

TD 3 : CHÂÎNES DE MARKOV, ALGORITHME DE METROPOLIS

EXERCICE 1 (Metropolis vs Rejet)

On considère la densité f sur $\mathbb{R}^2$ définie par

$$f(u, v) \propto (\cos u)^2 (\sin v)^2 e^{-\frac{u^2+v^2}{20}}.$$

1. On veut simuler selon la densité f en utilisant l'algorithme de Metropolis. Partant de $x = (u, v)$, on considère le noyau de proposition $Q(x, dx')$ défini de sorte que sa densité $q(x, \cdot)$ corresponde à celle de la loi $\mathcal{N}(x, \sigma^2 I_2)$, où I_2 est la matrice identité de taille 2 et $\sigma > 0$ est un paramètre de réglage de l'algorithme (autrement dit, $Q(x, \cdot) = \mathcal{N}(x, \sigma^2 I_2)$). Expliciter le rapport de Metropolis $\rho(x, x') = \rho((u, v), (u', v'))$.
2. Proposer une méthode de rejet pour simuler suivant la loi de densité f à partir d'une loi instrumentale gaussienne.

EXERCICE 2 (Convergence en variation totale)

Soit $E = [0, 1]$ muni de sa tribu borélienne $\mathcal{B}$ et le noyau P défini comme suit : s'il existe $m \in \mathbb{N}^*$ tel que $x = 1/m$, alors

$$P(x, \cdot) = x^2 \text{Unif}([0, 1]) + (1 - x^2) \delta_{1/(m+1)},$$

sinon $P(x, \cdot) = \text{Unif}([0, 1])$. Soit (X_n) une chaîne de noyau P .

1. Question préliminaire : on note λ la mesure de Lebesgue sur $[0, 1]$, f une densité, $p \in [0, 1]$ et $x_0 \in]0, 1[$. Montrer que si $\mu = f \cdot \lambda$ et $\nu = p \delta_{x_0} + (1 - p)(f \cdot \lambda)$, alors $\|\mu - \nu\| = p$.
2. Montrer que λ est invariante pour P .
3. Montrer que pour tout couple (x, B) de $[0, 1] \times \mathcal{B}$ tel que $\lambda(B) > 0$, on a $P(x, B) > 0$. En déduire que P est λ -irréductible et apériodique.
4. Soit $X_0 = x \in [0, 1]$ tel qu'il n'existe pas $m \geq 2$ pour lequel $x = 1/m$. Préciser la loi de X_n pour tout n .
5. Soit $X_0 = 1/m$ avec $m \geq 2$. Préciser la loi de X_n pour tout n .
6. Pour quelles valeurs initiales $X_0 = x$ a-t-on convergence de $\|P^n(x, \cdot) - \lambda\|$ vers 0 ? Ceci est-il en accord avec le théorème de convergence en variation totale vu en cours ?

EXERCICE 3 (Metropolis indépendant et uniforme ergodicité)

On se place sur $\mathbb{R}$. Avec l'abus de notation usuel $\pi(dx) = \pi(x)dx$ pour la loi cible, on applique un algorithme de Metropolis indépendant où le noyau de proposition est $Q(x, dy) = q(y)dy$. On suppose que les densités π et q sont strictement positives sur $\mathbb{R}$ et qu'il existe une constante M telle que $\pi(x) \leq Mq(x)$.

1. Montrer que $M \geq 1$. À quelle condition a-t-on $M = 1$? Dans toute la suite, on exclut ce cas d'égalité.
2. Expliciter le noyau de transition de Metropolis sous la forme $P(x, dy) = p(x, y)dy + r(x)\delta_x(dy)$ avec $p(x, y) \geq \frac{\pi(y)}{M}$ pour tout couple (x, y) et $r(x) \geq 0$ pour tout x .

- Montrer que pour tout x et toute fonction $\varphi : \mathbb{R} \rightarrow \mathbb{R}_+$ borélienne, on a $(P\varphi)(x) \geq \frac{1}{M}\pi(\varphi)$. On résume ceci par l'inégalité entre mesures $P(x, \cdot) \geq \frac{1}{M}\pi(\cdot)$ et on dit que le noyau P vérifie une condition de Doeblin.
- Justifier que l'application $R : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \rightarrow \mathbb{R}$ définie par

$$R(x, dy) = \frac{M}{M-1} \left(P(x, dy) - \frac{1}{M}\pi(dy) \right) \iff R(x, B) = \frac{M}{M-1} \left(P(x, B) - \frac{1}{M}\pi(B) \right)$$

est un noyau de transition.

