Examen du cours de M2 :
Méthodes numériques probabilistes

Mardi 5 janvier 2021, 13h00-16h00.

Seuls les notes de cours et le polycopié sont autorisés. Certaines questions sont facultatives : leur résolution éventuelle apportera des "points bonus".

Problème 1 : variance asymptotique pour des chaînes de Markov non réversibles

Soit π une mesure de probabilité strictement positive sur l'ensemble de cardinal fini $E = \{1, \dots, \mathsf{m}\}$. On note $I \in \mathbb{R}^{\mathsf{m}\times\mathsf{m}}$ la matrice identité.

Une formule pour la variance asymptotique. On considère une matrice stochastique $P \in \mathbb{R}^{\mathsf{m}\times\mathsf{m}}$ irréductible et apériodique, de mesure invariante π. On rappelle le théorème central limite énoncé en cours : pour toute fonction[1] $f : E \to \mathbb{R}$,

$$\lim_{n\to\infty} \sqrt{n} \left(\frac{1}{n} \sum_{k=0}^{n-1} f(X_k^P) - \pi f \right) = \mathcal{N}\left(0, \sigma^2(f, P)\right) \text{ en loi}$$

où $(X_k^P)_{k\geq 0}$ désigne une chaîne de Markov de matrice de transition P et

$$\sigma^2(f, P) = \mathrm{Var}_\pi(f(X_0^P)) + 2 \sum_{k=1}^{\infty} \mathrm{Cov}_\pi(f(X_0^P), f(X_k^P))$$

est la variance asymptotique. On introduit $\pi^\perp = \{g \in \mathbb{R}^{\mathsf{m}},\, \pi g = 0\}$ et $\tilde{f}(x) = f(x) - \pi f$.

1. Vérifier que $\pi^\perp$ est laissé invariant par P.
2. Montrer que le problème suivant : trouver $g \in \pi^\perp$ tel que

   $$\forall x \in E, \quad (I - P)g(x) = \tilde{f}(x)$$

   admet une unique solution.
3. Montrer que pour tout $x \in E$,

   $$g(x) = \sum_{n=0}^{\infty} \mathbb{E}_x(\tilde{f}(X_n^P)).$$

4. En déduire que

   $$\sigma^2(f, P) = 2\langle g, \tilde{f}\rangle_\pi - \langle \tilde{f}, \tilde{f}\rangle_\pi$$

   où on utilise le produit scalaire scalaire, pour toutes fonctions $u : E \to \mathbb{R}$ et $v : E \to \mathbb{R}$,

   $$\langle u, v\rangle_\pi = \sum_{x\in E} u(x)\, v(x)\, \pi(x).$$

1. Dans toute la suite, on identifiera une fonction $f : E \to \mathbb{R}$ avec un vecteur colonne de $\mathbb{R}^{\mathsf{m}}$ et de même une fonction $f : E \to \mathbb{C}$ avec un vecteur colonne de $\mathbb{C}^{\mathsf{m}}$. En particulier, Pf désigne soit un vecteur de dimension m, soit de manière équivalente une fonction scalaire définie sur E.

Chaînes réversibles et non-réversibles. On considère une matrice de transition $S \in \mathbb{R}^{\mathsf{m}\times\mathsf{m}}$ d'une chaîne de Markov irréductible, apériodique et réversible par rapport à la mesure π :

$$\forall (x,y) \in E \times E, \quad \pi(x)S(x,y) = \pi(y)S(y,x).$$

Soit une matrice $A \in \mathbb{R}^{\mathsf{m}\times\mathsf{m}}$ qui vérifie les trois propriétés suivantes :

$$\forall (x,y) \in E \times E, \quad \pi(x)A(x,y) = -\pi(y)A(y,x)\,, \tag{1}$$

$$\forall (x,y) \in E \times E, \quad \begin{cases} A(x,y) = 0 \text{ si } S(x,y) = 0, \\ |A(x,y)| < S(x,y) \text{ si } S(x,y) \neq 0, \end{cases} \tag{2}$$

$$\forall x \in E, \quad \sum_{y \in E} A(x,y) = 0. \tag{3}$$

On introduit la matrice $P \in \mathbb{R}^{\mathsf{m}\times\mathsf{m}}$ définie par

$$\forall (x,y) \in E \times E, \quad P(x,y) = S(x,y) + A(x,y).$$

5. Montrer que pour tout $(x,y) \in E \times E$, $S(x,y) > 0 \iff P(x,y) > 0$.
6. Montrer que P est une matrice stochastique apériodique, qui admet π comme unique mesure invariante.

