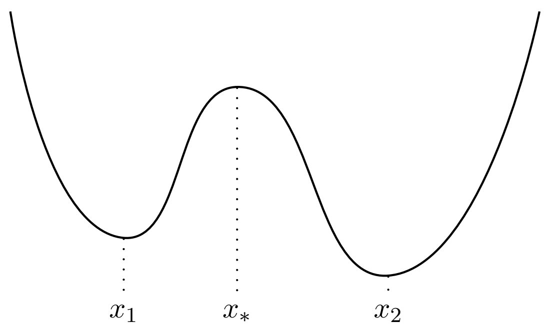
Examen du cours de M2 :
Méthodes numériques probabilistes

16 janvier 2018, 13h00-16h00.

Les notes de cours sont autorisées. Certaines questions sont facultatives : les résultats de ces questions peuvent être admis, et leur résolution éventuelle apportera des "points bonus".

Exercice : Métastabilité à basse température

Soit $V : \mathbb{R} \to \mathbb{R}$ une fonction de classe C^2, croissant vers $+\infty$ en $-\infty$ et $+\infty$, et possédant exactement deux minima locaux en $x_1 < x_2$, séparés par un point-selle $x_* \in]x_1, x_2[$ (voir la figure ci-dessous).



La diffusion d'une particule dans le *potentiel* V à la température $T > 0$ est modélisée par l'équation différentielle stochastique

$$\mathrm{d}X_t^x = -V'(X_t^x)\mathrm{d}t + \sqrt{2T}\mathrm{d}B_t, \qquad X_0^x = x,$$

où $(B_t)_{t\geq 0}$ est un mouvement Brownien standard dans $\mathbb{R}$, défini sur un espace de probabilité filtré $(\Omega, \mathcal{A}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. On suppose que V est tel que cette équation possède une unique solution forte, définie pour tout temps $t \geq 0$. On suppose également que V croît assez vite à l'infini pour que :

$$\exists T_0 > 0, \forall T \leq T_0, \int_{\mathbb{R}} \exp\left(-\frac{V(x)}{T}\right) dx < \infty.$$

Lorsque T est petit, la particule reste piégée "au fond des puits de potentiel", c'est-à-dire au voisinage de x_1 ou de x_2, sur des temps longs. Le but de l'exercice est de quantifier le temps moyen de transition d'un puits à l'autre, disons de x_2 vers x_1 pour fixer les idées. Pour tout $x \in \mathbb{R}$, on pose donc

$$\tau^x = \inf\{t \geq 0 : X_t^x = x_1\},$$

et l'on va chercher à estimer $\mathbb{E}[\tau^x]$ lorsque $T \to 0$, pour $x > x_*$.

Pour toute fonction v de classe C^2 sur $\mathbb{R}$, on définit

$$Lv = Tv'' - V'v'.$$

1. Quelle propriété possède la variable aléatoire τ^x ?

2. Pour $T \leq T_0$ et pour tout $x \in \mathbb{R}$, on pose

$$u(x) = \frac{1}{T} \int_{y=x_1}^{x} \int_{z=y}^{+\infty} \exp\left(\frac{V(y) - V(z)}{T}\right) \mathrm{d}z\mathrm{d}y.$$

Calculer $u(x_1)$ et, pour tout $x \in \mathbb{R}$, $Lu(x)$.

3. *(Question facultative.)* Soit $\phi : \mathbb{R} \to \mathbb{R}$ une fonction de classe C^2 qui vaut 0 sur $]-\infty, x_1 - 1[$ et 1 sur $[x_1, +\infty[$. On suppose également qu'il existe $\bar{x} > \max(x_2, 0)$, $a \in \mathbb{R}$, $b > 0$ et $p > 2$ tels que

$$\forall x \geq \bar{x}, \qquad V(x) = a + bx^p.$$

Montrer que la fonction $\tilde{u}$ définie par $\tilde{u}(x) = \phi(x)u(x)$ est bornée, de classe C^2, et à dérivées première et seconde bornées.

4. En déduire que pour tout $t \geq 0$,

$$\tilde{u}(X_t^x) = \tilde{u}(x) + \int_{s=0}^{t} L\tilde{u}(X_s^x)\mathrm{d}s + \sqrt{2T} \int_{s=0}^{t} \tilde{u}'(X_s^x)\mathrm{d}B_s.$$

5. On rappelle qu'on note $s \wedge t = \min(s, t)$. Vérifier que, pour tout $t \geq 0$,

$$\int_{s=0}^{t\wedge\tau^x} \tilde{u}'(X_s^x)\mathrm{d}B_s = \int_{s=0}^{t} \mathbf{1}_{\{s\leq\tau^x\}}\tilde{u}'(X_s^x)\mathrm{d}B_s,$$

et que l'espérance du terme de droite est nulle.

