

TD 3 : CHAÎNES 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 $\rho(x, x') = \rho((u, v), (u', v'))$.

Correction : On a donc $X_{n+1} = X_n + \sigma W_{n+1}$, où $(W_n)_{n \geq 1}$ est une suite i.i.d. de vecteurs gaussiens standards indépendants de X_0 . En notant $\|x\|$ la norme euclidienne dans $\mathbb{R}^2$, il vient

$$q(x, x') = \frac{1}{2\pi\sigma^2} e^{-\frac{\|x' - x\|^2}{2\sigma^2}} = q(x, x'),$$

et nous avons affaire à un noyau de Metropolis symétrique, si bien que le rapport de Metropolis se simplifie comme suit :

$$\rho(x, x') = \rho((u, v), (u', v')) = \frac{f(u', v')}{f(u, v)} = \frac{(\cos u')^2 (\sin v')^2}{(\cos u)^2 (\sin v)^2} e^{\frac{\|x\|^2 - \|x'\|^2}{20}}.$$

2. Proposer une méthode de rejet pour simuler suivant la loi de densité f à partir d'une loi instrumentale gaussienne.

Correction : Notons $f(u, v) = c \times f_0(u, v)$ où f_0 correspond à la version non normalisée

$$f_0(u, v) = (\cos u)^2 (\sin v)^2 e^{-\frac{u^2+v^2}{20}}.$$

On commence par noter que

$$\forall (u, v) \in \mathbb{R}^2 \quad f_0(u, v) \leq \exp\left(-\frac{1}{20}(u^2 + v^2)\right).$$

En considérant

$$h(u, v) = \frac{1}{20\pi} \exp\left(-\frac{1}{20}(u^2 + v^2)\right)$$

la densité de la loi normale centrée et de matrice de covariance $10I_2$, on a donc

$$\forall (u, v) \in \mathbb{R}^2 \quad f(u, v) = c \times f_0(u, v) \leq 20\pi \times c \times h(u, v),$$

ou encore

$$\forall (u, v) \in \mathbb{R}^2 \quad \frac{f(u, v)}{h(u, v)} = 20\pi c (\cos u)^2 (\sin v)^2 \leq 20\pi c.$$

En prenant $m = 20\pi c$, le rapport d'acceptation vaut donc tout simplement

$$r(u, v) = (\cos u)^2 (\sin v)^2.$$

La simulation suivant la densité f par méthode de rejet avec h comme loi instrumentale s'en déduit simplement.

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 $\|\mu - \nu\| = p$.

Correction : Pour la mesure dominante $\alpha(dx) = \lambda(dx) + \delta_{x_0}(dx)$ les densités sont les fonctions

$$\frac{d\mu}{d\alpha}(x) = f(x) \mathbf{1}_{x \neq x_0}$$

et

$$\frac{d\nu}{d\alpha}(x) = p \mathbf{1}_{x=x_0} + (1 - p) f(x) \mathbf{1}_{x \neq x_0}.$$

On applique alors la formule vue en cours :

$$\|\mu - \nu\| = \frac{1}{2} \int_{\mathbb{R}} \left| \frac{d\mu}{d\alpha}(x) - \frac{d\nu}{d\alpha}(x) \right| \alpha(dx),$$

d'où

$$\|\mu - \nu\| = \frac{1}{2} \left(\left| \frac{d\mu}{d\alpha}(x_0) - \frac{d\nu}{d\alpha}(x_0) \right| \alpha(\{x_0\}) + \int_{\mathbb{R} \setminus \{x_0\}} \left| \frac{d\mu}{d\alpha}(x) - \frac{d\nu}{d\alpha}(x) \right| \alpha(dx) \right),$$

donc

$$\|\mu - \nu\| = \frac{1}{2} \left(p + \int_{\mathbb{R} \setminus \{x_0\}} p f(x) dx \right) = p.$$

2. Montrer que λ est invariante pour P .

Correction : Notons $\mathcal{N} := \{x \in [0, 1] : \exists m \geq 2, x = 1/m\}$, ensemble dénombrable qui est donc négligeable pour la mesure de Lebesgue. On veut montrer que $\lambda P = \lambda$, c'est-à-dire que pour toute fonction test φ , on a $(\lambda P)(\varphi) = \lambda(\varphi)$. Or, par définition de l'action d'un noyau et par Fubini, on a

$$(\lambda P)(\varphi) = \iint dx P(x, dy) \varphi(y) = \int \left(\int_{[0,1]} dx P(x, dy) \right) \varphi(y) = \int \left(\int_{[0,1] \setminus \mathcal{N}} dx P(x, dy) \right) \varphi(y).$$

Si $x \notin \mathcal{N}$, alors $P(x, dy) = dy$ donc

$$(\lambda P)(\varphi) = \int \left(\int_{[0,1] \setminus \mathcal{N}} dx dy \right) \varphi(y) = \int \left(\int_{[0,1] \setminus \mathcal{N}} dx \right) dy \varphi(y) = \int dy \varphi(y) = \lambda(\varphi).$$

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.

