Echantillonnage d'importance adaptatif en Population MonteCarlo

Proposé par Randal Douc : randal.douc@polytechnique.edu

Dans tout ce projet,

$\boxed{\Pi \text{ désigne une loi sur } \mathbb{R} \text{ et } \pi \text{ désigne sa densité par rapport à la mesure de Lebesgue.}}$

On s'interesse ici à approcher $\Pi(h) = \int_{\mathbb{R}} h(x)\pi(x)dx$. La méthode que nous décrirons pourrait s'appliquer en toute généralité sur $\mathbb{R}^p$ plutôt que sur $\mathbb{R}$ mais pour simplifier, nous traiterons uniquement le cas $p = 1$.

1 Echantillonnage d'importance et optimalité au sens de la variance asymptotique

1.1 Echantillonnage d'importance non normalisée

Soient $(X_i)_{1\leq i\leq n}$ une famille de variables aléatoires indépendantes et de même loi ayant g pour densité par rapport à la mesure de Lebesgue. On supposera que $\int_{\{x;g(x)=0\}} \pi(x)dx = 0$. Notons

$$\tilde{\Pi}_{g,n}(h) = \frac{1}{n}\sum_{i=1}^{n} h(X_i)\frac{\pi(X_i)}{g(X_i)} \tag{1}$$

On appellera g la **densité instrumentale.**

(T1) On suppose que $\int |h(x)|\pi(x)dx < \infty$, quel théorème peut-on utiliser pour prouver que $\tilde{\Pi}_{g,n}(h)$ converge avec probabilité 1 vers $\Pi(h)$?

L'approximation de $\Pi(h)$ par $\tilde{\Pi}_{g,n}(h)$ est couramment appelé *échantillonnage d'importance non normalisée*. Une façon de voir cette approximation est de considérer que la loi Π est approchée à l'aide d'une famille de $(X_i,\omega_i)_{1\leq i\leq n}$ avec $\omega_i = \frac{\pi(X_i)}{ng(X_i)}$ qui est donc une famille de points (X_i) pondérés par les poids (ω_i). La quantité $\Pi(h)$ est alors approchée par $\sum_{i=1}^n \omega_i h(X_i)$.

(T2) Quel hypothèse faut-il ajouter pour obtenir la propriété : pour tout intervalle I,

$$\mathbb{P}\left(\sqrt{n}\left(\tilde{\Pi}_{g,n}(h) - \Pi(h)\right)/\sqrt{\tilde{V}_g(h)} \in I\right) \to \int_I \frac{e^{-x^2/2}}{\sqrt{2\pi}}dx\ ,$$

avec $\tilde{V}_g(h) = \int \pi(x)h^2(x)\frac{\pi(x)}{g(x)}dx - [\Pi(h)]^2$.

On dit aussi que $\tilde{V}_g(h)$ est la variance asymptotique de $\sqrt{n}\left(\tilde{\Pi}_{g,n}(h) - \Pi(h)\right)$. Une bonne approximation de $\Pi(h)$ par $\tilde{\Pi}_{g,n}(h)$ est ainsi liée à une variance asymptotique faible. Cela pourra nous guider dans le choix d'une "bonne" densité instrumentale. Cela dit, *le meilleur choix de densité instrumentale n'est pas nécessairement* π. Nous allons le constater dans un exemple très simple (questions S1 et S2 ci-dessous).

(S1) On prend pour Π la loi uniforme sur $[0,1]$ et on pose $h(x)=x$. Nous allons d'abord illustrer graphiquement en fonction de n, l'approximation de $\Pi(h)$ par $\tilde{\Pi}_{g,n}(h)$ où on a posé $g=\pi$. Pour cela, tirer $m=2000$ variables aléatoires $(X_i)_{1\leq i\leq m}$ iid suivant une loi uniforme et tracer les valeurs prises par $\tilde{\Pi}_{g,n}(h)$ pour $n=1,2,\ldots,m$. On pourra aussi représenter les bornes inférieures et supérieures d'un intervalle de confiance à 95% de $\Pi(h)$ pour ces mêmes valeurs de n.