6. En déduire que, pour tout $t \geq 0$, pour tout $x \geq x_1$,

$$\mathbb{E}\left[\tilde{u}(X_{t\wedge\tau^x}^x)\right] = u(x) - \mathbb{E}\left[t \wedge \tau^x\right].$$

7. En conclure une formule explicite pour $\mathbb{E}[\tau^x]$ lorsque $x \geq x_1$.

8. *(Question facultative.)* Soient D un ouvert de $\mathbb{R}^2$ et $M : D \to \mathbb{R}$ une fonction de classe C^2 tels que

$$\exists \eta_0 > 0, \qquad \int_{\xi\in D} \exp\left(\frac{M(\xi)}{\eta_0}\right) \mathrm{d}\xi < +\infty.$$

On suppose que :
— M atteint son maximum sur D en un unique point $\xi_0 \in D$ et la matrice hessienne $\nabla^2 M(\xi_0)$ de M en ξ_0 est définie négative,
— pour tout $r > 0$, il existe $\delta > 0$ tels que pour tout $\xi \in D$ tel que $|\xi - \xi_0| \geq r$, alors $M(\xi) \leq M(\xi_0) - \delta$.

Montrer que pour tout $\eta \in (0, \eta_0]$, $\int_{\xi\in D} \exp\left(\frac{M(\xi)}{\eta}\right) \mathrm{d}\xi < +\infty$. Montrer que, lorsque $\eta \to 0$,

$$\int_{\xi\in D} \exp\left(\frac{M(\xi)}{\eta}\right) \mathrm{d}\xi \sim \exp\left(\frac{M(\xi_0)}{\eta}\right) \frac{2\pi\eta}{\sqrt{|\det \nabla^2 M(\xi_0)|}}.$$

9. En supposant $V''(x_2) > 0$ et $V''(x_*) < 0$, en déduire un équivalent, lorsque $T \to 0$, de $\mathbb{E}[\tau^x]$ pour $x > x_*$.

10. *(Question bonus.)* En chimie, le processus $(X_t^x)_{t\geq 0}$ est souvent utilisé pour modéliser l'évolution d'une coordonnée de réaction. Dans ce contexte, connaissez-vous le nom de la formule que vous venez d'établir ?

Problème : Algorithme de Metropolis Hastings

On considère un espace d'états fini

$$M = \{1, \dots, d\}$$

muni d'une mesure de probabilité π que l'on suppose strictement positive sur M. On identifie π à un vecteur ligne dans $\mathbb{R}^d$. On souhaite calculer la moyenne d'une observable $f : M \to \mathbb{R}$ par rapport à π, qui s'écrit en identifiant f à un vecteur colonne dans $\mathbb{R}^d$:

$$\mathbb{E}_\pi(f) = \sum_{x=1}^{d} f(x)\pi(x) = \pi f.$$

On notera de même $\mathrm{Var}_\pi(f) = \mathbb{E}_\pi((f - \mathbb{E}_\pi(f))^2)$.

Première partie : l'algorithme de Metropolis Hastings. L'algorithme de Metropolis Hastings nécessite deux ingrédients :

- Une proposition de changement $Q : M \times M \to \mathbb{R}$ qui est le noyau d'une chaîne de Markov telle que
$$\forall x, y \in M^2,\ Q(x, y) > 0 \iff Q(y, x) > 0.$$
On suppose que l'on sait échantillonner les mesures $Q(x, \cdot)$.
- Une fonction d'acceptation $a : M \times M \to (0, 1]$.

L'algorithme construit alors une chaîne de Markov $(X_n)_{n\geq 0}$ de la façon suivante : à l'itération n, X_n étant connu,

- On propose un déplacement vers $\tilde{X}_{n+1}$ suivant la loi $Q(X_n, \cdot)$;
- On accepte ce déplacement avec probabilité $a(X_n, \tilde{X}_{n+1})$, c'est-à-dire : On tire U_n de loi uniforme sur $(0, 1)$ et
  - Si $U_n \leq a(X_n, \tilde{X}_{n+1})$, alors $X_{n+1} = \tilde{X}_{n+1}$;
  - Si $U_n > a(X_n, \tilde{X}_{n+1})$, alors $X_{n+1} = X_n$.