Correction : Soit un borélien $B \subseteq [0, 1]$ tel que $\lambda(B) > 0$. Que le réel x de $]0, 1]$ soit dans $\mathcal{N}$ ou non, on peut toujours écrire

$$P(x, B) \geq x^2 \lambda(B) > 0.$$

De plus, si $x = 0$, on a $P(x, B) = \lambda(B) > 0$. Un résultat du cours assure alors 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 .

Correction : Puisque x n'est pas dans $\mathcal{N}$, on a

$$\mu_1 := \text{Loi}(X_1) = P(x, \cdot) = \text{Unif}([0, 1]) = \lambda.$$

Or on a prouvé que $\lambda P = \lambda$. Ceci et la relation de Chapman-Kolmogorov impliquent que, pour tout $n \geq 1$,

$$\mu_n := \text{Loi}(X_n) = \mu_1 P^{n-1} = \lambda P^{n-1} = \lambda.$$

5. Soit $X_0 = 1/m$ avec $m \geq 2$. Préciser la loi de X_n pour tout n .

Correction : En notant

$$p_n(m) := \prod_{j=0}^{n-1} \left(1 - \frac{1}{(m+j)^2} \right),$$

on va montrer par récurrence que, pour tout $n \geq 1$,

$$\mu_n = \text{Loi}(X_n) = (1 - p_n(m))\lambda + p_n(m)\delta_{1/(m+n)}.$$

Pour $n = 1$

$$\mu_1 = \text{Loi}(X_1) = P(x, \cdot) = x^2 \text{Unif}([0, 1]) + (1 - x^2)\delta_{1/(m+1)} = \frac{1}{m^2}\lambda + \left(1 - \frac{1}{m^2}\right)\delta_{1/(m+1)},$$

qui donne bien

$$\mu_1 = (1 - p_1(m))\lambda + p_1(m)\delta_{1/(m+1)}.$$

Supposons que c'est vrai au rang n . Puisque $\lambda P = P$, on a

$$\mu_{n+1} = \mu_n P = (1 - p_n(m))\lambda P + p_n(m)\delta_{1/(m+n)}P = (1 - p_n(m))\lambda + p_n(m)\delta_{1/(m+n)}P,$$

avec

$$\delta_{1/(m+n)}P = P(1/(m+n), \cdot) = \frac{1}{(m+n)^2}\lambda + \left(1 - \frac{1}{(m+n)^2}\right)\delta_{1/(m+n+1)},$$

donc

$$\mu_{n+1} = \left((1 - p_n(m)) + p_n(m)\frac{1}{(m+n)^2} \right) \lambda + p_{n+1}(m)\delta_{1/(m+n+1)},$$

ce qui aboutit bien à

$$\mu_{n+1} = (1 - p_{n+1}(m))\lambda + p_{n+1}(m)\delta_{1/(m+n+1)}.$$

Remarque : Un produit vide faisant 1 par convention, la relation est également vérifiée pour $n = 0$.

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 ?

Correction : Puisque $\mu_n = P^n(x, \cdot)$, ce qui précède montre que

— Si $x \notin \mathcal{N}$, alors $\mu_n = \lambda$ pour tout $n \geq 1$ donc $\|P^n(x, \cdot) - \lambda\| = 0$.

— Si $x = 1/m \in \mathcal{N}$, alors pour tout $n \geq 0$, d'après la question préliminaire

$$\|P^n(x, \cdot) - \lambda\| = p_n(m),$$

avec

$$\log p_n(m) = \sum_{j=0}^{n-1} \log \left(1 - \frac{1}{(m+j)^2} \right).$$

et

$$\log \left(1 - \frac{1}{(m+j)^2} \right) \underset{j \rightarrow \infty}{\sim} -\frac{1}{(m+j)^2}.$$

Puisque

$$\sum_{j=0}^{\infty} \frac{1}{(m+j)^2} < \infty,$$

le critère d'équivalence sur les séries de signe constant implique qu'il existe $p(m) > 0$ tel que

$$\log p_n(m) \xrightarrow{n \rightarrow \infty} \log p(m),$$

donc

$$\|P^n(x, \cdot) - \lambda\| \xrightarrow{n \rightarrow \infty} p(m) > 0,$$

et il n'y a pas convergence en variation totale vers la mesure de Lebesgue.

Bilan : Il n'y a convergence en variation totale vers la mesure invariante λ que si la condition initiale x n'est pas dans l'ensemble $\mathcal{N}$. Ceci est cohérent avec le théorème vu en cours : puisque P est λ -irréductible, λ -invariante et apériodique, on savait sans faire de calculs que pour presque toute condition initiale x de $[0, 1]$, on aurait

$$\|P^n(x, \cdot) - \lambda\| \xrightarrow{n \rightarrow \infty} 0.$$

On a bien $\lambda(\mathcal{N}) = 0$ et les calculs ont permis de préciser pour quels x ça ne marche effectivement pas.

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.

Correction : Par intégration, on déduit de la condition $\pi(x) \leq Mq(x)$:

$$1 = \int \pi(x)dx \leq M \int q(x)dx = M, \quad \text{i.e.} \quad 1 \leq M.$$

Ensuite,

$$\pi \leq q \quad \Longleftrightarrow \quad \begin{cases} q \geq \pi \\ \int (q(x) - \pi(x))dx = 1 - 1 = 0 \end{cases} \quad \Longleftrightarrow \quad \pi(x) = q(x),$$

puisque l'intégrale d'une fonction positive est nulle si et seulement si l'intégrand est nul presque partout.

