Méthodes de Monte-Carlo par chaînes de Markov et algorithmes particulaires

Examen du mardi 30 avril 2019 (9h00-12h00)

L'objectif de ce problème est d'étudier la méthode proposée par Hobert, Jones, Presnell et Rosenthal en 2002 dans *Biometrika* en vue de construire, à partir d'une seule trajectoire, des intervalles de confiance pour les calculs d'intégrales par des méthodes de Monte Carlo par chaîne de Markov sous la condition de minoration du noyau (1) qui n'est pas trop restrictive.

Soit $(Y_j)_{j\geq 1}$ une suite de variables aléatoires réelles i.i.d. de carré intégrable $\mathbb{E}[Y_1^2] < \infty$ et centrées $\mathbb{E}[Y_1] = 0$. On pose $S_0 = 0$ et pour $n \in \mathbb{N}^*$, $S_n = \sum_{i=1}^n Y_j$. Soit $(N_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires à valeurs dans $\mathbb{N}$ telle que $\frac{N_n}{n}$ converge en probabilité vers une constante $c \in]0, +\infty[$ lorsque $n \to \infty$. On note $\lfloor x \rfloor$ la partie entière de $x \in \mathbb{R}$ ($\lfloor x \rfloor$ est l'entier t.q. $\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$).

1. Énoncer le théorème de la limite centrale pour la suite $(Y_j)_{j\geq 1}$. En déduire que $\frac{S_{\lfloor cn \rfloor}}{\sqrt{n}}$ converge en loi vers $\xi \sim \mathcal{N}_1(0, c\mathrm{Var}(Y_1))$ lorsque $n \to \infty$.

2. Soit $x > 0$. Pour $j \geq 1$ on note $A_j = \{|S_j| \geq x\}$ et $B_j = A_j \cap \{\bigcap_{i=1}^{j-1} A_i^c\}$. Soit $n \geq 1$.

   (a) Montrer que $\mathbb{P}(\max_{1\leq j\leq n} |S_j| \geq x) \leq \frac{1}{x^2} \sum_{j=1}^n \mathbb{E}[S_j^2 1_{B_j}]$.

   (b) Remarquer que pour $1 \leq j \leq n$, $S_j^2 \leq S_n^2 + 2S_j(S_j - S_n)$ et en déduire que $\mathbb{E}[S_j^2 1_{B_j}] \leq \mathbb{E}[S_n^2 1_{B_j}]$.

   (c) Conclure que $\mathbb{P}(\max_{1\leq j\leq n} |S_j| \geq x) \leq \frac{n\mathbb{E}[Y_1^2]}{x^2}$.

Soient $\varepsilon, \alpha > 0$ et $n \geq 1$.

3. En remarquant que, puisque N_n est à valeurs entières, $\left|\frac{N_n}{n} - c\right| < \alpha$ implique $|N_n - \lfloor cn \rfloor| \leq \lfloor \alpha n \rfloor + 1$, montrer que

$$\mathbb{P}\left(\frac{|S_{N_n} - S_{\lfloor cn \rfloor}|}{\sqrt{n}} \geq \varepsilon\right) \leq \mathbb{P}\left(\left|\frac{N_n}{n} - c\right| \geq \alpha\right) + \mathbb{P}\left(\max_{1\leq j\leq \lfloor \alpha n \rfloor + 1} \left|\sum_{i=1}^j Y_{\lfloor cn \rfloor + i}\right| \geq \varepsilon\sqrt{n}\right)$$
$$+ \mathbb{P}\left(\max_{1\leq j\leq \lfloor \alpha n \rfloor + 1} \left|\sum_{i=1}^j Y_{\lfloor cn \rfloor + 1 - i}\right| \geq \varepsilon\sqrt{n}\right),$$

   où, par convention, $Y_i = 0$ pour $i \leq 0$.

4. En déduire que $\mathbb{P}\left(\frac{|S_{N_n} - S_{\lfloor cn \rfloor}|}{\sqrt{n}} \geq \varepsilon\right) \leq \mathbb{P}\left(\left|\frac{N_n}{n} - c\right| \geq \alpha\right) + \frac{2(\lfloor \alpha n \rfloor + 1)\mathbb{E}[Y_1^2]}{n\varepsilon^2}$.

5. En déduire que $\frac{S_{N_n} - S_{\lfloor cn \rfloor}}{\sqrt{n}}$ converge en probabilité vers 0 lorsque $n \to \infty$ et conclure que $\frac{S_{N_n}}{\sqrt{n}}$ converge en loi vers $\xi \sim \mathcal{N}_1(0, c\mathrm{Var}(Y_1))$ lorsque $n \to \infty$.

