

Méthodes de Monte-Carlo et réduction de la variance

Introduction et notations : Ce TP aborde présente diverses méthodes d'estimation des intégrales. Ces méthodes ont comme outil principal la méthode de Monte-Carlo. On notera dans la suite pour f une fonction réelle, $I(f)$ son intégrale sur \mathbb{R} par rapport à la mesure de Lebesgue sous réserve de convergence. Pour estimer $I(f)$, on peut estimer $\mathbb{E}_w[\frac{f}{w}(X)]$ où X est une variable aléatoire dont la densité associée est w , à l'aide de l'estimateur de la moyenne : $\hat{I}_{w,n} = \frac{1}{n} \sum_{i=1}^n \frac{f(x_i)}{w(x_i)}$ où $(x_i)_{i=1..N}$ est un échantillon de la v.a. X . On rappelle que si $\frac{f}{w}$ est intégrable par rapport à w alors cet estimateur converge presque sûrement vers $I(f)$. Enfin, on note $Var_w(f) = \int \frac{f^2(x)}{w(x)} dx - I(f)^2$. L'estimation de certaines intégrales est cruciale dans de nombreux domaines, en particulier en finance ou en assurance. Dans ce dernier domaine en particulier, on doit modéliser l'occurrence d'évènements rares qui "coûtent cher", ce qui peut induire une forte variance sur les calculs d'espérance de primes par exemple.

1 Echantillonnage pondéré (Importance Sampling)

Cette partie est dédiée à la méthode dite de l'importance sampling qui a pour but d'améliorer la convergence de l'estimateur de la moyenne en choisissant une densité réduisant la variance de l'estimateur. Bien sûr, si la fonction à intégrer est de signe constant, le choix optimal pour w est $f/I(f)$, ce qui suppose donc de connaître la valeur cherchée. En effet, ce choix rend constante la variable aléatoire $\frac{f}{w}(X)$ avec X de loi associée w et donc la variance est nulle.

Exercice 1 : Un exemple très simple d'importance sampling, soit $f_1 = \mathbf{1}_{x>a}$ et $X \sim \mathcal{N}(0,1)$ une variable aléatoire gaussienne centrée réduite.

1. Calculer par une méthode de Monte-Carlo (moyenne avec $\hat{I}_{0,n}$ et variance empirique) directe $\mathbb{E}[f_1(X)]$ pour $a = 0$ et $n = 1000$ trajectoires et pour $a = 8$, $n = 1000$ et $n = 10^6$ trajectoires.
2. On se donne la famille de poids w_α la densité associée à une variable X_α de loi $\mathcal{N}(\alpha,1)$. Pour quel paramètre la réduction de variance sera-t-elle optimale? Calcul par Monte-carlo.

Correction 2 : On cherche à calculer $\int_{\mathbb{R}} \mathbf{1}_{x>a} \exp(-x^2/2) \frac{dx}{\sqrt{2\pi}}$. Si on note $g(x) = \frac{1}{\sqrt{2\pi}} \mathbf{1}_{x>a} \exp(-x^2/2)$, on calcule en prenant un poids w_α , $\mathbb{E}[\frac{g}{w_\alpha}(X_\alpha)]$, on a donc $\frac{g}{w_\alpha}(x) = f_1(x) \exp(\alpha^2/2 - \alpha x)$. Le poids optimum est obtenu par l'annulation de

$$\exp(-\frac{(a+\alpha)^2}{2}) = 2\alpha\sqrt{2\pi}(1 - Ndist(a+\alpha)),$$

où $Ndist$ est la fonction de répartition de la loi normale. La détermination du paramètre optimal fait intervenir la fonction de répartition, quantité que l'on souhaite justement calculer en a . Cette égalité montre que si α existe $\alpha > a$. En effectuant un développement limité de la fonction

de répartition lorsque a est grand, on s'aperçoit que α doit être proche de a . Expérimentalement, $\alpha = a$ est un bon choix pour la précision de l'estimateur.

On considère maintenant la fonction réelle f_2 linéaire par morceaux qui vaut 0 en dehors de $[a, +\infty]$ et $M > 0$ en dehors de $[-\infty, b]$ avec $a < b$.

1. Proposer une méthode de réduction de variance en utilisant l'importance sampling pour calculer $\mathbb{E}[f_2(X)]$ où $X \sim \mathcal{N}(0, 1)$ et l'implémenter.
2. Comparer cette méthode à un calcul exact pour une valeur de a très grande, par exemple, $a = 7$. (Indication : la fonction f_2 est une combinaison linéaire de fonction du type $(x - K)_+ = f_3$.)

Correction : Cette seconde partie de l'exercice est une application directe de la première.