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 .

Correction : On a

$$p(x, y) = \alpha(x, y)q(y) = \min\left(1, \frac{\pi(y)q(x)}{\pi(x)q(y)}\right) q(y) = \pi(y) \min\left(\frac{q(y)}{\pi(y)}, \frac{q(x)}{\pi(x)}\right) \geq \frac{\pi(y)}{M},$$

et

$$r(x) = 1 - \int_{\mathbb{R}} p(x, y)dy \geq 1 - \int_{\mathbb{R}} q(y)dy = 0.$$

3. 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.

Correction : Par définition, on a

$$(P\varphi)(x) = \int_{\mathbb{R}} P(x, dy)\varphi(y) = \int_{\mathbb{R}} p(x, y)\varphi(y)dy + \int_{\mathbb{R}} r(x)\delta_x(dy)\varphi(y) = \int_{\mathbb{R}} p(x, y)\varphi(y)dy + r(x)\varphi(x)$$

avec $r(x)\varphi(x) \geq 0$ et $p(x, y) \geq \frac{\pi(y)}{M}$ donc

$$(P\varphi)(x) \geq \frac{1}{M} \int_{\mathbb{R}} \varphi(y)\pi(y)dy = \frac{1}{M}\pi(\varphi).$$

4. 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.

Correction : Fixons le réel x . Pour tout borélien B , on a $P(x, B) = (P\mathbf{1}_B)(x)$ et $\pi(B) = \pi(\mathbf{1}_B)$ donc la question précédente implique

$$R(x, B) = \frac{M}{M-1} \left((P\mathbf{1}_B)(x) - \frac{1}{M}\pi(\mathbf{1}_B) \right) \geq 0.$$

La sigma-additivité de $R(x, \cdot)$ est assurée par les sigma-additivités des mesures de probabilité $P(x, \cdot)$ et π . En particulier, pour tout B , on a

$$R(x, B) + R(x, \mathbb{R} \setminus B) = R(x, \mathbb{R}) = \frac{M}{M-1} \left(P(x, \mathbb{R}) - \frac{1}{M}\pi(\mathbb{R}) \right) = 1,$$

d'où $0 \leq R(x, B) \leq 1$ et $R(x, \cdot)$ est bien une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Par ailleurs, pour tout borélien B fixé, l'application $x \mapsto R(x, B)$ est borélienne car P est un noyau de transition.

5. 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

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

Correction : Par définition du noyau R , on a $P(x, dy) = \frac{1}{M}\pi(dy) + (1 - \frac{1}{M})R(x, dy)$, donc pour toute mesure de probabilité μ et toute fonction test φ

$$(\mu P)(\varphi) = \iint \mu(dx)P(x, dy)\varphi(y) = \frac{1}{M} \iint \mu(dx)\pi(dy)\varphi(y) + \left(1 - \frac{1}{M}\right) \iint \mu(dx)R(x, dy)\varphi(y),$$

avec, par Fubini et puisque μ est une mesure de probabilité,

$$\frac{1}{M} \iint \mu(dx)\pi(dy)\varphi(y) = \frac{1}{M} \int \left(\int \mu(dx) \right) \pi(dy)\varphi(y) = \frac{1}{M} \int \pi(dy)\varphi(y) = \frac{1}{M}\pi(\varphi),$$

donc

$$(\mu P)(\varphi) = \frac{1}{M}\pi(\varphi) + \left(1 - \frac{1}{M}\right) (\mu R)(\varphi),$$

ce qui prouve que les deux mesures de probabilité μP et $\frac{1}{M}\pi + (1 - \frac{1}{M})\mu R$ sont égales. On peut alors établir l'égalité suivante par récurrence. La relation précédente appliquée à $\mu = \delta_x$ montre qu'elle est vraie pour $n = 1$. Supposons-la vraie au rang n , c'est-à-dire

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

C'est une égalité entre mesures, conservée lorsqu'on leur applique le noyau P (si on était en espace d'états dénombrable, penser à la multiplication d'un vecteur ligne par une matrice), ce qui donne

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

avec $\pi P = P$ donc

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

Or $\delta_x R^n$ est une mesure de probabilité et $\mu P = \frac{1}{M}\pi + (1 - \frac{1}{M})\mu R$ d'où

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

ce qui donne

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

et la récurrence est établie.