Nous avons besoin pour la suite d'utiliser le théorème spectral, dont on rappelle un énoncé adapté à notre contexte. On commence par étendre le produit scalaire $\langle \cdot, \cdot \rangle_\pi$ à des fonctions à valeurs complexes : pour toutes fonctions $u : E \to \mathbb{C}$ et $v : E \to \mathbb{C}$,

$$\langle u, v \rangle_\pi = \sum_{x \in E} \overline{u(x)}\, v(x)\, \pi(x),$$

où $\bar{z}$ désigne le complexe conjugué de $z \in \mathbb{C}$. Pour $M \in \mathbb{C}^{\mathsf{m}\times\mathsf{m}}$ une matrice à valeurs complexes, on note $M^* \in \mathbb{C}^{\mathsf{m}\times\mathsf{m}}$ son adjoint défini par : pour toutes fonctions $u : E \to \mathbb{C}$ et $v : E \to \mathbb{C}$,

$$\langle M^* u, v \rangle_\pi = \langle u, Mv \rangle_\pi.$$

D'après le théorème spectral, si M est hermitienne (i.e. $M^* = M$), alors, M est diagonalisable dans une base orthonormée (pour le produit scalaire hermitien $\langle \cdot, \cdot \rangle_\pi$) de $\mathbb{C}^{\mathsf{m}}$ avec des valeurs propres réelles.

7. Montrer que A est une matrice diagonalisable, et que ses valeurs propres sont purement imaginaires et de module strictement inférieur à 1.
8. *(Question facultative)* Que dire des résultats des trois questions précédentes si on remplace (2) par l'hypothèse plus faible $\forall (x,y) \in E \times E$, $|A(x,y)| \leq S(x,y)$?

L'objectif des questions qui suivent est d'étudier en quoi l'introduction de la matrice A permet d'améliorer les performances en terme de variance asymptotique d'un algorithme de type *Markov Chain Monte Carlo*, consistant à échantillonner π en générant une chaîne de Markov de matrice de transition $P = S + A$ plutôt que S.

On rappelle que les matrices S, A, P envoient $\pi^\perp$ dans $\pi^\perp$. Ces matrices sont vues dans toute la suite comme des endomorphismes de $\pi^\perp$, qui est un espace vectoriel de dimension $\mathsf{m} - 1$.

9. L'opérateur $S : \pi^\perp \to \pi^\perp$ est symétrique pour le produit scalaire $\langle \cdot, \cdot \rangle_\pi$ et donc diagonalisable : on note $(d_i)_{1 \le i \le \mathsf{m}-1} \in \mathbb{R}^{\mathsf{m}-1}$ les valeurs propres et $(f_i : E \to \mathbb{R})_{1 \le i \le \mathsf{m}-1}$ les vecteurs propres associés, qui forment une base orthonormée de $\pi^\perp$. Pour tous $\alpha \in \mathbb{R}$, on définit dans la suite l'opérateur $(I - S)^\alpha : \pi^\perp \to \pi^\perp$ par

$$\forall i \in \{1, \dots, \mathsf{m} - 1\}, \quad (I - S)^\alpha f_i = (1 - d_i)^\alpha f_i.$$

Vérifier que pour tout $\alpha \in \mathbb{R}$, $(I - S)^\alpha$ est un opérateur symétrique pour le produit scalaire $\langle \cdot, \cdot \rangle_\pi$, et que $(I - S)^\alpha (I - S)^\beta = (I - S)^{\alpha+\beta}$, pour tout $\alpha, \beta \in \mathbb{R}$.

10. Soit $Q = (I - S)^{-1/2} A (I - S)^{-1/2}$. Vérifier que $I - Q$ est inversible. Montrer que pour tout $g \in \pi^\perp$,

$$\langle (I - Q)^{-1} g, g \rangle_\pi \le \langle g, g \rangle_\pi.$$

Dans quel cas a-t-on égalité ?

11. Montrer que pour tout $f : E \to \mathbb{R}$,

$$\langle (I - P)^{-1} \tilde{f}, \tilde{f} \rangle_\pi = \langle (I - Q)^{-1} (I - S)^{-1/2} \tilde{f}, (I - S)^{-1/2} \tilde{f} \rangle_\pi.$$

12. Déduire des questions précédentes que pour toute fonction $f : E \to \mathbb{R}$,

$$\sigma^2(f, P) \le \sigma^2(f, S).$$

Dans quel cas a-t-on égalité ?

