par une méthode de Monte-Carlo et on considére donc

$$\tilde{K}_n = \frac{1}{n}\sum_{i=1}^{n}(f(U_i) - \alpha_n Y_i)$$

où

$$\alpha_n = \frac{\frac{1}{n}\sum_{i=1}^n f(U_i)Y_i - \left(\frac{1}{n}\sum_{i=1}^n f(U_i)\right)\left(\frac{1}{n}\sum_{i=1}^n Y_i\right)}{\frac{1}{n}\sum_{i=1}^n (Y_i)^2 - \left(\frac{1}{n}\sum_{i=1}^n Y_i\right)^2}.$$

Que dire de la limite de $\tilde{K}_n$ quand $n \to \infty$? Pour le même choix des fonctions qu'au point précédant et $n = 10, 25$, comparer empiriquement les estimateurs $K_n^{\alpha^\star}, \tilde{K}_n$. Quel estimateur se comporte le mieux? On utilisera toujours $Y \overset{\text{loi}}{=} U - \frac{1}{2}$ comme variable de contrôle.

(S) **Facultatif.** Proposer une méthode pour évaluer numériquement $\mathrm{Var}(\tilde{K}_n)$. Pour $n = 10, 25$ et $f(x) = x^2, \exp(x)$, vérifier empiriquement que $\mathrm{Var}(\tilde{K}_n) \geq \mathrm{Var}(K_n^{\alpha^\star})$.
