

Méthodes de Monte Carlo

Thierry Jeantheau

1 Méthode générale

Commençons par rappeler la loi des grands nombres. Soient X_1, X_2, \dots, X_n une suite de vecteurs aléatoires indépendants de même loi telle que $E(|X_1|) < \infty$, on a

$$S_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{p.s.}} E(X_1).$$

On voit qu'une quantité aléatoire (S_n) converge vers une quantité déterministe ($E(X_1)$). Si cette quantité déterministe n'est pas calculable explicitement, la simulation et la loi des grands nombres donne donc un moyen de l'approcher numériquement (avec n assez grand). Ces méthodes qui utilisent la simulation pour calculer une valeur numérique s'appelle "méthode de Monte Carlo", en référence à ce haut lieu du jeu de hasard.

Voyons immédiatement d'autres quantités que l'on peut estimer à l'aide de cette méthode.

1. **Calcul d'une probabilité.** On suppose que X est une variable aléatoire simulable, et on veut estimer $P(X \in B)$ (B sous ensemble du support de X . On considère alors

$$S_n = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \in B},$$

o u les X_i sont simulés suivant la même loi que X .

2. **Intégrale d'une fonction à support compact.** Soit f une fonction de $K \subset \mathbb{R}^d$ dans \mathbb{R} , où K est un compact de volume $\text{Vol}(K)$ connu. Pour estimer $\int_K f(t) dt$, il faut simuler, par exemple par une méthode de rejet, des variables aléatoires U_i uniforme sur K , et considérer

$$S_n = \frac{\text{Vol}(K)}{n} \sum_{i=1}^n f(U_i).$$

L'analyse numérique fournit d'autres méthodes pour estimer ce type de quantités (par exemple la méthode des trapèzes pour estimer une intégrale). Il est donc important d'avoir un indicateur de la précision de la méthode de Monte Carlo. On obtient celui-ci en utilisant le théorème central limite. Rappelons le dans le cas uni-dimensionnel. Si X_1, X_2, \dots, X_n est une suite de variables aléatoires indépendants de même loi de variance finie (notée σ^2), on a

$$\sqrt{n} \frac{S_n - E(X_i)}{\sigma} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

La vitesse de convergence est donc en $1/\sqrt{n}$. Si on intègre sur \mathbb{R} ($d = 1$), cette méthode de calcul n'est pas compétitive par rapport aux méthodes classiques, comme par exemple la méthode des trapèzes. Mais nous verrons qu'elle le devient en dimension plus grande.

2 Calcul d'intégrale

On cherche à calculer l'intégrale multiple

$$I = \int_{\mathbb{R}^d} f(x) dx.$$

On note $S \subset \mathbb{R}^d$ le support de f ($f(x) = 0 \Rightarrow x \notin S$), et soit $p(x)$ une densité de probabilité sur \mathbb{R}^d qui ne s'annule pas sur S . On peut écrire

$$I = \int_S \frac{f(x)}{p(x)} p(x) dx.$$

Cette intégrale est estimée par

$$S_n = \frac{1}{n} \sum_{i=1}^n \frac{f(X_i)}{p(X_i)},$$

où les X_i sont simulés suivant la loi de densité X . On pose

$$\sigma^2 = \text{var} \left(\frac{f(X_1)}{p(X_1)} \right)$$

A l'aide du théorème central limite, on a l'intervalle de confiance suivant à 95% pour I :

$$I \in \left[S_n - 2 \frac{\sigma}{\sqrt{n}}, S_n + 2 \frac{\sigma}{\sqrt{n}} \right]$$

Cette intervalle ne dépend pas de la dimension d du problème, cette méthode est donc particulièrement adaptée aux problèmes en grande dimension.

La densité $p(x)$ utilisée est largement arbitraire. La seule contrainte est que le support de p contienne le support de f . Pour améliorer l'estimation, on peut chercher la densité p qui rende σ le plus petit possible. Voici un résultat théorique qui permet de nous éclairer sur le choix de p :

Proposition 2.1 *La variance σ^2 de $f(X_1)/p(X_1)$ est minimale pour*

$$p(x) = \frac{|f(x)|}{\int_S |f(x)| dx}.$$

Démonstration. On a

$$\sigma^2 = \int \left(\frac{f(x)}{p(x)} \right)^2 p(x) dx - I^2$$

Il s'agit donc de rendre minimum $\int \frac{f^2(x)}{p(x)} dx$. D'après l'inégalité de Schwarz, nous avons, pour toute densité $p(x)$,

$$\left(\int |f(x)| dx \right)^2 = \left(\int \frac{|f(x)|}{\sqrt{p(x)}} \sqrt{p(x)} dx \right)^2 \leq \int \frac{f^2(x)}{p(x)} dx$$

Or, en posant $p(x) = \frac{|f(x)|}{\int |f(x)| dx}$, on obtient

$$\int \frac{f^2(x)}{p(x)} dx = \left(\int |f(x)| dx \right)^2.$$

Pour cette densité la borne minimale est atteinte, ce qui démontre la proposition. ■

Ce résultat n'est pas utilisable directement, car pour l'utiliser il faudrait connaître I et on tourne en rond... Il montre cependant que l'écart-type est d'autant plus petit que la densité $p(x)$ ressemble à la fonction $f(x)$. Il faut aussi choisir la densité $p(x)$ de manière à ce que la suite de variables aléatoires X_n soit facilement simulable.