Soient E un espace d'états muni d'une tribu $\mathcal{E}$ et $P(x,dy)$ un noyau markovien sur cet espace qui vérifie

(D1) $\exists V : E \to \mathbb{R}_+$ mesurable, $\exists K \in \mathbb{R}_+,\ \exists \gamma \in]0,1[,\ \forall x \in E,\ PV(x) \leq \gamma V(x) + K$

(D2′) $\exists R > \dfrac{4K}{(1-\sqrt{\gamma})^2},\ \exists \alpha \in]0,1],\ \forall x,y$ t.q. $V(x)+V(y) \leq R,\ P(x,.) \wedge P(y,.)(E) \geq \alpha.$

Soit $(X_k)_{k\in\mathbb{N}}$ une chaîne de Markov de noyau de transition P.

6. Pourquoi le noyau P admet-il une unique probabilité invariante π?
7. Pour $f : \mathbb{E} \to \mathbb{R}$ mesurable et telle que $\pi(|f|) < \infty$, quel est le comportement asymptotique de $\frac{1}{n}\sum_{k=0}^{n-1} f(X_k)$ lorsque $n \to \infty$?
8. Donner une hypothèse sur f sous laquelle $\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}(f(X_k) - \pi(f))$ converge en loi vers une gaussienne centrée dont on précisera la variance $\sigma^2(f)$.

On suppose désormais également l'existence de $\nu \in \mathcal{P}(E)$ et de $\rho : E \to \mathbb{R}_+$ mesurable telles que

$$\pi(\rho) > 0 \text{ et } \forall x \in E,\ \forall A \in \mathcal{E},\ P(x,A) \geq \rho(x)\nu(A). \tag{1}$$

9. Montrer que $\forall x \in E,\ \rho(x) \leq 1$. Lorsque $\inf_{x\in E}\rho(x) > 0$, quel nom porte la condition (1)?
10. Lorsque P est le noyau de transition d'un algorithme de Metropolis-Hastings sur $\mathbb{R}^d$ de densité cible $\eta(x)$ par rapport à la mesure de Lebesgue et de densité de proposition $q(x,y)$, donner des hypothèses sur η, q qui assurent que la condition (1) est satisfaite.

Soit $g : E \times E \to \mathbb{R}$ mesurable et telle que pour tout $x \in E$, $\rho(x)\nu(dy) = g(x,y)P(x,dy)$. On note $R(x,dy) = 1_{\{\rho(x)=1\}}\delta_x(dy) + 1_{\{\rho(x)<1\}}\frac{P(x,dy)-\rho(x)\nu(dy)}{1-\rho(x)}$. Soit $x \in E$, $(X_1,X_2) \sim P(x,dx_1)P(x_1,dx_2)$ et $(U_1,U_2) \sim \mathcal{U}[0,1]^2$ indépendants.

11. Montrer que R est un noyau markovien.
12. Soit $Z_1 = 1_{\{U_2 \leq g(X_1,X_2)\}}$ et $\tilde{\varphi} : E \times \{0,1\} \to \mathbb{R}$ mesurable bornée. Montrer que

$$\mathbb{E}[\tilde{\varphi}(X_1,Z_1)1_{\{U_1>g(x,X_1)\}}] = (1-\rho(x))\int_E (\tilde{\varphi}(x_1,0)(1-\rho(x_1)) + \tilde{\varphi}(x_1,1)\rho(x_1))R(x,dx_1).$$

En déduire que

$$\mathcal{L}((X_1,Z_1)|U_1 > g(x,X_1)) = R(x,dx_1)((1-\rho(x_1))\delta_0(dz_1) + \rho(x_1)\delta_1(dz_1))$$

Montrer de manière analogue que

$$\mathcal{L}((X_1,Z_1)|U_1 \leq g(x,X_1)) = \nu(dx_1)((1-\rho(x_1))\delta_0(dz_1) + \rho(x_1)\delta_1(dz_1)).$$

Soit $(U_k)_{k\in\mathbb{N}^*}$ une suite de variables aléatoires i.i.d. suivant la loi uniforme sur $[0,1]$ indépendante de la chaîne de Markov $(X_k)_{k\in\mathbb{N}}$ de noyau de transition P. Pour $k \in \mathbb{N}$, on pose $Z_k = 1_{\{U_{k+1}\leq g(X_k,X_{k+1})\}}$ et pour $n \in \mathbb{N}^*$, $N_n = \sum_{k=0}^{n-1} Z_k$. En généralisant le résultat de la question 12, on peut montrer, et on le supposera désormais, que $((X_k, Z_k))_{k\in\mathbb{N}}$ est une chaîne de Markov sur l'espace étendu $\tilde{E} = E \times \{0,1\}$ de noyau de transition

$$\tilde{P}((x,w),(dy,dz)) =(1_{\{w=0\}}R(x,dy) + 1_{\{w=1\}}\nu(dy))((1-\rho(y))\delta_0(dz) + \rho(y)\delta_1(dz)).$$

