Méthodes de Monte Carlo et processus stochastiques

Pierre Del Moral - Stefano De Marco - Massimiliano Gubinelli - Benjamin Jourdain

Map 564 - PC 3 - TP Scilab - 25 Janvier 2013

Les programmes scilab associées aux énoncés en rouge sont fournis. Les énoncés en bleu sont des questions d'analyse, à résoudre une fois les travaux pratiques terminés.

EXERCICE 1 -: Aiguille de Buffon

Le but de cet exercice est de refaire l'une des expériences de simulation de Monte Carlo les plus anciennes, proposée en 1733 par Georges-Louis Leclerc de Buffon. On lance des aiguilles de longueur 1 au hasard sur un parquet dont les lattes ont une largeur 1. Nous conviendrons qu'une aiguille ne peut couper au plus qu'une de ces lattes. L'objectif de cette expérience est de calculer numériquement la probabilité pour une aiguille touche le bord d'une latte.

1. **Méthode de Monte Carlo :** Vérifier numériquement que la probabilité pour qu'une aiguille touche le bord d'une latte est de $2/\pi \simeq 0,64$.
2. **Loi des grands nombres :** Proposer un estimateur empirique X_n sans biais de $2/\pi$, et calculer sa variance σ^2. Tracer la courbe des estimations empiriques moyennes au cours de n simulations.
3. **Théorème de la limite centrale :** On répète I fois l'expérience, et l'on note X_n^i l'estimation empirique associé à l'expérience $i = 1, \ldots, I$. On note $V_n^i := \sqrt{n}(\frac{X_n^i - 2/\pi}{\sigma})$ les termes de fluctuation des erreurs à chaque expérience.

   Tracer l'histogramme des valeurs $(V_n^i)_{1 \leq i \leq I}$. Comparer ce dernier à celui de I variables aléatoires $(W_n^i)_{1 \leq i \leq I}$ indépendantes de densité $g(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2}$.
4. **Intervalles de confiance :** Déterminer empiriquement les valeurs des probabilités $p_n(\delta) = \text{Proba}(|V_n| > \delta)$, et $p(\delta) = \text{Proba}(|W^1| > \delta)$, pour différentes valeurs de $\delta > 0$. Comparer ces valeurs pour différentes valeurs des paramètres (n, I, δ). Illustrer graphiquement les encadrements fournis par l'inégalité de Mills :

$$\forall \delta > 0 \qquad \frac{1}{\delta + 1/\delta}\, g(\delta) \leq \text{Proba}(W^1 > \delta) \leq \frac{1}{\delta}\, g(\delta)$$

5. **Vérifier les inégalités de Mills.**

On se donne par la suite un modèle "boîte noire" représentant un système entrée-sortie réel $(X \xrightarrow{f} Y)$ défini par la relation $Y = f(X)$. L'objectif est d'estimer la valeur $E(g(Y))$, pour une certaine fonction g sur l'ensemble des sorties Y. On suppose que $g(Y)$ est de carré intégrable

EXERCICE 2 -: Échantillonnage d'importance

On suppose que les variables d'entrée sont de loi gaussienne centrée réduite. On se place dans la situation $f(x) = x$, et $g(x) = (x-2)_+$.

1. **Méthode de Monte Carlo :** Proposer un estimateur empirique I_n fondé sur la simulation de n copies indépendantes de X. Décrire un programme scilab simulant plusieurs fois cet estimateur.

   Vérifier par intégration numérique, puis par simulation les assertions suivantes : $E(g(Y)) \simeq m = 8.5\ 10^{-3} \simeq I_n$, et $\mathrm{Var}(I_n) = \frac{\sigma^2}{n}$, avec $\sigma^2 \simeq 5,7 \times 10^{-3}$. Tracer les courbes des moyennes empiriques obtenues sur plusieurs réalisations, comparer les moyennes des réalisations avec la quantité recherchée.

2. On note $(I_n^i)_{1\leq i\leq N}$, les estimations empiriques associées à N expériences. Tracer l'histogramme des variables $V_n^i = \sqrt{n}\left(\frac{I_n^i - m}{\sigma}\right)$, puis le comparer graphiquement avec la densité gaussienne centré réduite pour $(n, N) = (500, 1000)$, puis pour $n = N = 1000$.