1. Ecrire la matrice de transition P de la chaîne de Markov $(X_n)_{n\geq 0}$.
2. On souhaite que $(X_n)_{n\geq 0}$ soit réversible par rapport à π, c'est-à-dire :
$$\forall x, y,\ \pi(x)P(x, y) = \pi(y)P(y, x).$$
Soit $F : \mathbb{R}_+^* \to (0, 1]$ une fonction telle que
$$\forall z > 0,\ F(z) = zF\left(\frac{1}{z}\right).$$
Montrer que $(X_n)_{n\geq 0}$ est réversible si et seulement si pour tout $x \neq y$ tel que $Q(x, y) > 0$,
$$a(x, y) = \lambda(x, y)F(r(x, y)) \tag{1}$$
où $\lambda : M \times M \to \mathbb{R}$ est une fonction symétrique (pour tout $x \neq y$ tel que $Q(x, y) > 0$, $\lambda(x, y) = \lambda(y, x)$) et
$$r(x, y) = \frac{\pi(y)Q(y, x)}{\pi(x)Q(x, y)}$$
est le rapport de Metropolis-Hastings. Vérifier que les deux règles d'acceptation suivantes sont bien de la forme (1) : la règle de Barker $\lambda = 1$ et $F(z) = \frac{z}{1+z}$ et la règle de Metropolis-Hastings $\lambda = 1$ et $F(z) = \min(1, z)$.

3. Ecrire l'algorithme dans le cas particulier où Q est symétrique ($Q(x,y) = Q(y,x)$), la règle d'acceptation est celle de Metropolis-Hastings et la loi cible est une loi de Boltzmann-Gibbs de la forme $\pi(x) = Z^{-1}\exp(-\beta V(x))$ pour une fonction potentiel $V : M \to \mathbb{R}$, une température inverse $\beta > 0$ et $Z = \sum_{x\in M}\exp(-\beta V(x))$. Dans quelle situation a-t-on $a(x,y) = 1$? Commenter.

Deuxième partie : généralités sur les chaînes de Markov réversibles. Soit $(X_n)_{n\geq 0}$ une chaîne de Markov sur M, réversible par rapport à π, de matrice de transition P.

4. Montrer que si X_0 est distribué sous π, alors, pour tout $n \geq 1$, $(X_0, X_1, \ldots, X_n)$ a même loi que $(X_n, X_{n-1}, \ldots, X_0)$.

On introduit le produit scalaire sur $\mathbb{R}^d$:

$$\forall (f,g) \in (\mathbb{R}^d)^2,\ \langle f, g\rangle_\pi = \sum_{x=1}^{d} f(x)g(x)\pi(x)$$

et on note $\|\cdot\|_{L^2(\pi)}$ la norme associée.

5. Montrer que $(X_n)_{n\geq 0}$ est une chaîne réversible par rapport à π si et seulement si P est auto-adjoint pour ce produit scalaire.

On note $(\lambda_i)_{i=1,\ldots,d}$ les valeurs propres de P ordonnées de manière décroissante, et $(e_i)_{i=1,\ldots,d}$ les vecteurs propres associés, que l'on suppose orthonormés pour le produit scalaire $\langle\cdot,\cdot\rangle_\pi$.

6. Montrer que pour tout $i \in \{1,\ldots,d\}$, $|\lambda_i| \leq 1$, que $\lambda_1 = 1$ et que $e_1 = (1,\ldots,1)^T$.
7. On suppose que 1 est valeur propre simple. Montrer que P admet π comme unique mesure invariante.
8. On suppose que 1 est valeur propre simple et que $\lambda_d > -1$. Montrer que pour tout $f : M \to \mathbb{R}$,
$$\|P^n \bar{f}\|^2_{L^2(\pi)} \leq \rho^{2n}\|\bar{f}\|^2_{L^2(\pi)}$$
où $\bar{f} = f - \mathbb{E}_\pi(f)$ et $\rho = \max(|\lambda_2|, |\lambda_d|) < 1$. En déduire [1] que pour tout $x \in M$
$$|\mathbb{E}^x(f(X_n)) - \mathbb{E}_\pi(f)|^2 \leq \frac{\rho^{2n}}{\pi(x)}\mathrm{Var}_\pi(f).$$

Dans les deux questions (facultatives) qui suivent, on fait le lien entre les hypothèses spectrales ci-dessus et les propriétés d'irréductibilité et d'apériodicité.