Exercice 2 : On considère $N = 100$ variables B_i de Bernoulli i.i.d. de probabilité $p = \mathbf{P}(B_i = 1)$, et on considère $L = \sum_{i=1}^n B_i$. On cherche à estimer avec une erreur relative de 1% la quantité suivante : $\mathbb{E}_Q[f_i(L)]$ où f_i est l'une des fonctions définies dans l'exercice précédent, et Q est la loi de L .

1. Après avoir fait un calcul exact, faire une estimation directe par Monte-Carlo pour des paramètres qui introduiront une variance forte, pour f_3 par exemple.

Correction 1 : Ce calcul exact en matlab peut se faire plus généralement pour une série de Bernoulli indépendantes qui n'ont pas les mêmes probabilités, en faisant explicitement les 100 convolutions pour le calcul de la loi. (Son support est de cardinal 101.) On peut le faire avec une boucle en Matlab, en utilisant toutefois la structure vectorielle.

2. On considère la famille de densités suivante : $\frac{dQ_y}{dQ} = \exp(yL) / \mathbb{E}_Q[\exp(yL)]$ pour $y \in \mathbb{R}$. On cherche le paramètre y qui minimise la variance, i.e. minimiser l'expression $\mathbb{E}[\mathbf{1}_{L>k} (\frac{dQ}{dQ_y})^2]$. Trouver un majorant simple de cette dernière expression et en déduire un paramètre qui semble réduire la variance.

Correction 2 : Le majorant demandé est : $\mathbb{E}_Q[\exp(yL)]^2 \exp(-2yK)$. Comme $L > K$, on a $\exp(yL) \geq \exp(yK)$, c'est suffisant pour obtenir l'inégalité. Pour obtenir un bon paramètre candidat, on cherche à minimiser le majorant trouvé. Pour cela, notons $\mathbf{K}_L(y) = \log(\mathbb{E}_Q[\exp(yL)])$. En dérivant le majorant par rapport à y , on obtient $K = \mathbf{K}'(y)$. Il existe un unique $y \in \mathbf{R}$ qui vérifie cette égalité. Explicitement, on a : $\mathbf{K}(u) = N \log(1 + p(\exp(u) - 1))$, et on en déduit $y = \log(\frac{(1-p)k}{p(N-k)})$. Si cette valeur est négative, on a aucun intérêt à faire un changement de probabilité, car cela signifie que $\mathbb{E}[L] > K$. Dans la question suivante, la méthode proposée est à appliquer lorsque $y > 0$.

3. En écrivant explicitement le changement de densité, proposer une méthode de réduction de variance et l'implémenter.

Correction 3 : On traite un cas plus général ici, où la somme considérée est $L = \sum_{i=1}^n m_i B_i$ et $p_i = \mathbf{P}(B_i = 1)$. Le point y peut se calculer par

dichotomie par exemple. Dans le cas où $K_L(y) = \sum_{i=1}^n \log(p_i \exp(m_i) + (1 - p_i))$, on a l'égalité suivante qui montre l'algorithme utilisé :

$$\frac{dQ}{dQ_y} = \prod_{i=1}^n (p_i/\tilde{p}_i)^{D_i} \left(\frac{1-p_i}{1-\tilde{p}_i}\right)^{1-D_i} = \exp(-yL + K_L(y)),$$

en notant $\tilde{p}_i = p_i \frac{\exp(y m_i)}{1 + p_i(\exp(y m_i) - 1)}$ et D_i sont des variables de Bernoulli indépendantes de probabilité \tilde{p}_i . L'algorithme de réduction de variance s'en déduit naturellement en exécutant la méthode de Monte-Carlo sur les D_i . Notons que le calcul exact ne peut plus être mené lorsqu'on introduit des poids sur les variables de Bernoulli, par exemple pour des valeurs \mathbb{Q} linéairement indépendantes le support de la loi de L est de cardinal 2^{100} et le calcul exact est équivalent à un problème combinatoire de complexité exponentielle. Dans ces situations, cette méthode est envisageable, elle dérive de la méthode dite du point selle.

2 Utilisation des symétries (systematic sampling)

Cette partie donne deux exemples d'utilisation des symétries après avoir présenté la démarche.

Exercice 3 : Soit A un domaine de \mathbb{R}^k et w une densité sur A , on considère un ensemble de transformations (difféomorphismes) de A , $\tilde{T} = T_i, i = 1 \dots n$ qui préservent la forme volume et telles que $w \circ T_i = w$. On considère,

$$f^{\tilde{T}} = \frac{1}{n} \sum_{i=1}^n f \circ T_i(x)$$

1. Montrer que $Var\left(\frac{f^{\tilde{T}}}{w}\right) = Var\left(\frac{f}{w}\right) - \mathbb{E}_w\left[\left(\frac{f-f^{\tilde{T}}}{w}\right)^2\right]$. Donner une interprétation géométrique de cette égalité.