3. **Échantillonnage d'importance :** Proposer une technique de simulation en utilisant la loi d'importance gaussienne réduite, et centrée en 2. On note J_n l'estimateur correspondant. Vérifier par intégration numérique, puis par simulation, l'estimation suivante $\mathrm{Var}(J_n) \simeq \frac{1}{n}\ \sigma_J^2$, avec $\sigma_J^2 \simeq\ 9,7 \times 10^{-5}$. Tracer les courbes des moyennes empiriques obtenues sur plusieurs réalisations, comparer les moyennes des réalisations avec la quantité recherchée. Quel est le gain en temps de calcul de cet estimateur, pour la même précision de variance que la méthode de Monte Carlo?

4. On note $(J_n^i)_{1\leq i\leq N}$, les estimations empiriques associées à N expériences. Tracer l'histogramme des variables $W_n^i = \sqrt{n}\left(\frac{J_n^i - m}{\sigma_J}\right)$, puis le comparer graphiquement avec la densité gaussienne centré réduite pour $n = N = 1000$.

EXERCICE 3 -: Échantillonnage antithétique & Méthodes de stratification

On suppose que les variables d'entrée sont de loi uniformes entre -1 et $+1$. On se place dans la situation $f(x) = e^x$, et $g(x) = x$.

1. **Méthode de Monte Carlo :** Proposer un estimateur empirique I_n fondé sur la simulation de n copies indépendantes de X. Vérifier que l'on a $E(g(Y)) = \sinh(1) \simeq 1,18 \simeq I_n$, et $\mathrm{Var}(I_n) = \frac{1}{2n}\ (1-e^{-2}) \simeq \frac{1}{n}\ 0,43$. Tracer les courbes des moyennes empiriques obtenues sur plusieurs réalisations, et comparer les moyennes des réalisations avec la quantité recherchée.

2. **Échantillonnage antithétique :** Vérifier que $(f(u)-f(v))(f(-u)-f(-v)) \leq 0$, pour tout $(u,v) \in [-1,1]$, et en déduire une technique d'échantillonnage antithétique fondée sur la simulation de n variables uniformes sur $[-1,1]$. On note I'_n l'estimateur empirique correspondant. Vérifier que l'on a $\mathrm{Var}(I') \simeq \frac{1}{n}\ 0,025$. Tracer les courbes des moyennes empiriques obtenues. Quel est le gain en temps de calcul de cet estimateur, pour la même précision de variance que la méthode de Monte Carlo?

3. **Stratification proportionnelle :** Proposer une technique de stratification proportionnelle fondée sur la simulation de variables uniformes sur $[-1,0]$, et uniformes sur $[0,1]$. On note J_n l'estimateur correspondant. Vérifier que l'on a $\mathrm{Var}(J_n) \simeq \frac{1}{n}\, 0,14$. Tracer les courbes des moyennes empiriques obtenues. Quel est le gain en temps de calcul de cet estimateur, pour la même précision de variance que la méthode de Monte Carlo?

4. **Stratification non proportionnelle :** Proposer une technique de stratification proportionnelle fondée sur la simulation d'une proportion $n_1 = rn$ uniformes sur $[-1,0]$, et $n2 = (1-r)n$ de variables uniformes sur $[0,1]$, avec $r \in]0,1[$ tel que $n1, n2$ soient entiers. Tracer les courbes des moyennes empiriques obtenues, pour différentes valeurs de r. Quels sont les gains en temps de calcul de ces estimateurs, pour la même précision de variance que la méthode de Monte Carlo?

5. **Stratification optimale sur les deux strates** $(S_1, S_2) = ([-1,0],[0,1])$**:** Calculer les valeurs des paramètres

$$\begin{aligned}
\sigma_1^2 &:= E(e^{2X} \mid X \in [-1,0]) - E(e^X \mid X \in [-1,0])^2 \\
\sigma_2^2 &:= E(e^{2X} \mid X \in [0,1]) - E(e^X \mid X \in [0,1])^2
\end{aligned}$$

   et proposer une technique de stratification optimale. Déterminer les estimateurs empiriques, ainsi que leur variances. Tracer les courbes des moyennes empiriques obtenues. Comparer numériquement ces différentes méthodes de stratification et d'échantillonnage antithétique.

EXERCICE 4 -: Variables de contrôle

Les modèles "boîte noire" représentant des système entrée-sortie de codes numériques sont souvent très couteux en temps de calcul. Supposons que nous ayons un modèle entrée-sortie réduit $(X \xrightarrow{f_r} Y_r)$ facile à simuler pour différentes valeurs des entrées. On supposera connue la quantité $m_r = E(g(Y_r))$, et $E(g_r(Y)^2) < \infty$.

