

## TD 4 : ECHANTILLONNEUR DE GIBBS

### EXERCICE 1 (Cas d'école pour Gibbs)

On considère la densité

$$f(x, y) = C \exp \left( -\frac{y^2}{2} - \frac{x^2(1 + y + y^2)}{2} \right),$$

où  $C$  est la constante de normalisation.

1. Soient  $(x, y) \in \mathbb{R}^2$ . Quelle est la loi de  $X$  sachant  $Y = y$ ? Et celle de  $Y$  sachant  $X = x$ ?

**Correction :** Pour tout réel  $y$ , on a

$$f(x|y) \propto f(x, y) \propto \exp \left( -\frac{x^2(1 + y + y^2)}{2} \right),$$

donc

$$\text{Loi}(X|Y = y) = \mathcal{N} \left( 0, \frac{1}{1 + y + y^2} \right).$$

De même, pour tout réel  $x$ , l'écriture

$$f(x|y) \propto f(x, y) \propto \exp \left( -\frac{1 + x^2}{2} \left( y + \frac{x^2}{2(1 + x^2)} \right)^2 \right)$$

montre que

$$\text{Loi}(Y|X = x) = \mathcal{N} \left( -\frac{x^2}{2(1 + x^2)}, \frac{1}{1 + x^2} \right).$$

- 
2. En déduire un échantillonneur de Gibbs séquentiel de loi cible  $f$  produisant une chaîne de Markov  $((X_n, Y_n))_{n \geq 0}$ .

**Correction :** Un échantillonneur séquentiel de Gibbs de loi cible  $f$  est alors très facile à construire puisqu'il suffit de simuler des variables gaussiennes. Sachant  $(X_n, Y_n) = (x, y)$  :

— on simule

$$X' = x' \sim \mathcal{N} \left( 0, \frac{1}{1 + y + y^2} \right);$$

— ceci fait, on simule

$$Y' = y' \sim \mathcal{N} \left( -\frac{x'^2}{2(1 + x'^2)}, \frac{1}{1 + x'^2} \right);$$

— on pose  $(X_{n+1}, Y_{n+1}) = (x', y')$ .

Pour l'initialisation, on peut par exemple partir de  $X_0 = (0, 0)$ .

- 
3. La chaîne  $((X_n, Y_n))_{n \geq 0}$  ainsi construite vérifie-t-elle la convergence vers  $f$  en variation totale et le Théorème ergodique ?

**Correction :** Le support  $\mathbb{R}^2$  de la loi jointe étant le produit cartésien des supports des lois marginales, un résultat du cours assure que convergence en variation totale et Théorème ergodique sont vérifiés pour presque toute condition initiale  $X_0 = (x_0, y_0)$  appartenant au support de  $f$ , c'est-à-dire pour presque toute condition initiale dans  $\mathbb{R}^2$  puisque  $f$  étant strictement positive, la loi de densité  $f$  et la mesure de Lebesgue sont équivalentes.

---

### EXERCICE 2 (Somme d'exponentielles contrainte)

Soient  $X_1, \dots, X_d$  des variables exponentielles indépendantes de paramètres  $\lambda_1, \dots, \lambda_d > 0$  donnés. Soit  $S = X_1 + \dots + X_d$  leur somme et  $s > 0$  une constante fixée.

1. Donner la densité  $\pi$  du vecteur aléatoire  $(X_1, \dots, X_d)$  sachant  $S \geq s$ .

**Correction :** La densité  $\pi$  est la restriction de la densité du vecteur  $(X_1, \dots, X_d)$  au domaine

$$\mathcal{S} = \left\{ (x_1, \dots, x_d) \in \mathbb{R}_+^d : x_1 + \dots + x_d \geq s \right\},$$

c'est-à-dire

$$\pi(x_1, \dots, x_d) = \left( \prod_{k=1}^d \lambda_k e^{-\lambda_k x_k} \mathbf{1}_{x_k \geq 0} \right) \frac{\mathbf{1}_{x_1 + \dots + x_d \geq s}}{\mathbb{P}(S \geq s)}.$$

2. Soit  $x_{-1} := (x_2, \dots, x_d) \in \mathbb{R}_+^{d-1}$  fixé. Déterminer la densité conditionnelle  $\pi(\cdot | x_{-1})$  de  $\text{Loi}(X_1 | X_{-1} = x_{-1}, S \geq s)$ . Quelle loi reconnaissez-vous ?