On note $\tilde{\mu}_n(dx,dz)$ la loi de (X_n,Z_n) on définit également la probabilité sur $\tilde{E}$

$$\tilde{\pi}(dx,dw) = \pi(dx)((1-\rho(x))\delta_0(dw) + \rho(x)\delta_1(dw)).$$

13. Montrer que $\tilde{\pi}$ est invariante par $\tilde{P}$.

14. Soit $\tilde{\sigma}$ une probabilité sur $\tilde{E}$ invariante par $\tilde{P}$ et

$$\sigma(dy) = \int_{(x,w)\in\tilde{E}} \tilde{\sigma}(dx,dw)(1_{\{w=0\}}R(x,dy) + 1_{\{w=1\}}\nu(dy)).$$

Montrer que $\tilde{\sigma}(dy,dz) = \sigma(dy)((1-\rho(y))\delta_0(dz) + \rho(y)\delta_1(dz))$ puis que σ est invariante par P. Conclure que $\tilde{\pi}$ est l'unique probabilité invariante de $\tilde{P}$.

15. Montrer que pour $n \geq 1$, $\tilde{\mu}_n(dx,dz) = \mu_n(dx)((1-\rho(x))\delta_0(dz) + \rho(x)\delta_1(dz))$. Comment s'interprète la probabilité μ_n?

16. Soit $\tilde{\varphi} : \tilde{E} \to \mathbb{R}$ mesurable et bornée par $1/2$. Vérifier que pour $n \geq 1$, $\tilde{\mu}_n(\tilde{\varphi}) - \tilde{\pi}(\tilde{\varphi}) = \mu_n(\varphi) - \pi(\varphi)$ où $\varphi(x) = (1-\rho(x))\tilde{\varphi}(x,0) + \rho(x)\tilde{\varphi}(x,1)$. En déduire que $\lim_{n\to\infty} d_{\mathrm{TV}}(\tilde{\mu}_n, \tilde{\pi}) = 0$.

17. Pour $z \in \{0,1\}$ et $f : E \to \mathbb{R}$ mesurable telle que $\pi(|f|) < \infty$, quel est le comportement asymptotique de $\frac{1}{n}\sum_{k=0}^{n-1} 1_{\{Z_k=z\}}f(X_k)$ lorsque $n \to \infty$? En déduire que $\frac{N_n}{n}$ converge p.s. vers $\pi(\rho)$.

Avec la convention $\min\emptyset = +\infty$, on pose $\tau_1 = \min\{k \in \mathbb{N} : Z_k = 1\}$ et pour $j \in \mathbb{N}^*$, lorsque $\tau_j < \infty$, on définit par récurrence, $\tau_{j+1} = \min\{k > \tau_j : Z_k = 1\}$.

18. Pourquoi a-t-on $\mathbb{P}(\forall j \in \mathbb{N}^*, \tau_j < \infty) = 1$? En remarquant que $\tau_n \geq n-1$ et que $\frac{\tau_n}{n} = \frac{n-1}{n} \times \frac{\tau_n}{N_{\tau_n}}$, donner le comportement presque sûr de $\frac{\tau_n}{n}$ lorsque $n \to \infty$.

19. Soient $f, h : \bigcup_{n\in\mathbb{N}^*} \{\{n\} \times E^n\} \to \mathbb{R}$, mesurables bornées. Montrer que pour $k_1 \in \mathbb{N}$, $k_2, k_3 \in \mathbb{N}^*$,

$$\mathbb{E}\left[1_{\{\tau_1=k_1,\tau_2-\tau_1=k_2,\tau_3-\tau_2=k_3\}}f(\tau_2-\tau_1, X_{\tau_1+1},\ldots,X_{\tau_2})h(\tau_3-\tau_2, X_{\tau_2+1},\ldots,X_{\tau_3})\right]$$
$$= \mathbb{P}(\tau_1 = k_1)\int_{E^{k_2}} f(k_2,x_1,\ldots,x_{k_2})\nu(dx_1)\prod_{j=1}^{k_2-1}(1-\rho(x_j))R(x_j,dx_{j+1})\rho(x_{k_2})$$
$$\times \int_{E^{k_3}} h(k_3,x_1,\ldots,x_{k_3})\nu(dx_1)\prod_{j=1}^{k_3-1}(1-\rho(x_j))R(x_j,dx_{j+1})\rho(x_{k_3}).$$

En déduire que τ_1, $(\tau_2-\tau_1, \sum_{k=\tau_1+1}^{\tau_2}(f(X_k)-\pi(f)))$ et $(\tau_3-\tau_2, \sum_{k=\tau_2+1}^{\tau_3}(f(X_k)-\pi(f)))$ sont indépendantes avec les deux dernières indentiquement distribuées suivant une loi qui ne dépend pas de la loi de la condition initiale (X_0, Z_0).