(S2) Reprendre la question précédente, en remplaçant la densité g par $g:[0,1]\mapsto\mathbb{R}^+$ telle que $g(x)=\frac{5}{2}x^{3/2}$ pour $x\in[0,1]$. Pour simuler une variable aléatoire X suivant cette densité g, on pourra utiliser (en la justifiant) la méthode suivante : tirer une variable aléatoire U suivant une loi uniforme sur $[0,1]$ et poser $X=U^{2/5}$. Vérifier expérimentalement qu'il vaut mieux utiliser la densité instrumentale définie sur $[0,1]$ par $g(x)=\frac{5}{2}x^{3/2}$ plutôt que la densité instrumentale uniforme : $g(x)=1$ pour $x\in[0,1]$.

(T3) En utilisant l'inégalité de Cauchy-Schwarz, montrer que $\int\pi(x)h^2(x)\frac{\pi(x)}{g(x)}dx\geq(\int\pi(x)|h(x)|dx)^2$

(T4) En déduire que $\tilde{V}_g(h)\geq\tilde{V}_{\tilde{g}}(h)$ avec $\tilde{g}(x)=\frac{|h(x)|\pi(x)}{\int|h(y)|\pi(y)dy}$. Déterminez les cas d'égalité.

(S3) **Simulation**. En guise d'illustration de la question précédente, on prendra pour Π la loi $\mathcal{N}(0,1)$ et on posera $h(x)=x$. Ecrire la densité $\tilde{g}$. Calculer la fonction de répartition de $\tilde{g}$. En déduire une façon de simuler un échantillon $(X_i)_{1\leq i\leq n}$ de variables aléatoires indépendantes et de même loi de densité $\tilde{g}$. Vérifier par simulation l'inégalité de la question 4 en prenant pour g la densité de la loi $\mathcal{N}(0,1)$.

La loi conduisant à la variance asymptotique minimale est donc de densité $\tilde{g}$ par rapport à la mesure de Lebesgue. Dans le cas où h est une fonction positive, il est facile de vérifier que $\tilde{\Pi}_{\tilde{g},n}(h)=\Pi(h)$. Néanmoins, dans ce cas, la densité optimale $\tilde{g}$ fait intervenir précisément la quantité $\int h(x)\pi(x)dx$ inconnue (c'est elle que l'on cherche à calculer) et l'on ne sait pas en général simuler directement suivant $\tilde{g}$. Nous allons maintenant déterminer une famille de densités associées à des lois sous lesquelles nous pouvons simuler et chercher dans cette famille de densités celle qui conduit la variance optimale (dans cette famille). Dans la pratique, si on sait simuler suivant des lois ayant une densité dans la famille finie de densités $\{g_1,\ldots,g_d\}$, alors, on sait aussi simuler suivant une loi de densité $\sum_{i=1}^d\alpha_i g_i$ avec $\forall i$, $\alpha_i\geq 0$ et $\sum_{i=1}^d\alpha_i=1$ (il suffit de simuler une variable Z à valeurs dans $\{1,\ldots,d\}$ telle que $\mathbb{P}(Z=k)=\alpha_k$ puis si Z a été tirée et vaut k, il suffit de simuler X conditionnellement à $Z=k$ suivant la loi de densité g_k). Nous chercherons par la suite un algorithme qui permet de déterminer les coefficients $\tilde{\alpha}=(\tilde{\alpha}_i)_{1\leq i\leq d}$ associés à une variance asymptotique minimale dans cette famille. En d'autres termes, $\tilde{\alpha}$ sera défini de la façon suivante : $\forall\alpha=(\alpha_1,\ldots,\alpha_d)\in\mathcal{S}=\{\beta=(\beta_1,\ldots,\beta_d);\sum_{i=1}^d\beta_i=1,\ \beta_i\geq 0\}$,

$$\tilde{V}_{\sum_{i=1}^d\alpha_i g_i}(h)\geq\tilde{V}_{\sum_{i=1}^d\tilde{\alpha}_i g_i}(h)\tag{2}$$