6. On rappelle que $\|\mu - \nu\| = \sup_{\varphi: E \rightarrow [0,1]} |\mu(\varphi) - \nu(\varphi)|$. Dédurre 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

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

Correction : Pour toute fonction borélienne $\varphi : \mathbb{R} \rightarrow [0, 1]$, on a

$$|(\delta_x P^n)(\varphi) - \pi(\varphi)| = \left(1 - \frac{1}{M}\right)^n |(\delta_x R^n)(\varphi) - \pi(\varphi)| \leq \left(1 - \frac{1}{M}\right)^n \|\delta_x R^n - \pi\| \leq \left(1 - \frac{1}{M}\right)^n,$$

car la distance en variation totale entre deux mesures de probabilités est toujours inférieure ou égale à 1. Il reste à prendre le supremum à gauche sur toutes les fonctions φ et à dire que $\delta_x P^n = P^n(x, \cdot)$ pour conclure que

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

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

Correction : La condition $\pi(x) \leq Mq(x)$ est celle qui est requise pour appliquer la méthode de rejet lorsqu'on veut simuler selon π à partir de propositions faites selon q . Par la méthode de rejet, chaque fois que l'on accepte, on obtient une variable distribuée exactement selon π . En revanche, par Metropolis indépendant, on construit une chaîne (X_n) : pour n donné, la loi de X_n n'est pas π (sauf si X_0 a pour loi π), mais la loi limite est π et il y a convergence vers celle-ci à vitesse géométrique. Par ailleurs, on retrouve le fait que plus M est proche de 1, mieux c'est.

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.

1. Rappeler la densité f correspondante. Quelle est sa moyenne ? Quel est son écart-type ?

Correction : La densité f correspondante est

$$f(x) = \frac{1}{2} \times \frac{1}{\sqrt{2\pi}} e^{-\frac{(x+3)^2}{2}} + \frac{1}{2} \times \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-3)^2}{2}}.$$

Par symétrie, sa moyenne est bien entendu 0. Notons X de densité f , $Y \sim \mathcal{N}(-m, 1)$ de densité notée g et $T \sim \mathcal{N}(m, 1)$ de densité notée h , donc $f = \frac{1}{2}(g + h)$ et

$$\text{Var}(X) = \mathbb{E}[X^2] = \int_{\mathbb{R}} x^2 f(x) dx = \frac{1}{2} \left(\int_{\mathbb{R}} x^2 g(x) dx + \int_{\mathbb{R}} x^2 h(x) dx \right) = \frac{1}{2} (\mathbb{E}[Y^2] + \mathbb{E}[T^2]),$$

or

$$\mathbb{E}[Y^2] = \text{Var}(Y) + \mathbb{E}[Y]^2 = 1 + m^2 = \mathbb{E}[T^2],$$

donc $\text{Var}(X) = 1 + m^2 = 10$ et l'écart-type de X vaut donc $\sigma(X) = \sqrt{10}$.

2. 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 ?

Correction : Puisque Q et la loi cible (de densité f) sont absolument continues par rapport à la mesure de Lebesgue sur $\mathbb{R}$ et comme

$$(x, x') \mapsto q(x, x') = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x'-x)^2}{2}}$$

est continue et strictement positive sur $\mathbb{R} \times \mathbb{R}$, on sait par un résultat du cours que le noyau de Metropolis est irréductible et apériodique donc en théorie il n'y a aucun problème pour la convergence en variation totale, ainsi que pour le Théorème ergodique. Néanmoins, partant de $X_0 = m = 3$, les propositions faites par le noyau instrumental Q vont proposer d'aller vers un point moins favorable pour la densité f donc seront majoritairement refusées, et il faudra beaucoup de temps avant d'arriver dans le voisinage du second mode, c'est-à-dire autour de $-m = -3$.

3. 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.

- (a) Quelles transitions sont favorisées par ce noyau instrumental ?

Correction : Ce noyau instrumental favorise les propositions vers des valeurs plus élevées de f , donc vers des régions plus intéressantes : ceci est dû au signe de $f'(x)$. De plus, on notera que le rapport $f'(x)/f(x)$ assure qu'on peut appliquer cette méthode à des densités connues uniquement à une constante multiplicative près, donc qu'elle a le même avantage que l'algorithme de Metropolis de ce point de vue.

- (b) Expliciter le rapport de Metropolis $\rho(x, y)$.

Correction : On n'est pas dans un cas Metropolis symétrique, donc le rapport de Metropolis ne se simplifie pas :

$$\rho(x, y) = \frac{f(y)q(y, x)}{f(x)q(x, y)},$$

avec

$$q(x, y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2} \left[y - x - \frac{\sigma^2}{2} \frac{f'(x)}{f(x)}\right]^2\right).$$

On a au final

$$\log \rho(x, y) = \log f(y) - \log f(x) + \frac{x - y}{2} \left(\frac{f'(x)}{f(x)} + \frac{f'(y)}{f(y)} \right) + \frac{\sigma^2}{8} \left(\frac{f'(x)^2}{f(x)^2} - \frac{f'(y)^2}{f(y)^2} \right).$$

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.

1. 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 $\Phi_{\mathcal{N}(0,1)}(t) = e^{-t^2/2}$.