En posant $Y_j = \sum_{k=\tau_j+1}^{\tau_{j+1}}(f(X_k) - \pi(f))$ pour $j \geq 1$, on montre plus généralement que les variables aléatoires $((\tau_{j+1} - \tau_j), Y_j)_{j\geq 1}$ sont indépendantes et identiquement distribuées. On supposera que les Y_j sont de carré intégrable : $\mathbb{E}[Y_1^2] < \infty$. On pose $S_0 = 0$ et pour $n \in \mathbb{N}^*$, $S_n = \sum_{j=1}^{n} Y_j$. On suppose que $\sup_{x\in E} \frac{|f(x)|}{1+\sqrt{V(x)}} < \infty$.

20. À l'aide de la question 18, montrer que $\mathbb{E}[\tau_2 - \tau_1] = \frac{1}{\pi(\rho)}$.

21. Vérifier que

$$\frac{S_n}{n} = \frac{\tau_{n+1}}{N_{\tau_{n+1}}} \times \frac{1+\tau_{n+1}}{\tau_{n+1}} \times \frac{1}{1+\tau_{n+1}} \sum_{k=0}^{\tau_{n+1}}(f(X_k) - \pi(f)) - \frac{1}{n}\sum_{k=0}^{\tau_1}(f(X_k) - \pi(f)),$$

et en déduire que $\frac{S_n}{n}$ tend presque sûrement vers 0 lorsque $n \to \infty$. Que peut-on en conclure pour $\mathbb{E}[Y_1]$?

22. Vérifier que

$$\frac{1}{\sqrt{n}}\sum_{k=0}^{n-1}(f(X_k) - \pi(f)) = \frac{S_{N_n}}{\sqrt{n}} + \frac{1}{\sqrt{n}}\sum_{k=0}^{\tau_1}(f(X_k) - \pi(f)) - \frac{1}{\sqrt{n}}\sum_{k=n}^{\tau_{N_n+1}}(f(X_k) - \pi(f)).$$

Quel est le comportement de $\frac{1}{\sqrt{n}}\sum_{k=0}^{\tau_1}(f(X_k) - \pi(f))$ lorsque $n \to \infty$?

23. En admettant que $\frac{1}{\sqrt{n}}\sum_{k=n}^{\tau_{N_n+1}}(f(X_k) - \pi(f))$ converge en probabilité vers 0 lorsque $n \to \infty$, résultat qui fait l'objet de la question suivante, conclure que $\sigma^2(f) = \pi(\rho)\mathrm{Var}(Y_1)$. En déduire un estimateur de $\sigma^2(f)$ calculé à partir de $((X_k, Z_k))_{0\leq k\leq n-1}$ et qui converge lorsque $n \to \infty$. Donner un intervalle de confiance à 95% pour $\pi(f)$ toujours calculé à partir de $((X_k, Z_k))_{0\leq k\leq n-1}$.

24. Soit $\varepsilon > 0$. Montrer que

$$\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{k=n}^{\tau_{N_n+1}}|f(X_k) - \pi(f)| \geq \varepsilon\right) \leq \mathbb{P}(\tau_1 \geq n)$$
$$+\sum_{j=0}^{n-1}\mathbb{P}\left(Z_{j+1} = 0, Z_{j+2} = 0, \ldots, Z_{n-1} = 0, \sum_{k=j+1}^{\tau_{N_n+1}}|f(X_k) - \pi(f)| \geq \varepsilon\sqrt{n}\middle| Z_j = 1\right)$$

À l'aide de la propriété de Markov, vérifier que le terme d'indice j dans la somme vaut $\mathbb{E}_j(1_{\{\tau_2-\tau_1\geq n-j\}}1_{\{\sum_{k=\tau_1+1}^{\tau_2}|f(X_k)-\pi(f)|\geq\varepsilon\sqrt{n}\}}|\tau_1 = 0)$ où l'indice j du signe espérance signifie que la condition initiale (X_0, Z_0) est choisie distribuée suivant $\tilde{\mu}_j$. En déduire, à l'aide de la question 19, que

$$\mathbb{P}\left(\frac{1}{\sqrt{n}}\sum_{k=n}^{\tau_{N_n+1}}|f(X_k) - \pi(f)| \geq \varepsilon\right) \leq \mathbb{P}(\tau_1 \geq n)+\mathbb{E}\left((\tau_2 - \tau_1)1_{\{\sum_{k=\tau_1+1}^{\tau_2}|f(X_k)-\pi(f)|\geq\varepsilon\sqrt{n}\}}\right).$$

Conclure que $\frac{1}{\sqrt{n}}\sum_{k=n}^{\tau_{N_n+1}}(f(X_k) - \pi(f))$ converge en probabilité vers 0 lorsque $n \to \infty$.