9. *(Question facultative.)* On rappelle que P est irréductible si et seulement si :
$$\forall (x,y) \in M^2, \exists n \geq 1,\ P^n(x,y) > 0.$$
   (a) Montrer que P est irréductible si et seulement si, pour tout $(x,y) \in M^2$, il existe $n \geq 1$ et un chemin $(x_0, x_1, \ldots, x_n)$ avec $x_0 = x$ et $x_n = y$ et tel que $\mathbb{P}((X_0,\ldots,X_n) = (x_0,\ldots,x_n)) > 0$.
   (b) Soit
   $$\mathcal{E}(f,P) = \sum_{x,y}(f(y) - f(x))^2\pi(x)P(x,y)$$
   la forme de Dirichlet associée à P. Montrer que [2]
   $$\mathcal{E}(f,P) = \mathbb{E}^\pi((f(X_1) - f(X_0))^2) = 2\mathbb{E}^\pi(f\cdot(f - Pf)(X_0))$$
   où $(f\cdot(f-Pf))(x) = f(x)(f(x) - Pf(x)) = f^2(x) - f(x)\sum_{y\in M}P(x,y)f(y)$.

1. L'exposant x dans $\mathbb{E}^x$ indique qu'on considère la chaîne de Markov démarrant en x : $X_0 = x$.
2. L'exposant π dans $\mathbb{E}^\pi$ indique qu'on considère la chaîne de Markov démarrant sous π : $X_0 \sim \pi$.

(c) En utilisant la question précédente, montrer que si P est irréductible,

$$Pf = f \implies \forall (x,y) \in M^2, f(x) = f(y).$$

En déduire que si P est irréductible alors 1 est valeur propre simple.

10. *(Question facultative)* On rappelle que P est irréductible et apériodique[3] si et seulement si

$$\exists n_0 \geq 1, \forall (x,y) \in M^2, P^{n_0}(x,y) > 0.$$

(a) Soit $\epsilon > 0$ tel que $\forall (x,y) \in M^2, P^{n_0}(x,y) > \epsilon\pi(y)$. Vérifier que $\epsilon < 1$. Soit l'application

$$\mathcal{T} : \begin{cases} \mathcal{P}(M) \to \mathcal{P}(M) \\ \quad \nu \mapsto \nu P \end{cases}$$

où $\mathcal{P}(M)$ désigne les mesures de probabilité sur M et pour tout $y \in M$, $(\nu P)(y) = \sum_{x \in M} \nu(x) P(x,y)$. Montrer que pour tout $(\mu, \nu) \in (\mathcal{P}(M))^2$,

$$\|\mathcal{T}^{n_0}(\nu) - \mathcal{T}^{n_0}(\mu)\|_{L^1} \leq (1-\epsilon)\|\nu - \mu\|_{L^1},$$

où pour tout $\nu \in \mathcal{P}(M)$, $\|\nu\|_{L^1} = \sum_{x=1}^d |\nu(x)|$. En déduire que π est l'unique point fixe de $\mathcal{T}$ et que pour tout $\mu \in \mathcal{P}(M)$, pour tout $n \geq 1$,

$$\|\mu P^n - \pi\|_{L^1} \leq 2(1-\epsilon)^{\left\lfloor \frac{n}{n_0} \right\rfloor}$$

où $\left\lfloor \frac{n}{n_0} \right\rfloor$ désigne la partie entière de $\frac{n}{n_0}$.

(b) Déduire du résultat précédent que si P est irréductible et apériodique, alors 1 est valeur propre simple et $\lambda_d > -1$.

Troisième partie : variances asymptotiques pour les chaînes de Markov réversibles. On suppose dans la suite que P est irréductible et apériodique. En terme de spectre (cf. questions 9 et 10), on a donc

$$1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_d > -1.$$

On déduit du fait que $(X_n)_{n \geq 0}$ admet π comme unique mesure invariante la loi forte des grands nombres (c'est une conséquence du théorème de Birkhoff) : presque sûrement,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \mathbb{E}_\pi(f).$$

On admettra également qu'il existe un théorème central limite associé à cette convergence :

$$\lim_{n \to \infty} \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} (f(X_k) - \mathbb{E}_\pi(f)) = \mathcal{N}(0, v(f,P)) \text{ en loi}$$

où $v(f,P) = \lim_{n \to \infty} \mathbb{E}\left(\frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} (f(X_k) - \mathbb{E}_\pi(f))\right)^2$ est la variance asymptotique.