3 Méthodes des variables "antithétiques"

Il existe différentes manières d'améliorer les méthodes de Monte Carlo. Nous en décrivons ici une. Pour expliquer la démarche, commençons par traiter un exemple. Soit f une fonction, on cherche à estimer

$$I = \int_0^1 f(x) dx.$$

On peut donc estimer I par Monte Carlo en simulant U_1, \dots, U_n suivant la loi $\mathcal{U}_{[0,1]}$, et en considérant

$$S_n = \frac{1}{n} \sum_{i=1}^n f(U_i).$$

Puisque U_1 et $1 - U_1$ ont même loi (ce sont des variables "antithétiques"), on a

$$I = \frac{1}{2} E[f(U_1) + f(1 - U_1)].$$

On en déduit une autre approximation de I , donnée par

$$\widehat{S}_{2n} = \frac{1}{2n} \sum_{i=1}^n f(U_i) + f(1 - U_i).$$

Comparons maintenant les deux méthodes, c'est à dire les variances de S_{2n} et \widehat{S}_{2n} . Evidemment, on a $\text{Var} S_{2n} = \text{Var} f(U_1)/2n$. D'autre part, on a

$$\text{Var} (f(U_1) + f(1 - U_1)) = 2 \left(\text{Var} (f(U_1)) + \text{Cov} (f(U_1), f(1 - U_1)) \right),$$

et

$$\text{Var} (\widehat{S}_{2n}) = \frac{1}{2n} \left(\text{Var} (f(U_1)) + \text{Cov} (f(U_1), f(1 - U_1)) \right).$$

La méthode présente donc un gain si $\text{Cov}(f(U_1), f(1 - U_1))$ est négative.

Généralisons maintenant le principe. Soient X_1, X_2, \dots, X_n une suite de vecteurs aléatoires de \mathbb{R}^d , et T une fonction de \mathbb{R}^d dans \mathbb{R}^d . On dit que les vecteurs X_1 et $T(X_1)$ sont antithétiques si ils ont la même loi. On pose

$$S_{2n} = \frac{1}{2n} \sum_{i=1}^{2n} f(X_i) \quad \text{et} \quad \widehat{S}_{2n} = \frac{1}{2n} \sum_{i=1}^n f(X_i) + f(T(X_i)).$$

Si $\text{Cov}(f(X_1), f(T(X_1)))$ est négative, alors on obtient une meilleure approximation en utilisant \widehat{S}_{2n} . En dimension 1, le résultat suivant donne des hypothèses générales d'application de ce résultat.

Proposition 3.1 *Si $f : \mathbb{R} \rightarrow \mathbb{R}$ est monotone, et si $T : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction décroissante, alors*

$$\text{Var} (\widehat{S}_{2n}) \leq \text{Var} (S_{2n}).$$

Démonstration. Il suffit de montrer que $\text{Cov}((f(X_1), f(T(X_1))) \leq 0$. On a

$$\text{Cov} ((f(X_1), f(T(X_1))) = E \left[\left(f(X_1) - E[f(X_1)] \right) \left(f(T(X_1)) - E[f(T(X_1))] \right) \right],$$

et, puisque $E[f(T(X_1))] = E[f(X_1)]$, il faut montrer que

$$E \left[f(X_1) f(T(X_1)) \right] \leq \left(E \left[f(X_1) \right] \right)^2. \quad (1)$$

Considérons maintenant la quantité

$$E \left[\left(f(X_1) - f(X_2) \right) \left(f(T(X_1)) - f(T(X_2)) \right) \right].$$

Comme T est décroissante et f est monotone, ce terme est toujours négatif. En développant et en utilisant l'indépendance de X_1 et X_2 , on obtient

$$E \left[f(X_1) f(T(X_1)) \right] + E \left[f(X_2) f(T(X_2)) \right] \leq E \left[f(X_1) \right] E \left[f(T(X_2)) \right] + E \left[f(X_2) \right] E \left[f(T(X_1)) \right].$$

L'identité en loi de X_1 et X_2 donne l'inégalité (1) cherchée. ■