Construction de A. Pour ces deux dernières questions, on se demande comment construire, à partir de S, une matrice A qui satisfait les trois propriétés (1)-(2)-(3). Pour cela, on considère un cycle pour S, i.e. un chemin dans E de probabilité strictement positive pour S partant d'un état et revenant à ce même état, et visitant au moins un autre état.

13. Pour tout $x \in E$, justifier l'existence d'un cycle pour S partant de x et revenant en x.

Quitte à renuméroter les états, on a donc :

$$\text{pour un } \ell \ge 2,\ S(1,2) S(2,3) \dots S(\ell, 1) > 0.$$

On considère alors

$$\alpha = \min \left(\pi(1) S(1,2),\ \pi(2) S(2,3), \dots, \pi(\ell-1) S(\ell-1, \ell),\ \pi(\ell) S(\ell, 1) \right).$$

Pour $t \in (-\alpha, \alpha)$, on définit A_t par :

$$A_t(x,y) = \begin{cases} \frac{t}{\pi(x)} & \text{si } (x,y) = (\ell, 1) \text{ ou } (x,y) = (i, i+1) \text{ pour un } i \in \{1, \dots, \ell - 1\}, \\ -\frac{t}{\pi(x)} & \text{si } (x,y) = (1, \ell) \text{ ou } (x,y) = (i, i-1) \text{ pour un } i \in \{2, \dots, \ell\}, \\ 0 & \text{sinon.} \end{cases}$$

14. Montrer que la matrice A_t construite ci-dessus satisfait (1)-(2)-(3).

Problème 2 : une méthode d'échantillonnage d'importance pour des diffusions

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité muni d'une filtration $(\mathcal{F}_s)_{s\geq 0}$ et d'un $(\mathcal{F}_s)_{s\geq 0}$-mouvement brownien d-dimensionnel $(B_s)_{s\geq 0}$. Dans tout ce problème, on fixe $T > 0$ et on se donne des fonctions $b : \mathbb{R}^n \to \mathbb{R}^n$ et $\sigma : \mathbb{R}^n \to \mathbb{R}^{n\times d}$ régulières, bornées et globalement lipschitziennes. Pour tous $t \in [0, T]$ et $x \in \mathbb{R}^n$, on note $(X_s^{t,x})_{s\in[t,T]}$ l'unique solution (forte) de l'équation différentielle stochastique (EDS)

$$\begin{cases} \mathrm{d}X_s^{t,x} = b(X_s^{t,x})\mathrm{d}s + \sigma(X_s^{t,x})\mathrm{d}B_s, \qquad s \in [t, T], \\ X_t^{t,x} = x. \end{cases} \tag{4}$$

Pour $f : \mathbb{R}^n \to \mathbb{R}$ une fonction régulière bornée, le but de ce problème est d'étudier une méthode d'échantillonnage d'importance pour le calcul de la fonction $u : [0, T] \times \mathbb{R}^n \to \mathbb{R}$ définie par

$$u(t, x) = \mathbb{E}\left[f(X_T^{t,x})\right]. \tag{5}$$

Nous allons pour cela introduire un processus de diffusion $(Y_s^{t,x})_{s\in[t,T]}$, dépendant d'une fonction ψ qui sera précisée plus bas, tel que pour tous (t, x) et ψ il existe une variable aléatoire $W_\psi^{t,x}$, qui est une fonction déterministe de la trajectoire $(Y_s^{t,x})_{s\in[t,T]}$, et qui vérifie

$$u(t, x) = \mathbb{E}\left[f(Y_T^{t,x})W_\psi^{t,x}\right]. \tag{6}$$

Nous montrerons ensuite qu'il est possible de choisir ψ de manière à ce que $f(Y_T^{t,x})W_\psi^{t,x}$ soit de variance nulle, ce qui fournit donc un estimateur optimal, de manière analogue au résultat du Lemme 2.3.6 vu en cours.

Si vous ne vous sentez pas à l'aise avec la manipulation des opérateurs différentiels dans $\mathbb{R}^n$, vous pouvez supposer $n = d = 1$ dans tout le problème afin de ne pas perdre de temps dans les calculs. Dans ce cas, merci de le signaler au début de votre rédaction. La notation ne pénalisera pas ce choix.