2 Réduction systématique de la variance

2.1 La fonction F

Dans la suite, on se fixe les densités $g_1, \ldots, g_d$ et on considère la fonction $\sigma^2 : \mathcal{S} \to \mathcal{S}$ définie par $\sigma^2(\alpha) = \tilde{V}_{\sum_{i=1}^d \alpha_i g_i}(h) + (\Pi(h))^2 = \int \pi(x) h^2(x) \frac{\pi(x)}{\sum_{i=1}^d \alpha_i g_i(x)} dx$. Pour que les intégrales soient bien définies, on suppose que

$$\int \pi(x) h^2(x) \frac{\pi(x)}{\inf\{g_i(x)\,, i = 1, \ldots, d\}} dx < \infty,$$

et on supposera de plus que $\int_{\{x; g_i(x)=0\}} \pi(x) dx = 0$ pour tout $i = 1, \ldots, d$. Posons maintenant $F : \mathcal{S} \to \mathcal{S}$ la fonction définie par : pour tout $k \in \{1, \ldots, d\}$,

$$[F(\alpha)]_k = \frac{\alpha_k \int g_k(x) \pi(x) h^2(x) \frac{\pi(x)}{[\sum_{i=1}^d \alpha_i g_i(x)]^2} dx}{\sigma^2(\alpha)} \tag{3}$$

où $[F(\alpha)]_k$ désigne la k-ième composante de $F(\alpha)$.

(T5) (Facultatif) Montrer que σ^2 admet un unique minimum et que ce minimum est atteint en un point de $\mathcal{S}$ que nous noterons $\tilde{\alpha}$. Montrer que $\tilde{\alpha}$ vérifie (2).

(T6) Vérifier que

$$\frac{\sigma^2(F(\alpha))}{\sigma^2(\alpha)} = \int \frac{\pi^2(x) h^2(x)}{\sum_{i=1}^d \alpha_i g_i(x)} \left(\sum_{i=1}^d \frac{\alpha_i g_i(x)}{\sum_{j=1}^d \alpha_j g_j(x)} \int \frac{g_i(u) \pi^2(u) h^2(u)}{[\sum_{k=1}^d \alpha_k g_k(u)]^2} du \right)^{-1} dx$$

(T7) En déduire en utilisant une inégalité de convexité que $\sigma^2(F(\alpha)) \leq \sigma^2(\alpha)$.

Ainsi la suite définie itérativement par

$$\begin{cases} \alpha^1 = (1/d, \ldots, 1/d)\,, \\ \alpha^{t+1} = F(\alpha^t) \quad \text{pour } t \geq 1. \end{cases} \tag{4}$$

est une suite de points de $\mathcal{S}$ telle qu'à chaque étape l'expression de la variance asymptotique diminue (i.e. $\tilde{V}_{\sum_{i=1}^d \alpha_i^{t+1} g_i}(h) \leq \tilde{V}_{\sum_{i=1}^d \alpha_i^t g_i}(h)$). Il est de plus possible de montrer que $\alpha^t \to_{t\to\infty} \tilde{\alpha}$.

Malheureusement, dans la formule de mise à jour de α^t (voir (3)), intervient la quantité $\int g_k(x) \pi(x) h^2(x) \frac{\pi(x)}{(\sum_{i=1}^d \alpha_i g_i(x))^2} dx$ qui n'est pas calculable a priori. Il nous faut alors l'approcher.

2.1.1 Approximation de la fonction F

Soit $\tilde{\alpha}^t = (\tilde{\alpha}^t_1, \ldots, \tilde{\alpha}^t_d) \in \mathcal{S}$. Soit une suite de vecteurs aléatoires $(X^t_i, Z^t_i)_{i \geq 0}$ indépendants et de même loi, la loi de (X^t_i, Z^t_i) étant définie de la façon suivante. On tire Z^t_i dans l'ensemble $\{1, \ldots, d\}$ telle que $\mathbb{P}(Z^t_i = l) = \tilde{\alpha}^t_l$. Puis si $Z^t_i = l$, on tire X^t_i conditionnellement à $Z^t_i = l$ suivant la loi de densité g_l. Notons $\mathbf{1}_{Z^t_i = l}$ la fonction qui vaut 1 si $Z^t_i = l$ et 0 sinon.