- Montrer que, pour toute mesure de probabilité μ , on a $\mu P = \frac{1}{M}\pi + (1 - \frac{1}{M})\mu R$. En déduire que pour tout entier naturel n

$$\delta_x P^n - \pi = \left(1 - \frac{1}{M}\right)^n (\delta_x R^n - \pi).$$

- On rappelle que $\|\mu - \nu\| = \sup_{\varphi: E \rightarrow [0,1]} |\mu(\varphi) - \nu(\varphi)|$. Déduire de la question précédente que la chaîne (X_n) est uniformément ergodique, c'est-à-dire que pour tout réel x et tout entier n on a

$$\|P^n(x, \cdot) - \pi\| \leq \left(1 - \frac{1}{M}\right)^n.$$

- Que vous rappelle la condition $\pi(x) \leq Mq(x)$?

EXERCICE 4 (Metropolis–Adjusted Langevin Algorithm (MALA))

La loi cible considérée dans cet exercice est le mélange équiprobable de deux lois normales centrées respectivement en $m = 3$ et $-m = -3$ et de même écart-type égal à 1.

- Rappeler la densité f correspondante. Quelle est sa moyenne? Quel est son écart-type?
- Si on applique l'algorithme de Metropolis avec le noyau instrumental $Q(x, \cdot) = \mathcal{N}(x, 1)$ et partant de $X_0 = m$, quel va être le problème?
- On considère une densité cible $f > 0$ sur $\mathbb{R}$ et un noyau de proposition qui n'est pas symétrique : sachant $X_n = x$, la proposition est

$$Y_{n+1} = x + \frac{\sigma^2}{2} \times \frac{f'(x)}{f(x)} + \sigma W_{n+1},$$

où W_{n+1} suit une loi normale centrée réduite et $\sigma > 0$ est un paramètre de réglage.

- Quelles transitions sont favorisées par ce noyau instrumental?
- Expliciter le rapport de Metropolis $\rho(x, y)$.

EXERCICE 5 (Noyau gaussien)

Soit $-1 < r < 1$ un réel fixé et la chaîne de Markov (X_n) définie par $X_0 = x$ et la récurrence aléatoire $X_{n+1} = rX_n + \sqrt{1-r^2}W_{n+1}$ où $(W_n)_{n \geq 1}$ est une suite de variables i.i.d. gaussiennes standards.

- Montrer que la loi gaussienne standard, notée π , est l'unique loi stationnaire de la chaîne. Indication : on pourra passer par les fonctions caractéristiques et se souvenir que celle de la loi gaussienne standard est $\Phi_{\mathcal{N}(0,1)}(t) = e^{-t^2/2}$.
- Quelle est la loi μ_n de X_n ?

3. Si $\mu = f \cdot \lambda$ et $\nu = g \cdot \lambda$ sont deux lois absolument continues par rapport à la mesure de Lebesgue sur $\mathbb{R}$, on définit leur distance de Hellinger par

$$H(\mu, \nu) = \left\| \sqrt{f} - \sqrt{g} \right\|_2 = \left(\int_{\mathbb{R}} \left(\sqrt{f(x)} - \sqrt{g(x)} \right)^2 dx \right)^{\frac{1}{2}}.$$

On note comme d'habitude $\|\mu - \nu\|$ la distance en variation totale. Montrer que $\|\mu - \nu\| \leq H(\mu, \nu)$ et que $H(\mu, \nu)^2 \leq 2\|\mu - \nu\|$.

4. Montrer que si $\mu = \mathcal{N}(m_1, \sigma_1^2)$ et $\nu = \mathcal{N}(m_2, \sigma_2^2)$ avec σ_1 et σ_2 strictement positifs, alors

$$H(\mu, \nu)^2 = 2 \left\{ 1 - \sqrt{2 \frac{\sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2}} \exp \left(-\frac{(m_1 - m_2)^2}{4(\sigma_1^2 + \sigma_2^2)} \right) \right\}.$$

5. En déduire un encadrement de $\|\mu_n - \pi\|$ en fonction de x et r . On donnera des équivalents des deux bornes lorsque n tend vers l'infini, à x fixé.