Correction : On a $X_1 = rX_0 + \sqrt{1-r^2}W_1$ avec W_1 indépendante de X_0 . En passant par les fonctions caractéristiques, on a donc pour tout réel t

$$\mathbb{E} [e^{itX_1}] = \mathbb{E} [e^{it(rX_0 + \sqrt{1-r^2}W_1)}] = \mathbb{E} [e^{i(tr)X_0}] \mathbb{E} [e^{i(t\sqrt{1-r^2})W}] = e^{-\frac{(1-r^2)t^2}{2}} \mathbb{E} [e^{i(tr)X_0}].$$

Notons Φ_π la fonction caractéristique de la loi π : rappelons que celle-ci est nécessairement continue sur $\mathbb{R}$ (et même uniformément) et qu'elle vaut 1 en 0. Alors π est invariante si et seulement si, lorsque X_0 a Φ_π pour fonction caractéristique, il en va de même pour X_1 , c'est-à-dire si et seulement si, pour tout réel t ,

$$\Phi_\pi(t) = e^{-\frac{(1-r^2)t^2}{2}} \Phi_\pi(rt) \iff \Phi_\pi(t)e^{\frac{t^2}{2}} = \Phi_\pi(rt)e^{\frac{(rt)^2}{2}}.$$

Notons $\psi(t) := \Phi_\pi(t)e^{t^2/2}$. La fonction ψ est continue sur $\mathbb{R}$, vérifie $\psi(0) = 1$ et $\psi(rt) = \psi(t)$ pour tout réel t . On en déduit que $\psi(r^n t) = \psi(t)$ pour tout n , or $|r| < 1$ donc par continuité en 0 il

s'ensuit que $\psi(t) = 1$ pour tout réel t , ce qui signifie que la loi normale standard est la seule loi invariante.

2. Quelle est la loi μ_n de X_n ?

Correction : Rappelons que si $Y_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ est indépendante de $Y_2 \sim \mathcal{N}(m_2, \sigma_2^2)$, alors $Y_1 + Y_2 \sim \mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$. On montre alors sans difficulté, par exemple par récurrence, que

$$X_n \sim \mathcal{N}(r^n x, 1 - r^{2n}).$$

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 $H(\mu, \nu)^2 \leq 2\|\mu - \nu\|$.

Correction : Partant de

$$\|\mu - \nu\| = \frac{1}{2} \|f - g\|_1 = \frac{1}{2} \int_{\mathbb{R}} |f(x) - g(x)| dx = \frac{1}{2} \int_{\mathbb{R}} \left| \sqrt{f(x)} + \sqrt{g(x)} \right| \times \left| \sqrt{f(x)} - \sqrt{g(x)} \right| dx,$$

l'inégalité de Cauchy-Schwarz donne

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

Or $(a + b)^2 \leq 2(a^2 + b^2)$ donc, puisque f et g sont des densités :

$$\int_{\mathbb{R}} \left| \sqrt{f(x)} + \sqrt{g(x)} \right|^2 dx \leq 2 \left(\int_{\mathbb{R}} f(x) dx + \int_{\mathbb{R}} g(x) dx \right) = 4,$$

et l'inégalité $\|\mu - \nu\| \leq H(\mu, \nu)$ en découle. Pour la seconde inégalité, $H(\mu, \nu)^2 \leq 2\|\mu - \nu\|$ équivaut à

$$\int_{\mathbb{R}} \left(|f(x) - g(x)| - \left(\sqrt{f(x)} - \sqrt{g(x)} \right)^2 \right) dx \geq 0.$$

Introduisons les deux réels positifs $a = \sqrt{f(x)}$ et $b = \sqrt{g(x)}$, alors l'intégrand s'écrit

$$|a^2 - b^2| - (a - b)^2 = |a^2 - b^2| - |a - b|^2 = |a - b|(a + b - |a - b|) = 2|a - b| \min(a, b) \geq 0,$$

ce qui permet de conclure.

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

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

Correction : De façon générale on a

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

donc il s'agit de prouver que

$$\int_{\mathbb{R}} \sqrt{f(x)g(x)} dx = \sqrt{2 \frac{\sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2}} \exp \left(-\frac{(m_1 - m_2)^2}{4(\sigma_1^2 + \sigma_2^2)} \right).$$

Comme toujours, cela revient à écrire $\sqrt{f(x)g(x)}$ comme la densité d'une gaussienne à un facteur multiplicatif près. Allons-y :

$$\sqrt{f(x)g(x)} = \frac{1}{\sqrt{2\pi\sigma_1\sigma_2}} \exp \left(-\frac{1}{4} (\sigma_1^{-2}(x - m_1)^2 + \sigma_2^{-2}(x - m_2)^2) \right),$$

avec

$$\sigma_1^{-2}(x - m_1)^2 + \sigma_2^{-2}(x - m_2)^2 = (\sigma_1^{-2} + \sigma_2^{-2}) \left(x - \frac{\sigma_1^{-2}m_1 + \sigma_2^{-2}m_2}{\sigma_1^{-2} + \sigma_2^{-2}} \right)^2 + R,$$

où, après simplifications,

$$R = \frac{(m_1 - m_2)^2}{\sigma_1^2 + \sigma_2^2}.$$

Comme par ailleurs

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma_1\sigma_2}} \exp \left(-\frac{\sigma_1^{-2} + \sigma_2^{-2}}{4} \left(x - \frac{\sigma_1^{-2}m_1 + \sigma_2^{-2}m_2}{\sigma_1^{-2} + \sigma_2^{-2}} \right)^2 \right) dx = \sqrt{2 \frac{\sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2}},$$

la relation est établie.

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.