---

3. Attention ! Cette définition n'est valable que dans notre cadre, où M est un espace d'état fini.

11. Montrer que $\sum_{n\geq 1} \mathrm{Cov}^\pi(f(X_0), f(X_n)) < \infty$ où $\mathrm{Cov}^\pi(f(X_0), f(X_n)) = \mathbb{E}^\pi(\bar{f}(X_0)\bar{f}(X_n))$ (on rappelle que $\bar{f} = f - \mathbb{E}_\pi(f)$). Supposons que la chaîne démarre à l'état stationnaire : $X_0 \sim \pi$. Montrer que
$$v(f,P) = \mathrm{Var}_\pi(f) + 2\sum_{n\geq 1} \mathrm{Cov}^\pi(f(X_0), f(X_n)).$$
On admettra que ce résultat reste vrai pour toute condition initiale X_0. *(On pourra chercher à le démontrer en deuxième lecture.)*
12. Montrer que
$$v(f,P) = 2\langle \bar{f}, (\mathrm{Id} - P)^{-1}\bar{f}\rangle_\pi - \|\bar{f}\|^2_{L^2(\pi)}.$$
On justifiera précisément le sens donné à $(\mathrm{Id} - P)^{-1}\bar{f}$.
13. Montrer que
$$\langle \bar{f}, (\mathrm{Id} - P)^{-1}\bar{f}\rangle_\pi = \sup_{g\in L^2_0(\pi)} 2\langle \bar{f}, g\rangle_\pi - \langle g, (\mathrm{Id} - P)g\rangle_\pi$$
où $L^2_0(\pi)$ désigne le sous espace de $L^2(\pi)$ des vecteurs de moyenne nulle sous π. En déduire que
$$v(f,P) = \sup_{g\in L^2_0(\pi)} 4\langle \bar{f}, g\rangle_\pi - \mathcal{E}(g,P) - \|\bar{f}\|^2_{L^2(\pi)} \tag{2}$$
où on rappelle (cf. question 9) que pour tout $f : M \to \mathbb{R}$,
$$\mathcal{E}(f,P) = \sum_{x,y} (f(y) - f(x))^2 \pi(x) P(x,y) = \mathbb{E}^\pi((f(X_1) - f(X_0))^2)$$
est la forme de Dirichlet associée à P.
14. Soit P et Q deux matrices de transition réversibles par rapport à π. On introduit la notation :
$$P \succ Q \iff \forall x \neq y,\ P(x,y) \geq Q(x,y).$$
Montrer que
$$P \succ Q \Longrightarrow \forall f : M \to \mathbb{R},\ \mathcal{E}(f,P) \geq \mathcal{E}(f,Q).$$
15. Soit P et Q deux matrices de transition réversibles par rapport à π, avec P et Q irréductibles et apériodiques. Montrer que si $P \succ Q$, alors, pour tout $f : M \to \mathbb{R}$, $v(f,P) \leq v(f,Q)$. Donner une interprétation de ce résultat.
16. *(Question facultative.)* Montrer que
$$v(f,P) = \sum_{k=2}^{d} \frac{1+\lambda_k}{1-\lambda_k} |\langle f - \mathbb{E}_\pi(f), e_k\rangle_\pi|^2.$$
En déduire que si $P \succ Q$, alors $\lambda_2(P) \leq \lambda_2(Q)$, où on indique explicitement entre parenthèses la dépendance de la deuxième valeur propre en la matrice de transition considérée.

Quatrième partie : retour sur l'algoritme de Metropolis Hastings. On considère à nouveau l'algorithme de Metropolis Hastings présenté en première partie, pour une proposition de changement Q fixée, et une fonction d'acceptation $a : M \times M \to (0,1]$ telle que $(X_n)_{n\geq 0}$ soit réversible. On note P la matrice de transition de $(X_n)_{n\geq 0}$. On suppose que Q est irréductible et apériodique.

17. En utilisant le résultat de la question 10, montrer que P est irréductible et apériodique. On a donc $\lambda_2(P) < \lambda_1(P) = 1$ et $\lambda_d(P) > -1$.
18. Montrer que pour une proposition de changement Q fixée, le choix de Metropolis-Hastings $a_{\mathrm{MH}}(x,y) = \min(1, r(x,y))$ est optimal en terme de variance asymptotique.