EXERCICE 6 ((*) Metropolis avec propositions locales)

On considère une densité cible $\pi > 0$ sur $\mathbb{R}$ et un noyau de proposition $Q(x, dy) = q(y - x)dy$ avec q une densité paire sur $\mathbb{R}$ et supérieure ou égale à $\epsilon > 0$ sur $[-1, 1]$.

- On note P le noyau de Metropolis, x un réel et B un borélien. Rappeler ce que vaut $P(x, B)$.
- Montrer que pour tout borélien B tel que $\lambda(B \cap [x - 1; x + 1]) > 0$, on a $P(x, B) > 0$.
- Soit B un borélien non négligeable, i.e. $\lambda(B) > 0$. Justifier qu'il existe un entier n et une séquence d'intervalles $(I_1, \dots, I_n)$ non négligeables tels que $P(x, I_1) > 0$, pour tout k et tout $y \in I_k$ on ait $P(y, I_{k+1}) > 0$ et pour tout $y \in I_n$ on ait $P(y, B) > 0$.
- Soient $\varphi_1, \dots, \varphi_n$ des fonctions boréliennes bornées et (X_n) une chaîne de Markov de noyau de transition P . Montrer que

$$\mathbb{E}_x[\varphi_1(X_1) \dots \varphi_n(X_n)] = \int P(x, dx_1) \varphi_1(x_1) \int P(x_1, dx_2) \varphi_2(x_2) \dots \int P(x_{n-1}, dx_n) \varphi_n(x_n).$$

- En déduire que P est irréductible.
- Que vaut $P(x, \{x\})$? Justifier qu'on peut trouver un réel x tel que $P(x, \{x\}) > 0$.
- En déduire que la chaîne est apériodique.
- Application : on considère la chaîne de Metropolis (X_n) avec le noyau de proposition correspondant à $Y_{n+1} = X_n + U_{n+1}$ avec $U_{n+1} \sim \text{Unif}([-1, 1])$. La convergence en variation totale et le théorème ergodique sont-ils vérifiés?

EXERCICE 7 ((†) Modèle d'Ising)

Soit N un entier naturel fixé. On considère le carré $\mathcal{C} = \{(i, j) \in \{1, \dots, N\}^2\}$, chaque site (i, j) ne pouvant prendre que deux valeurs $S_{ij} = \pm 1$ (son spin). À une configuration de spins $S = (S_{ij})_{1 \leq i, j \leq N}$ donnée est alors associée l'énergie $V(S)$ définie par

$$V(S) = - \sum_{(i,j) \sim (k,\ell)} S_{ij} S_{k\ell} \quad \text{avec} \quad (i, j) \sim (k, \ell) \Leftrightarrow |i - k| + |j - \ell| = 1.$$

Dans la somme définissant V interviennent donc toutes les paires de sites voisins dans la grille $\mathcal{C}$.

1. Quelles sont les configurations minimisant l'énergie ?
2. Pour tenir compte de la température $T > 0$, on associe à V la mesure de Gibbs

$$\mu_T(S) = \frac{1}{Z_T} \exp\left(-\frac{1}{T}V(S)\right),$$

où Z_T est la constante de normalisation. Que dire de cette mesure de probabilité dans les deux situations extrêmes où $T \rightarrow 0$ et où $T \rightarrow +\infty$?

3. On considère T fixé et on veut simuler S suivant la mesure μ_T . Donner en fonction de N le nombre de configurations $S = (S_{ij})_{1 \leq i, j \leq N}$ possibles.
4. On applique la méthode de Metropolis pour simuler suivant μ_T . Pour cela, partant d'une configuration $S = (S_{ij})_{1 \leq i, j \leq N}$, on commence par tirer un site (i_0, j_0) au hasard (uniforme) et on propose de changer $S_{i_0 j_0}$ en $-S_{i_0 j_0}$. On note $S^{(i_0, j_0)}$ cette nouvelle configuration où seul un spin a changé de signe. Que vaut $q(S, S^{(i_0, j_0)})$? Et $q(S^{(i_0, j_0)}, S)$? Calculer le rapport de Metropolis $\rho(S, S^{(i_0, j_0)})$.
5. Expliquer brièvement pourquoi la chaîne de Markov ainsi construite est irréductible et apériodique.