Correction : Nous avons vu que $X_n \sim \mathcal{N}(r^n x, 1 - r^{2n})$. Puisque $|r| < 1$, la variance est strictement positive pour tout $n \geq 1$, donc

$$\frac{1}{2} H(\mathcal{N}(r^n x, 1 - r^{2n}), \mathcal{N}(0, 1))^2 \leq \|\mu_n - \pi\| \leq H(\mathcal{N}(r^n x, 1 - r^{2n}), \mathcal{N}(0, 1)),$$

avec

$$H(\mathcal{N}(r^n x, 1 - r^{2n}), \mathcal{N}(0, 1))^2 = 2 \left\{ 1 - \sqrt{2 \frac{\sqrt{1 - r^{2n}}}{2 - r^{2n}}} \exp \left(-\frac{x^2 r^{2n}}{4(2 - r^{2n})} \right) \right\}.$$

Puisque (r^n) tend vers 0 lorsque n tend vers l'infini, le calcul donne

$$H(\mathcal{N}(r^n x, 1 - r^{2n}), \mathcal{N}(0, 1))^2 \underset{n \rightarrow \infty}{\sim} \frac{x^2}{4} r^{2n},$$

et par conséquent

$$H(\mathcal{N}(r^n x, 1 - r^{2n}), \mathcal{N}(0, 1)) \underset{n \rightarrow \infty}{\sim} \frac{|x|}{2} |r|^n.$$

La vitesse de convergence en variation totale de μ_n vers la loi stationnaire est donc au moins géométrique de raison $|r|$ et d'autant plus rapide que la condition initiale $X_0 = x$ est proche de l'origine.

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 $[-1, 1]$.

1. On note P le noyau de Metropolis, x un réel et B un borélien. Rappeler ce que vaut $P(x, B)$.

Correction : On a $P(x, dy) = p(x, y)dy + r(x)\delta_x(dy)$ avec, puisque q est paire,

$$p(x, y) = \alpha(x, y)q(y - x) = q(y - x) \min \left(1, \frac{\pi(y)q(y - x)}{\pi(x)q(x - y)} \right) = q(y - x) \min \left(1, \frac{\pi(y)}{\pi(x)} \right),$$

et

$$r(x) = 1 - \int_{\mathbb{R}} p(x, y)dy \geq 1 - \int_{\mathbb{R}} q(y - x)dy = 0.$$

Ainsi, pour tout borélien B ,

$$P(x, B) = \int_B P(x, dy) = r(x)\mathbf{1}_B(x) + \int_B p(x, y)dy.$$

-
2. Montrer que pour tout borélien B tel que $\lambda(B \cap [x - 1; x + 1]) > 0$, on a $P(x, B) > 0$.

Correction : Notons $J := B \cap [x - 1; x + 1]$, alors $\lambda(J) > 0$. De plus, comme $J \subseteq B$ et $P(x, \cdot)$ est une mesure de probabilité, il vient

$$P(x, B) \geq P(x, J) = r(x)\mathbf{1}_J(x) + \int_J p(x, y)dy \geq \int_J q(y - x) \min \left(1, \frac{\pi(y)}{\pi(x)} \right) dy,$$

car $r(x) \geq 0$. Comme $y \in J$ implique $|y - x| \leq 1$ et puisque $q \geq \epsilon$ sur $[-1, 1]$, on en déduit que

$$P(x, J) \geq \epsilon \int_J \min \left(1, \frac{\pi(y)}{\pi(x)} \right) dy.$$

Etant donné que $\lambda(J) > 0$ et que l'intégrand est strictement positif, il s'ensuit que $P(x, B) > 0$.

-
3. 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 $P(y, B) > 0$.

Correction : Puisque B est non négligeable, il existe x_B tel que $\lambda(B \cap [x_B - 1/4; x_B + 1/4]) > 0$. Supposons sans perte de généralité que $x \leq x_B$ et, pour $k \geq 1$, considérons les intervalles définis par $I_k := [x + (k - 1)/2; x + k/2]$. Comme la suite $(x + k/2)_{k \geq 1}$ est croissante et que les intervalles I_k sont de longueur $1/2$ tout comme $[x_B - 1/4; x_B + 1/4]$, il existe n tel que, pour tout $y \in I_n$

$$[x_B - 1/4; x_B + 1/4] \subseteq [y - 1; y + 1] \implies \lambda(B \cap [y - 1; y + 1]) \geq \lambda(B \cap [x_B - 1/4; x_B + 1/4]) > 0.$$

La question précédente assure alors que $P(y, B) > 0$. Bien sûr $\lambda(I_1 \cap [x - 1; x + 1]) > 0$ donc $P(x, I_1) > 0$ et plus généralement, pour tout k et tout $y \in I_k$, on a aussi $\lambda(I_{k+1} \cap [y - 1; y + 1]) > 0$ donc $P(y, I_{k+1}) > 0$.

