Examen du cours de M2 : Méthodes numériques probabilistes

13 janvier 2010, 09h00 - 12h00.

Les notes de cours sont autorisées.

Problème 1 : Algorithme de Metropolis-Hastings

Remarque : les sections 2 et 3 peuvent être traitées indépendamment l'une de l'autre.

1 On désire échantillonner une mesure de probabilité $\pi : \mathcal{S} \to (0,1]$, sur un espace d'états dénombrable $\mathcal{S}$. On considère la chaîne de Markov suivante : X_0 est une variable aléatoire donnée et, pour X_n donné à l'itération n, X_{n+1} est construit de la façon suivante :

- La variable aléatoire X^1_{n+1} est tirée selon la loi $\mathbb{P}(X^1_{n+1} = y | X_n = x) = q(x,y)$ où q est un noyau de proposition symétrique :

$$q(x,y) = q(y,x)$$

  que l'on suppose irréductible.
- On tire U^1_n selon la loi uniforme sur $(0,1)$ et on pose
  - $X_{n+1} = X^1_{n+1}$ si $U^1_n \leq a(X_n, X^1_{n+1})$,
  - $X_{n+1} = X_n$ sinon,

  où

$$a(x,y) = \min\left(1, \frac{\pi(y)}{\pi(x)}\right).$$

Montrer que conditionnellement à X_n, X_{n+1} est distribué selon la loi $p_{\mathrm{MH}}(X_n, y)$, où

$$p_{\mathrm{MH}}(X_n, y) = q(X_n, y) a(X_n, y) 1_{X_n \neq y} + r(X_n) 1_{X_n = y}$$

où on précisera la fonction r. Montrer que π est une mesure réversible pour le noyau p_{MH}, c'est-à-dire que $\forall x, y \in \mathcal{S}$,

$$\pi(x) p_{\mathrm{MH}}(x,y) = \pi(y) p_{\mathrm{MH}}(y,x).$$

En déduire que π est une mesure invariante pour la chaîne de Markov $(X_n)_{n\geq 0}$. Rappeler pourquoi, p.s. $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \varphi(X_n) = \sum_{s \in \mathcal{S}} \varphi(s)\pi(s)$, où $\varphi : \mathcal{S} \to \mathbb{R}$ est une fonction test bornée.

2 On considère maintenant une modification de l'algorithme précédent, où X_{n+1} est "la première proposition de mouvement acceptée". On note $(\overline{X}_n)_{n\geq 0}$ cette nouvelle chaîne de Markov. Plus précisément, pour $\overline{X}_n$ donné à l'itération n, $\overline{X}_{n+1}$ est construit de la façon suivante :

$$\overline{X}_{n+1} = \overline{X}^T_{n+1},$$

où $(\overline{X}^k_{n+1})_{k\geq 1}$ sont i.i.d. selon la loi $\left(q(\overline{X}_n, y)\right)_{y\in\mathcal{S}}$, et

$$T = \inf\left\{k \geq 1,\, U^k_n \leq a(\overline{X}_n, \overline{X}^k_{n+1})\right\},$$

où les $(U_n^k)_{k\geq 1}$ sont i.i.d. de loi uniforme sur $(0,1)$.

2.a Conditionnellement à $\overline{X}_n$, quelle est la loi de la variable aléatoire T ? Calculer $\mathbb{E}(T|\overline{X}_n = x)$.

2.b Donner l'expression de la loi $(p(\overline{X}_n, y))_{y\in\mathcal{S}}$ de $\overline{X}_{n+1}$ conditionnellement à $\overline{X}_n$.

2.c Montrer qu'il existe une mesure réversible $(\overline{\pi}(x))_{x\in\mathcal{S}}$, que l'on précisera, pour la chaîne de Markov $(\overline{X}_n)_{n\geq 0}$.

2.d Sous quelle condition a-t-on $\overline{\pi} = \pi$? Montrer qu'une telle situation se produit par exemple dans le cas où $q(x,y) = 1/|\mathcal{S}|$ est la probabilité uniforme sur $\mathcal{S}$ et π est la probabilité uniforme sur un sous-ensemble $\mathcal{S}'$ de $\mathcal{S}$. Comme s'appelle un tel algorithme d'échantillonnage ?