(T8) Montrer que $n^{-1} \sum_{i=1}^n \left(\frac{h(X^t_i)\pi(X^t_i)}{\sum_{s=1}^d \tilde{\alpha}^t_s g_s(X^t_i)} \right)^2 \mathbf{1}_{Z^t_i = k}$ converge avec probabilité un vers

$$\tilde{\alpha}^t_k \int g_k(x) \frac{\pi^2(x) h^2(x)}{(\sum_{i=1}^d \tilde{\alpha}^t_i g_i(x))^2} dx.$$

(T9) En déduire que $n^{-1} \sum_{i=1}^n \left(\frac{h(X^t_i)\pi(X^t_i)}{\sum_{s=1}^d \tilde{\alpha}^t_s g_s(X^t_i)} \right)^2$ converge avec probabilité un vers $\sigma^2(\tilde{\alpha}^t)$.

Pour n suffisamment grand, on approchera donc $[F(\tilde{\alpha}^t)]_k$ par

$$\frac{\sum_{\{i; Z^t_i = k\}} \left(\frac{h(X^t_i)\pi(X^t_i)}{\sum_{s=1}^d \tilde{\alpha}^t_s g_s(X^t_i)} \right)^2}{\sum_{i=1}^n \left(\frac{h(X^t_i)\pi(X^t_i)}{\sum_{s=1}^d \tilde{\alpha}^t_s g_s(X^t_i)} \right)^2}.$$

2.2 L'algorithme d'échantillonnage adaptatif : Algorithme et simulation

Poser $\tilde{\alpha}^1 = (1/d, \ldots, 1/d)$ puis, pour $t = 1, \ldots, T$ faire successivement :

Algorithme :

(a) **Simulation**

(i) Simuler n variables $(Z^t_i)_{1 \leq i \leq n}$ à valeurs dans $\{1, \ldots, d\}$ indépendantes et de même loi telle que $\mathbb{P}(Z^t_i = l) = \tilde{\alpha}^t_l$.

(ii) Pour chaque $i \in \{1, \ldots, n\}$, si $Z^t_i = k$ alors simuler indépendamment la variable X^t_i suivant la loi g_k.

(b) **Mise à jour**.

Pour $k = 1, \ldots, d$, poser $\tilde{\alpha}^{t+1}_k = \frac{\sum_{\{i; Z^t_i = k\}} \left(\frac{h(X^t_i)\pi(X^t_i)}{\sum_{s=1}^d \tilde{\alpha}^t_s g_s(X^t_i)} \right)^2}{\sum_{i=1}^n \left(\frac{h(X^t_i)\pi(X^t_i)}{\sum_{s=1}^d \tilde{\alpha}^t_s g_s(X^t_i)} \right)^2}$ puis, poser $\tilde{\alpha}^{t+1} = (\tilde{\alpha}^{t+1}_1, \ldots, \tilde{\alpha}^{t+1}_d)$.

L'approximation de $\Pi(h)$ sera alors donné par $n^{-1} \sum_{i=1}^n h(X^T_i) \frac{\pi(X^t_i)}{\sum_{s=1}^d \tilde{\alpha}^T_i g(X^T_i)}$.

(S4) On mettra en oeuvre l'algorithme sur un exemple où la loi Π visée est une $\mathcal{N}(0,1)$, $h(x) = x$, $d = 3$, et les densités g_1 et g_2 sont respectivement la densité d'une $\mathcal{N}(0,1)$, la densité $\tilde{g}$ définie dans la partie 1 et g_3 est une densité simulable quelconque.

(S5) On cherchera aussi des exemples numériques pour Π où on mettra en évidence que les $\tilde{\alpha}^t_i$ peuvent converger vers des quantités différentes de 0 et 1.

(S6) On mettra en évidence aussi la décroissance systématique de la variance à chaque itération de l'algorithme sur un exemple simple que vous choisirez.