Remarques préliminaires sur u. Pour tout $x \in \mathbb{R}^n$, on pose $a(x) = \sigma(x)\sigma^\top(x) \in \mathbb{R}^{n\times n}$ et on note, pour toute fonction régulière $\Phi : \mathbb{R}^n \to \mathbb{R}$,

$$L\Phi(x) = \frac{1}{2}a(x) : \nabla^2\Phi(x) + b(x) \cdot \nabla\Phi(x)$$

l'opérateur différentiel associé à l'EDS (4). Dans cette expression, $\nabla^2\Phi(x) \in \mathbb{R}^{n\times n}$ désigne la matrice hessienne de Φ, et le symbole $:$ désigne le produit de Frobenius entre deux matrices :

$$a(x) : \nabla^2\Phi(x) = \mathrm{tr}\left(a(x)^\top \nabla^2\Phi(x)\right) = \sum_{i,j=1}^n a_{i,j}(x)\frac{\partial^2\Phi}{\partial x_i \partial x_j}(x).$$

On dit que la fonction $u : [0, T] \times \mathbb{R}^n \to \mathbb{R}$ est de classe $C_b^{1,2}$ si elle est bornée et ses dérivées $\frac{\partial u}{\partial t}$, ∇u et $\nabla^2 u$ existent, sont continues et bornées sur $[0, T] \times \mathbb{R}^n$.

On suppose que les fonctions a, b et f sont telles que u est de classe $C_b^{1,2}$ et vérifie

$$\begin{cases} \dfrac{\partial u}{\partial t}(t, x) + Lu(t, x) = 0, \qquad t \in [0, T], \quad x \in \mathbb{R}^n, \\ u(T, x) = f(x). \end{cases} \tag{7}$$

1. Quel nom porte alors la représentation probabiliste (5) de la solution u de l'équations aux dérivées partielles (7) ?
2. Décrire brièvement le principe du calcul de $u(t,x)$ par la méthode de Monte-Carlo.
3. Vérifier que pour toutes fonctions régulières $\Phi, \Psi : \mathbb{R}^n \to \mathbb{R}$,

   $$L(\Phi\Psi)(x) = \Psi(x)L\Phi(x) + \Phi(x)L\Psi(x) + \Gamma(\Phi,\Psi)(x),$$

   avec $\Gamma(\Phi,\Psi) := \sigma^\top \nabla\Phi \cdot \sigma^\top\nabla\Psi$.

Étude du processus Y. Soit $\psi : [0,T]\times\mathbb{R}^n \to \mathbb{R}$ une fonction de classe $C_b^{1,2}$. Pour tous $t \in [0,T]$ et $x \in \mathbb{R}^n$, on note $(Y_s^{t,x})_{s\in[t,T]}$ l'unique solution (forte) de l'EDS

$$\begin{cases} \mathrm{d}Y_s^{t,x} = \big(b(Y_s^{t,x}) + a(Y_s^{t,x})\nabla\psi(s,Y_s^{t,x})\big)\,\mathrm{d}s + \sigma(Y_s^{t,x})\mathrm{d}B_s, \qquad s \in [t,T], \\ \;\;Y_t^{t,x} = x, \end{cases} \tag{8}$$

et on note, pour toute fonction régulière $\Phi : \mathbb{R}^n \to \mathbb{R}$,

$$M_t^\psi \Phi(x) := L\Phi(x) + a(x)\nabla\psi(t,x)\cdot\nabla\Phi(x)$$

l'opérateur différentiel associé. Pour toutes fonctions $g : \mathbb{R}^n \to \mathbb{R}$ et $k : [0,T]\times\mathbb{R}^n \to \mathbb{R}$ régulières et bornées, on admet qu'il existe une fonction $v : [0,T]\times\mathbb{R}^n \to \mathbb{R}$ de classe $C_b^{1,2}$ qui vérifie

$$\begin{cases} \dfrac{\partial v}{\partial t}(t,x) + M_t^\psi v(t,x) = k(t,x)v(t,x), \qquad t\in[0,T], \quad x \in \mathbb{R}^n, \\ \qquad\qquad v(T,x) = g(x). \end{cases} \tag{9}$$

4. Justifier que, pour $0 \le t \le s \le T$,

   $$\mathrm{d}v(s,Y_s^{t,x}) = k(s,Y_s^{t,x})v(s,Y_s^{t,x})\mathrm{d}s + \sigma^\top(Y_s^{t,x})\nabla v(s,Y_s^{t,x})\cdot \mathrm{d}B_s.$$