3 On considère à nouveau l'algorithme de la première section.

3.a Soit $\varphi : \mathcal{S} \to \mathbb{R}$ une fonction test bornée. On propose de remplacer l'estimateur $I_N(\varphi) = \frac{1}{N}\sum_{n=1}^N \varphi(X_n)$ de la moyenne $I(\varphi) = \sum_{s\in\mathcal{S}} \varphi(s)\pi(s)$ par

$$I_N^{\mathrm{WR}}(\varphi) = \frac{1}{N}\sum_{n=0}^{N-1} a(X_n, X_{n+1}^1)\varphi(X_{n+1}^1) + (1 - a(X_n, X_{n+1}^1))\varphi(X_n),$$

appelé estimateur “waste recycling”. Montrer que $I_N^{\mathrm{WR}}(\varphi) = \dfrac{1}{N}\displaystyle\sum_{n=0}^{N-1} \varphi_{\mathrm{WR}}(X_n, X_{n+1}^1)$ où

$$\varphi_{\mathrm{WR}}(x, x^1) = \mathbb{E}\left(\varphi(X_1)\,|\,(X_0, X_1^1) = (x, x^1)\right),$$

et en déduire que $\mathbb{E}(\varphi_{\mathrm{WR}}(X_0, X_1^1)) = \mathbb{E}(\varphi(X_1))$. Montrer que $\mathbb{E}\left(I_N(\varphi)\right) = \mathbb{E}\left(I_N^{\mathrm{WR}}(\varphi)\right)$.

3.b On considère $Z_n = (X_n, X_{n+1}^1)$, pour $n \geq 0$. Montrer que $(Z_n)_{n\geq 0}$ est une chaîne de Markov. Donner l'expression de la loi $\left(p_{\mathrm{WR}}((X_{n+1}, X_{n+2}^1), (y, y^1))\right)_{(y,y^1)\in\mathcal{S}^2}$ de Z_{n+1} conditionnellement à Z_n. Quelle relation lie p_{WR} et p_{MH} ?

3.c Préciser un espace d'état sur laquelle la chaîne $(Z_n)_{n\geq 0}$ est irréductible. Montrer que la chaîne de Markov $(Z_n)_{n\geq 0}$ admet une mesure invariante π_{WR}, dont on donnera une expression en fonction de π et de q. Est-ce la seule mesure invariante ? Est-ce que π_{WR} est une mesure réversible pour $(Z_n)_{n\geq 0}$?

3.d Calculer la moyenne

$$\sum_{(x,x^1)\in\mathcal{S}^2} \varphi_{\mathrm{WR}}(x, x^1)\,\pi_{\mathrm{WR}}(x, x^1).$$

Montrer que p.s. $\lim\limits_{N\to\infty} I_N^{\mathrm{WR}}(\varphi) = I(\varphi)$.

3.e Montrer que $\mathrm{Var}_\pi(\varphi(X_{n+1})) \geq \mathrm{Var}_\pi(\varphi_{\mathrm{WR}}(X_n, X_{n+1}^1))$, où l'indice π indique que la loi de X_0 est π. Peut-on en déduire que la variance de $I_N^{\mathrm{WR}}(\varphi)$ est plus petite que celle de $I_N(\varphi)$?

Pour conclure sur un avantage possible de l'estimateur $I_N^{\mathrm{WR}}(\varphi)$ comparativement à $I_N(\varphi)$, il faut analyser la variance asymptotique. Des références sont données dans la correction.

Problème 2 : Interprétation probabiliste de problèmes elliptiques et dynamique réduite

Soit X_t un processus stochastique à valeurs dans $\mathbb{R}^d$ satisfaisant l'équation différentielle stochastique :

$$dX_t = -\nabla V(X_t)\,dt + \sqrt{2\beta^{-1}}dB_t \tag{1}$$

où β est un paramètre réel positif, B_t est un mouvement brownien d-dimensionnel, $V : \mathbb{R}^d \to \mathbb{R}$ est une fonction que l'on suppose suffisamment régulière pour que (1) admette une unique solution définie pour tout temps $t \geq 0$, pour une condition intiale X_0 donnée.

On indique par un indice dans l'espérance ou la probabilité la condition intiale de X_t. Ainsi, pour $x \in \mathbb{R}^d$ fixé, $\mathbb{E}_x(\varphi(X_t))$ désigne l'espérance de $\varphi(X_t)$, avec X_t solution de (1) et tel que $X_0 = x$.

1 Soit $\mathcal{D}$ un ouvert borné de $\mathbb{R}^d$. On considère

$$\tau_{\mathcal{D}} = \inf\{t \geq 0,\ X_t \notin \mathcal{D}\}.$$

1.a Rappeler pourquoi $\tau_{\mathcal{D}}$ est un temps d'arrêt. Soit $h(x) = -\mu\exp(\nu x_1)$ une fonction définie sur $\mathcal{D}$, où μ et ν sont deux réels positifs, et x_1 désigne la première composante de $x \in \mathcal{D} \subset \mathbb{R}^d$. Montrer que pour μ et ν assez grands,

$$-\nabla V \cdot \nabla h + \beta^{-1}\Delta h \leq -1 \text{ sur } \mathcal{D}.$$

Soit n un entier positif. Montrer que (on rappelle que pour deux réels a et b, $a \wedge b := \min(a, b)$),

$$h(X_{\tau_{\mathcal{D}}\wedge n}) \leq h(X_0) - \tau_{\mathcal{D}} \wedge n + \sqrt{2\beta^{-1}}\int_0^n 1_{t\leq\tau_{\mathcal{D}}}\nabla h(X_t)\cdot dB_t$$

et en déduire que $\mathbb{E}_x(\tau_{\mathcal{D}} \wedge n) \leq 2\sup_{x\in\mathcal{D}}|h(x)|$. Montrer que $\mathbb{E}_x(\tau_{\mathcal{D}}) < \infty$ et en déduire que $\mathbb{P}_x$-p.s. $\tau_{\mathcal{D}} < \infty$.