-
4. Soient $\varphi_1, \dots, \varphi_n$ des fonctions boréliennes bornées et (X_n) une chaîne de Markov de noyau de transition P . Montrer que

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

Correction : En procédant par conditionnement, on peut toujours écrire

$$\mathbb{E}_x[\varphi_1(X_1) \dots \varphi_n(X_n)] = \mathbb{E}_x[\mathbb{E}_x[\varphi_1(X_1) \dots \varphi_n(X_n) | X_{n-1}, \dots, X_1]],$$

avec, par la propriété classique de l'espérance conditionnelle,

$$\mathbb{E}_x[\varphi_1(X_1) \dots \varphi_n(X_n) | X_{n-1}, \dots, X_1] = \varphi_1(X_1) \dots \varphi_{n-1}(X_{n-1}) \mathbb{E}_x[\varphi_n(X_n) | X_{n-1}, \dots, X_1].$$

Or (X_n) est une chaîne de noyau P donc $\text{Loi}(X_n | X_{n-1}, \dots, X_0) = \text{Loi}(X_n | X_{n-1}) = P(X_{n-1}, \cdot)$ et

$$\mathbb{E}_x[\varphi_n(X_n) | X_{n-1}, \dots, X_1] = \mathbb{E}[\varphi_n(X_n) | X_{n-1}] = (P\varphi_n)(X_{n-1}) = \int P(X_{n-1}, dx_n) \varphi_n(x_n),$$

de sorte que

$$\mathbb{E}_x[\varphi_1(X_1) \dots \varphi_n(X_n)] = \mathbb{E}_x[\varphi_1(X_1) \dots \varphi_{n-1}(X_{n-1}) (P\varphi_n)(X_{n-1})].$$

Il ne reste plus qu'à itérer ce raisonnement pour arriver à la formule annoncée.

5. En déduire que P est irréductible.

Correction : Soient x un réel, B un borélien non négligeable. Notons (X_n) une chaîne de noyau P et partant de $X_0 = x$. Avec les notations ci-dessus, on peut écrire

$$P^{n+1}(x, B) = \mathbb{P}_x(X_{n+1} \in B) \geq \mathbb{P}_x(X_1 \in I_1, \dots, X_n \in I_n, X_{n+1} \in B) =: p_x(n),$$

avec

$$p_x(n) = \mathbb{E}_x[\mathbf{1}_{I_1}(X_1) \dots \mathbf{1}_{I_n}(X_n) \mathbf{1}_B(X_{n+1})],$$

et la question précédente donne

$$p_x(n) = \int_{I_1} P(x, dx_1) \int_{I_2} P(x_1, dx_2) \dots \int_B P(x_n, dx_{n+1}),$$

ou encore

$$p_x(n) = \int_{I_1} P(x, dx_1) \int_{I_2} P(x_1, dx_2) \dots \int_{I_n} P(x_{n-1}, dx_n) P(x_n, B).$$

Fixons $x_{n-1} \in I_{n-1}$ et considérons l'intégrale

$$\varphi(x_{n-1}) := \int_{I_n} P(x_{n-1}, dx_n) P(x_n, B).$$

On sait deux choses : pour tout $x_n \in I_n$, on a $P(x_n, B) > 0$; d'autre part, la mesure de probabilité $P(x_{n-1}, dx_n)$ n'est pas identiquement nulle sur I_n car $P(x_{n-1}, I_n) > 0$. Ceci implique $\varphi(x_{n-1}) > 0$, et c'est vrai pour tout $x_{n-1} \in I_{n-1}$. En itérant ce raisonnement, on en déduit que $p_x(n) > 0$, donc $P^{n+1}(x, B) > 0$ et la chaîne est irréductible.

6. Que vaut $P(x, \{x\})$? Justifier qu'on peut trouver un réel x tel que $P(x, \{x\}) > 0$.

Correction : La première question donne

$$P(x, \{x\}) = r(x) \mathbf{1}_{\{x\}}(x) + \int_{\{x\}} p(x, y) dy = r(x) = 1 - \int_{\mathbb{R}} p(x, y) dy,$$

donc, q étant une densité sur $\mathbb{R}$,

$$P(x, \{x\}) = 1 - \int_{\mathbb{R}} q(y-x) \min\left(1, \frac{\pi(y)}{\pi(x)}\right) dy = \int_{\mathbb{R}} q(y-x) \left(1 - \min\left(1, \frac{\pi(y)}{\pi(x)}\right)\right) dy.$$

Or q est supérieure ou égale à $\epsilon > 0$ sur $[-1, 1]$ donc

$$P(x, \{x\}) \geq \epsilon \int_{x-1}^{x+1} \left(1 - \min\left(1, \frac{\pi(y)}{\pi(x)}\right)\right) dy.$$

La fonction intégrée étant supérieure ou égale à 0, $P(x, \{x\}) = 0$ implique que, pour presque tout $y \in [x-1; x+1]$, elle est nulle, c'est-à-dire que $\pi(y) \geq \pi(x)$ pour presque tout y dans $[x-1; x+1]$. En particulier, il existerait deux points $y_1 \in [x-1; x-1/2]$ et $z_1 \in [x+1/2; x+1]$ tels que $\pi(y_1) \geq \pi(x)$ et $\pi(z_1) \geq \pi(x)$. Raisonnons par l'absurde : si on avait $P(x, \{x\}) = 0$ pour tout x , on aurait en particulier $P(y_1, \{y_1\}) = 0$ et $P(z_1, \{z_1\}) = 0$ donc par transitivité π serait supérieure ou égale à $\pi(x)$ presque partout sur $[y_1-1, y_1+1]$ et sur $[z_1-1, z_1+1]$. De proche en proche, on pourrait couvrir $\mathbb{R}$ par une suite de segments sur lesquels π serait presque partout supérieure ou égale à $\pi(x) > 0$. Ceci est absurde car π est une densité. Ainsi on peut trouver un réel x tel que $P(x, \{x\}) > 0$.