On note $(X_i)_{1 \leq i \leq n}$ une suite de copies indépendantes de la variable d'entrée X, et l'on pose

$$I_n = \frac{1}{n}\sum_{i=1}^{n} g(f(X_i)) \qquad I_n^c = m_r + \frac{1}{n}\sum_{i=1}^{n}\left(g(f(X_i)) - g(f_r(X_i))\right)$$

1. Vérifier que I_n et I_n^c sont des estimations non biasées de $E(g(Y))$, et calculer leur variances.

2. On suppose que les variables d'entrée X suivent une loi uniforme sur $[0,1]$, $f(x) = e^x$, $f_r(x) = 1 + x$, et $g(x) = x$. Tracer les courbes des moyennes empiriques I_n, I_n^c sur plusieurs réalisations, puis les histogrammes des fluctuations d'erreurs. Comparer les variances de ces deux estimateurs.

EXERCICE 5 -: Stratification gaussienne

On veut calculer $\mathbb{E}(f(X))$ où X est un vecteur aléatoire à valeurs dans $\mathbb{R}^d$ et $f : \mathbb{R}^d \to \mathbb{R}$ est une fonction telle que $f(X)$ est de carré intégrable. On suppose qu'il existe une partition $D_1, \ldots, D_I$ de $\mathbb{R}^d$ telle que pour tout $i \in \{1, \ldots, I\}$, $p_i = \mathbb{P}(X \in D_i)$ est strictement positif et connu et on sait simuler des variables aléatoires $(X_k^i)_{k\geq 1}$ suivant la loi conditionnelle de X sachant $X \in D_i$ avec les variables $(X_k^i)_{1\leq i\leq I, k\geq 1}$ indépendantes.

1. Pour $(n_1, \ldots, n_I) \in \mathbb{N}^{*I}$, on pose $Y_{n_1,\ldots,n_I} = \sum_{i=1}^I \frac{p_i}{n_i} \sum_{k=1}^{n_i} f(X_k^i)$. Calculer $\mathbb{E}(Y_{n_1,\ldots,n_I})$. Quel est le comportement de $Y_{n_1,\ldots,n_I}$ lorsque $\min_{1\leq i\leq I} n_i \to +\infty$?

2. On note $\sigma_i^2 = \mathrm{Var}\,(f(X_1^i))$. Montrer que $\mathrm{Var}\,(Y_{n_1,\ldots,n_I}) = \frac{1}{\sum_{i=1}^I n_i} v\left(\frac{n_1}{\sum_i n_i}, \ldots, \frac{n_I}{\sum_i n_i}\right)$ où $v(q_1, \ldots, q_I) = \sum_{i=1}^I \frac{(p_i \sigma_i)^2}{q_i}$.

3. Montrer que $\forall q = (q_1, \ldots, q_I)$ probabilité sur $\{1, \ldots, I\}$, $v(q) \geq \left(\sum_{i=1}^I p_i \sigma_i\right)^2$ où le minorant est atteint pour $q^\star = \left(\frac{p_1\sigma_1}{\sum_{i=1}^I p_i\sigma_i}, \ldots, \frac{p_I\sigma_I}{\sum_{i=1}^I p_i\sigma_i}\right)$. Pourquoi la clé de répartition optimale $q^\star$ des tirages dans les strates est-elle en général inconnue?

   En remarquant que $\sigma_i^2 = \mathbb{E}(f^2(X)|X \in D_i) - (\mathbb{E}(f(X)|X \in D_i))^2$, vérifier que $v(p_1, \ldots, p_I) \leq \mathrm{Var}\,(f(X))$ et interpréter ce résultat.

On suppose désormais que $X \sim \mathcal{N}_d(0, I_d)$ et que $D_i = \{y_{i-1} \leq u.X \leq y_i\}$ où $-\infty = y_0 < y_1 < \ldots < y_{I-1} < y_I = +\infty$ et $u \in \mathbb{R}^d$ est tel que $|u| = 1$. On note $N(y) = \int_{-\infty}^y e^{-\frac{z^2}{2}} \frac{dz}{\sqrt{2\pi}}$.

1. Quelle est la loi de $u.X$? Exprimer p_i à l'aide de la fonction N.

2. Montrer que $u.X$ et $X - (u.X)u$ sont indépendants.

3. Soit U une variable aléatoire uniforme sur $[0, 1]$ indépendante de X. Vérifier que $Z_i = N^{-1}[N(y_{i-1})(1 - U) + N(y_i)U]$ suit la loi conditionnelle de $u.X$ sachant $X \in D_i$. Conclure que $X + (Z_i - u.X)u$ suit la loi conditionnelle de X sachant $X \in D_i$.