5. Pour $0 \le t < s < T$, on pose

   $$Z_s^{t,x} = \exp\left(-\int_{r=t}^s k(r,Y_r^{t,x})\mathrm{d}r\right).$$

   Montrer que

   $$g(Y_T^{t,x})Z_T^{t,x} = v(t,x) + \int_{r=t}^T Z_s^{t,x}\sigma^\top(Y_s^{t,x})\nabla v(s,Y_s^{t,x})\cdot\mathrm{d}B_s.$$

6. En déduire la représentation probabiliste de v :

   $$v(t,x) = \mathbb{E}\left[g(Y_T^{t,x})\exp\left(-\int_{s=t}^T k(s,Y_s^{t,x})\mathrm{d}s\right)\right].$$

Définition de u^ψ. Pour toute fonction $\psi : [0,T]\times\mathbb{R}^n \to \mathbb{R}$ de classe $C_b^{1,2}$, on définit la fonction $u^\psi : [0,T]\times\mathbb{R}^n\to\mathbb{R}$ par

$$u^\psi(t,x) := \mathrm{e}^{\psi(t,x)}v^\psi(t,x),$$

avec

$$v^\psi(t,x) := \mathbb{E}\left[g^\psi(Y_T^{t,x})\exp\left(-\int_{s=t}^T k^\psi(s,Y_s^{t,x})\mathrm{d}s\right)\right],$$

et

$$g^\psi(y) = f(y)\mathrm{e}^{-\psi(T,y)}, \qquad k^\psi(s,y) := -\frac{\partial\psi}{\partial t}(s,y) - L\psi(s,y) - \frac{1}{2}\left|\sigma^\top(y)\nabla\psi(s,y)\right|^2.$$

7. Montrer que
$$M_t^{\psi} v^{\psi} = \mathrm{e}^{-\psi} \left(L u^{\psi} - u^{\psi} L\psi - \frac{1}{2} u^{\psi} \left| \sigma^{\top} \nabla \psi \right|^2 \right).$$

8. En admettant que les fonctions g^{ψ} et k^{ψ} sont telles que le problème associé (9) admet une solution de classe $C_b^{1,2}$, montrer que u^{ψ} est de classe $C_b^{1,2}$ et vérifie
$$\frac{\partial u^{\psi}}{\partial t}(t,x) + L u^{\psi}(t,x) = 0, \qquad t \in [0,T], \quad x \in \mathbb{R}^n.$$

9. Que dire par ailleurs de $u^{\psi}(T,x)$?

10. Conclure en explicitant une variable aléatoire $W_{\psi}^{t,x}$, qui est une fonction déterministe de la trajectoire $(Y_s^{t,x})_{s\in[t,T]}$, telle que l'identité (6) est vraie.

Choix optimal de ψ. On déduit de la section précédente que $u(t,x)$ peut être évalué, par la méthode de Monte-Carlo, en simulant N trajectoires indépendantes $(Y_s^{t,x,i})_{s\in[t,T]}$, $1 \leq i \leq N$, de l'EDS (8), puis en calculant la moyenne empirique
$$\hat{u}_N^{\psi}(t,x) := \frac{1}{N} \sum_{i=1}^{N} f(Y_T^{t,x,i}) W_{\psi}^{t,x,i},$$
avec $W_{\psi}^{t,x,i}$ calculé à partir de la trajectoire $(Y_s^{t,x,i})_{s\in[t,T]}$.

11. On note $V^{\psi}(t,x) = \mathrm{Var}(f(Y_T^{t,x}) W_{\psi}^{t,x})$ et on remarque que sous nos hypothèses, cette quantité est finie. Pour $\alpha \in (0, 1/2)$, donner un intervalle $[\mathcal{I}_N^-, \mathcal{I}_N^+]$, dont les bornes dépendent de $\hat{u}_N^{\psi}(t,x)$ et $V^{\psi}(t,x)$, tel que
$$\lim_{N\to+\infty} \mathbb{P}\left(u(t,x) \in [\mathcal{I}_N^-, \mathcal{I}_N^+] \right) = 1 - \alpha.$$

12. On suppose désormais, et jusqu'à la fin du sujet, qu'il existe $c > 0$ tel que pour tout $x \in \mathbb{R}^n$, $f(x) \geq c$. Que peut-on en déduire sur u ?

13. On pose $\psi(t,x) = \ln u(t,x)$. Vérifier que ψ est de classe $C_b^{1,2}$, puis calculer $L\psi$ et en déduire la valeur de $W_{\psi}^{t,x}$.

14. Conclure.