Correction 1 : Cette égalité est le théorème de pythagore sur les espérances conditionnelles. Il suffit de projeter la variable $f/w(X)$ sur les fonctions invariantes par le groupe de transformations envisagé. Le seul point qu'il faut étudier est : étant donnés n réels, quelle est la projection L^2 par rapport à la mesure de comptage de cette famille de réels sur les constantes : c'est la moyenne des réels. Pour le cas général, il suffit de conditionner par rapport à la plus grande tribu invariante par $T : \bigcap_{i=1}^n \mathbf{F}_{T_i(X)}$, avec \mathbf{F}_X la tribu engendrée par la variable aléatoire X .

2. Proposer une méthode qui pourrait permettre de réduire le temps de calcul et donner la condition d'application. On suppose que le temps de calcul est du aux seules évaluations de la fonction. On constate donc que si $f \circ T_i = f$ quelque soit i , cette méthode est plus coûteuse que la méthode directe.

Correction 2 : Pour avoir une réduction de variance efficace, il faut que le gain en variance soit plus important que simuler $n - 1$ autres séries d'estimations, en particulier, il est souhaitable d'avoir :

$$Var_w(f^{\tilde{T}}/w) < \frac{1}{n-1} \mathbb{E}_w\left[\left(\frac{f-f^{\tilde{T}}}{w}\right)^2\right].$$

Exercice 4 : Le contrôle antithétique est une application immédiate des propriétés précédentes.

1. On cherche à calculer $I(f)$ où f est une fonction définie sur $[0, 1]$: implémenter une méthode de réduction de variance et prendre des exemples représentatifs des différentes situations, en utilisant le groupe de transformations Id et $1 - Id$ qui laissent invariante la loi uniforme.

Correction 1 : On est dans le cas d'une vraie réduction de variance car :

$$Var\left(\frac{f^{\bar{T}}}{w}\right) = \frac{1}{2}Var\left(\frac{f}{w}\right) + \frac{1}{2}Cov_w[f \circ Id, f \circ 1 - Id].$$

Le deuxième terme est effectivement strictement négatif quand f est monotone et strictement monotone sur un sous intervalle non trivial. Plus généralement, on a $Cov(f(U), g(U)) < 0$ où U est une v.a. uniforme, f et g monotones, l'une croissante, l'autre décroissante, avec enfin la condition que chacune soit strictement monotone sur un sous intervalle de $[0, 1]$. C'est un exercice facile.

2. Implémenter cette méthode avec le groupe $\{Id, -Id\}$ dans le cadre de l'exercice 1, commenter.

3 Variable de contrôle

Cette technique repose sur l'idée d'introduire une variable Y d'espérance nulle telle que : $Var(X + Y) < Var(X)$, on estime alors $\mathbb{E}[X + Y]$.

Exercice 5 : On considère le prix d'une option sur deux actifs financiers.

$$\varphi(g_1, g_2) := (\lambda_1 \exp(\sigma_1 g_1) + \lambda_2 \exp(\sigma_2 g_2) - K)_+$$

où $\lambda_1, \lambda_2, \sigma_1, \sigma_2, K > 0$ et où g_1, g_2 sont deux variables indépendantes de loi $\mathcal{N}(0, 1)$.

Evaluer le prix $\mathbb{E}(\varphi(g_1, g_2))$ en programmant chacune des méthodes suivantes. Comparer leur variance empirique sur un échantillon fixé.

1. Méthode de Monte-Carlo sans réduction de variance, on pourra prendre $\lambda_1 = \lambda_2 = 5$, $K = 10$, $\sigma_{1,2} = 0, 5$.
2. Variables de contrôle.
 - (a) "Parité Call-Put" : considérer la variable de contrôle

$$(\lambda_1 \exp(\sigma_1 g_1) + \lambda_2 \exp(\sigma_2 g_2) - K)_+ - (K - \lambda_1 \exp(\sigma_1 g_1) - \lambda_2 \exp(\sigma_2 g_2))_+$$

(On rappelle la transformée de Laplace d'une variable gaussienne $\mathcal{N}(0, t)$ est $z \mapsto \exp(\frac{tz^2}{2})$).

- (b) Développement à l'ordre 1 : considérer la variable de contrôle

$$\left((\lambda_1 + \lambda_2) \exp\left(\frac{\lambda_1 \sigma_1}{\lambda_1 + \lambda_2} g_1 + \frac{\lambda_2 \sigma_2}{\lambda_1 + \lambda_2} g_2 \right) - K \right)_+$$

Références

- [1] MICHAEL EVANS AND TIM SWARTZ : "*Approximating Integrals via Monte Carlo and Deterministic Methods*", Oxford Statistical Science Series.
- [2] PIERRE COHORT : "*TP Monte-Carlo pour la finance*"