Soit une fonction $u : \mathcal{D} \to \mathbb{R}$ que l'on suppose régulière sur $\overline{\mathcal{D}}$ et telle que

$$-\nabla V \cdot \nabla u + \beta^{-1}\Delta u = 0 \text{ sur } \mathcal{D}. \tag{2}$$

1.b Montrer que pour tout entier n,

$$\mathbb{E}_x(u(X_{\tau_{\mathcal{D}}\wedge n})) = u(x).$$

En déduire que

$$\mathbb{E}_x(u(X_{\tau_{\mathcal{D}}})) = u(x).$$

1.c Montrer que si u est solution de (2) sur $\mathcal{D}$ avec conditions aux limites de Dirichlet :

$$u = f \text{ sur } \partial\mathcal{D}, \tag{3}$$

alors
$$u(x) = \mathbb{E}_x(f(X_{\tau_{\mathcal{D}}})).$$
En déduire un résultat d'unicité pour les solutions régulières de (2)–(3).

2 On considère maintenant la solution $p : \mathbb{R}^d \setminus (\overline{A} \cup \overline{B}) \to \mathbb{R}$ du problème

$$\begin{cases} -\nabla V \cdot \nabla p + \beta^{-1} \Delta p = 0 \text{ sur } \mathbb{R}^d \setminus (\overline{A} \cup \overline{B}), \\ p = 0 \text{ sur } \partial A, \\ p = 1 \text{ sur } \partial B, \end{cases} \tag{4}$$

où A et B sont deux ouverts de $\mathbb{R}^d$ disjoints. On introduit

$$T_A = \inf\{t > 0, X_t \in \overline{A}\},\ T_B = \inf\{t > 0, X_t \in \overline{B}\} \text{ et } \tau = \min(T_A, T_B).$$

On suppose dans toute la suite que A, B et V sont tels que pour tout $x \in \mathbb{R}^d$, $\mathbb{P}_x$-p.s., $T_A < \infty$ et $T_B < \infty$.

2.a Vérifier que $\tau = \inf\{t \geq 0, X_t \notin \mathbb{R}^d \setminus (\overline{A} \cup \overline{B})\}$. En utilisant les résultats de la section précédente, montrer que $p(x) = \mathbb{P}_x(T_B < T_A)$. Autrement dit, pour tout $x \in \mathbb{R}^d \setminus (\overline{A} \cup \overline{B})$,

$$p(x) = \mathbb{P}_x \left(X_t \text{ touche } B \text{ avant } A\right).$$

Le domaine $\mathbb{R}^d \setminus (\overline{A} \cup \overline{B})$ peut être vu comme la réunion des lignes de niveaux de p :

$$\mathbb{R}^d \setminus (\overline{A} \cup \overline{B}) = \bigcup_{0<z<1} \Sigma(z)$$

où, pour $z \in (0, 1)$,
$$\Sigma(z) = \{x \in \mathbb{R}^d \setminus (\overline{A} \cup \overline{B}),\, p(x) = z\}.$$
Soit $z_0 = 0 < z_1 < \ldots < z_m < z_{m+1} = 1$ une discrétisation de l'intervalle $[0, 1]$. On considère un processus X_t vérifiant (1) avec $X_0 \in \partial A$, et on introduit la suite des indices $(I_n)_{n\geq 0}$ des sous-variétés $\Sigma(z_{I_n})$ visitées par X_t, en imposant $I_n \neq I_{n+1}$. Plus précisément, $I_0 = 0$ (puisque $X_0 \in \Sigma(z_0)$), $T_1 = \inf \left\{ t > 0, X_t \in \bigcup_{i=1}^{m+1} \Sigma(z_i) \right\}$ et I_1 est défini par $X_{T_1} \in \Sigma(z_{I_1})$. De même, pour $n \geq 0$, $X_{T_n} \in \Sigma(z_{I_n})$ et on définit (T_{n+1}, I_{n+1}) par

$$T_{n+1} = \inf \left\{ t > T_n, X_t \in \bigcup_{i=0,\, i\neq I_n}^{m+1} \Sigma(z_i) \right\}$$

$$X_{T_{n+1}} \in \Sigma(z_{I_{n+1}}).$$
On renvoie à la Figure 1 pour une illustration.