7. En déduire que la chaîne est apériodique.

Correction : Si la chaîne n'était pas apériodique, il existerait $d \geq 2$ et une partition $E = \cup_{j=1}^d E_j$ tels que

$$\begin{cases} P(x, E_{j+1}) = 1 & \forall x \in E_j, \forall j \in \{1, \dots, d-1\} \\ P(x, E_1) = 1 & \forall x \in E_d. \end{cases}$$

Soit alors x tel que $P(x, \{x\}) > 0$ comme en question précédente. On peut sans perte de généralité supposer que x appartient à E_1 , or on devrait avoir $P(x, E_2) = 1$, ce qui contredit $P(x, \{x\}) > 0$.

8. 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 ?

Correction : Les résultats précédents s'appliquent avec $\epsilon = 1/2$ puisque $q(x) = \frac{1}{2}\mathbf{1}_{[-1, 1]}(x)$. Par conséquent la chaîne est π -irréductible, π -invariante et apériodique. Pour π -presque toute condition initiale x , on a donc

$$\|P^n(x, \cdot) - \pi\| \xrightarrow{n \rightarrow \infty} 0,$$

et pour toute fonction $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ telle que $\pi(|\varphi|) < \infty$

$$\frac{1}{n} \sum_{k=1}^n \varphi(X_k) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \pi(\varphi) = \int_{\mathbb{R}} \varphi(x) \pi(x) dx.$$

Puisque $\pi > 0$ sur $\mathbb{R}$, π est équivalente à la mesure de Lebesgue sur $\mathbb{R}$, donc on peut remplacer le π -presque toute condition initiale par presque toute condition initiale.

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 ?

Correction : Les configurations minimisant l'énergie sont celles pour lesquelles les spins sont tous égaux à $+1$ ou à -1 , configurations notées S^+ et S^- . Il y a donc deux minima globaux pour V .

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 $T \rightarrow +\infty$?

Correction : Lorsque $T \rightarrow +\infty$, μ_T tend vers la mesure uniforme : toutes les configurations deviennent équiprobables. A contrario, lorsque $T \rightarrow 0$, seuls les deux minima globaux auront une probabilité qui ne s'éteint pas, laquelle tend donc vers $1/2$ (principe de Laplace).

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.

Correction : Le nombre de configurations possibles est $2^{N \times N}$, qui correspond à la taille de l'espace d'états, lequel est donc fini mais immense dès que N n'est pas petit. Dès lors, il est exclu d'essayer de simuler directement selon μ_T , d'où le recours à une méthode de type MCMC dans la suite.

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 $\rho(S, S^{(i_0, j_0)})$.

Correction : Pour passer de S à $S^{(i_0, j_0)}$ via Q , il faut et il suffit d'avoir tiré le couple d'indices (i_0, j_0) , la réciproque étant tout aussi vraie, d'où

$$Q(S, S^{(i_0, j_0)}) = Q(S^{(i_0, j_0)}, S) = \frac{1}{N^2}.$$

Nous sommes donc dans le cas du Metropolis symétrique, avec

$$\rho(S, S^{(i_0, j_0)}) = \frac{\mu_T(S^{(i_0, j_0)})}{\mu_T(S)} = \exp \left(-\frac{1}{T} (V(S^{(i_0, j_0)}) - V(S)) \right),$$

c'est-à-dire

$$\rho(S, S^{(i_0, j_0)}) = \exp \left(-\frac{2S_{i_0 j_0}}{T} \sum_{(i,j) \sim (i_0, j_0)} S_{ij} \right),$$

donc une somme faisant intervenir au plus 4 termes (il y en a moins si (i_0, j_0) est au bord de la grille). La simplification vient du télescopage de tous les termes ne faisant pas intervenir le point (i_0, j_0) dans les sommes définissant $V(S^{(i_0, j_0)})$ et $V(S)$.

Remarque : Outre le fait qu'il ne subsiste qu'au plus 4 termes à sommer grâce au télescopage, on notera que l'épouvantable constante de normalisation Z_T a elle aussi disparu. Cette propriété générale de la méthode de Metropolis révèle ici tout son intérêt.

5. Expliquer brièvement pourquoi la chaîne de Markov ainsi construite est irréductible et apériodique.

Correction : Soient S et S' deux configurations quelconques et n_0 le nombre de sites en lesquelles elles diffèrent. On voit qu'il y a une probabilité non nulle d'aller de S à S' en exactement n_0 étapes, donc la chaîne est irréductible. Par ailleurs, partant de S^+ , toute configuration proposée par Q (laquelle contient donc un seul spin à -1) est moins vraisemblable pour μ_T , c'est-à-dire que la probabilité de rester sur place est strictement positive. Ainsi S^+ est apériodique, et puisque la chaîne est irréductible, elle est apériodique.