2.b Vérifier que, pour tout $n \geq 0$, si $I_n = 0$, alors $I_{n+1} = 1$, si $I_n = m+1$, alors $I_{n+1} = m$. Plus généralement, vérifier que pour tout $n \geq 0$, $|I_n - I_{n+1}| = 1$. Expliquer pourquoi, pour tout $n \geq 0$, p.s., $T_n < \infty$.

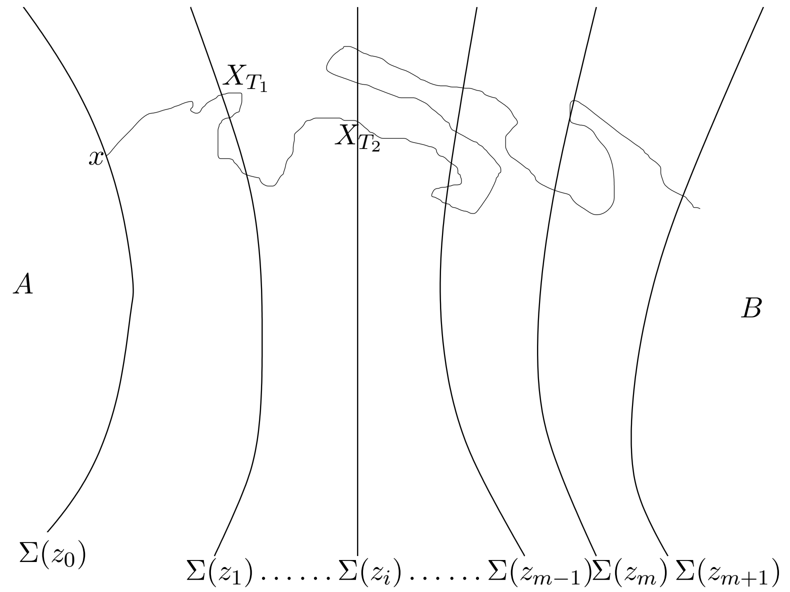

FIG. 1 – Une trajectoire de X_t traverse les sous-variétés $(\Sigma(z_i))_{0 \leq i \leq m+1}$.

2.c Soit $k \in \{1, \ldots, m\}$ fixé. Pour $x \in \bigcup_{z_{k-1} < z < z_{k+1}} \Sigma(z)$, on définit

$$\tilde{p}(x) = \mathbb{P}_x(X_t \text{ touche } \Sigma(z_{k+1}) \text{ avant } \Sigma(z_{k-1})).$$

Quelle équation aux dérivées partielles la fonction $\tilde{p}$ satisfait-elle (préciser les conditions aux limites) ? Montrer que $\tilde{p}(x) = \dfrac{p(x) - z_{k-1}}{z_{k+1} - z_{k-1}}$. En déduire que pour tout $x \in \Sigma(z_k)$,

$$\tilde{p}(x) = \frac{z_k - z_{k-1}}{z_{k+1} - z_{k-1}}.$$

2.d En déduire que I_n est une chaîne de Markov homogène en temps et de matrice de transition $q(i,j) = \mathbb{P}(I_{n+1} = j | I_n = i)$ telle que $q(i,i) = 0$, $q(0,1) = 1$, $q(m+1,m) = 1$, $q(i,j) = 0$ si $|i - j| \neq 1$, et, pour $1 \leq i \leq m$,

$$\begin{cases} q(i, i+1) = \dfrac{z_i - z_{i-1}}{z_{i+1} - z_{i-1}}, \\ q(i, i-1) = \dfrac{z_{i+1} - z_i}{z_{i+1} - z_{i-1}}. \end{cases}$$

2.e *(Question subsidiaire)* Montrer que pour une variable aléatoire T (non identiquement nulle) et à valeurs dans $[0, \infty)$, si, pour tout $s, t > 0$,

$$\mathbb{P}(T > t + s | T > s) = \mathbb{P}(T > t),$$

alors T est une loi exponentielle. On considère $J_t = I_{\tau(t)}$, où $\tau(t)$ est défini par

$$T_{\tau(t)} \leq t < T_{\tau(t)+1}.$$

Que représente J_t ? Montrer que si, conditionnellement à $(X_t)_{t \leq T_n}$, le temps T_{n+1} est exponentiellement distribué avec un paramètre qui ne dépend que de I_n, alors J_t est un processus de Markov : la loi de J_t pour $t \geq t_0$ ne dépend que de J_{t_0}.

L'intérêt de cette construction est d'obtenir une dynamique réduite, (une chaîne ou un processus de Markov à valeurs dans un espace d'états fini), à partir d'une dynamique à espace d'états continu. Des références sont données dans la correction.