**Correction :** Pour tout  $x_{-1} := (x_2, \dots, x_d) \in \mathbb{R}_+^{d-1}$  fixé, la densité conditionnelle s'écrit

$$\pi(x_1 | x_{-1}) \propto \pi(x_1, x_{-1}) = \pi(x_1, \dots, x_d) \propto e^{-\lambda_1 x_1} \mathbf{1}_{x_1 \geq 0} \mathbf{1}_{x_1 + \dots + x_d \geq s} = e^{-\lambda_1 x_1} \mathbf{1}_{x_1 \geq (s - (x_2 + \dots + x_d))_+},$$

avec la notation  $x_+ := \max(x, 0)$ . On reconnaît une loi exponentielle translatée :

$$\text{Loi}(X_1 | X_{-1} = x_{-1}) = (s - (x_2 + \dots + x_d))_+ + \text{Exp}(\lambda_1).$$

3. En déduire un échantillonneur de Gibbs séquentiel de loi cible  $\pi$ .

**Correction :** Pour l'initialisation, on peut par exemple partir du point  $X_0 = (s/d, \dots, s/d)$ . Sachant  $X_n = (x_1, \dots, x_d)$ , on simule tout d'abord

$$X'_1 = x'_1 \sim (s - (x_2 + \dots + x_d))_+ + \text{Exp}(\lambda_1),$$

puis

$$X'_2 = x'_2 \sim (s - (x'_1 + x_3 + \dots + x_d))_+ + \text{Exp}(\lambda_2),$$

et ainsi de suite sur les  $d$  coordonnées jusqu'à obtenir  $X_{n+1} = (x'_1, \dots, x'_d)$ .

4. Le support  $\mathcal{S}(\pi)$  de la loi jointe  $\pi$  correspond-il au produit cartésien des supports des marginales ?

**Correction :** Comme vu-ci-dessus, le support de  $\pi$  est

$$\mathcal{S}(\pi) = \left\{ (x_1, \dots, x_d) \in \mathbb{R}_+^d : x_1 + \dots + x_d \geq s \right\}.$$

Pour tout  $x_1 \in \mathbb{R}_+$ , notons

$$\mathcal{S}(x_1) := \left\{ (x_2, \dots, x_d) \in \mathbb{R}_+^{d-1} : x_1 + \dots + x_d \geq s \right\}.$$

Bien entendu, quel que soit  $x_1 \geq 0$ , l'ensemble  $\mathcal{S}(x_1)$  est non négligeable pour la mesure de Lebesgue sur  $\mathbb{R}^{d-1}$ , si bien que l'intégrale d'une fonction strictement positive sur  $\mathcal{S}(x_1)$  est non-nulle. Ainsi, la marginale en  $x_1$ , i.e.

$$\pi_1(x_1) = \lambda_1 e^{-\lambda_1 x_1} \int_{\mathcal{S}(x_1)} \left( \prod_{k=2}^d \lambda_k e^{-\lambda_k x_k} \right) dx_2 \dots dx_d$$

est strictement positive, d'où  $\mathcal{S}(\pi_1) = \mathbb{R}_+$ , et

$$\mathcal{S}(\pi) \subsetneq \mathcal{S}(\pi_1) \times \cdots \times \mathcal{S}(\pi_d).$$

Par conséquent, pour montrer que la chaîne est irréductible, on ne peut pas appliquer le critère des supports.

5. On veut montrer que la chaîne  $(X_n)$  ainsi construite, de noyau noté  $P$ , est irréductible. Pour simplifier les notations, on suppose  $d = 2$  et on considère un borélien  $B$  de  $\mathbb{R}_+^2$  tel que  $\pi(B) > 0$ .

- (a) Soit  $x \in \mathcal{S}(\pi)$ . Rappeler la densité du noyau  $P$  par rapport à la mesure de Lebesgue, c'est-à-dire la fonction  $p(x, x') = p((x_1, x_2), (x'_1, x'_2))$  telle que  $P(x, dx') = p(x, x')dx'$ .

**Correction :** On a, d'après le cours,

$$p(x, x') = \pi(x'_1|x_2)\pi(x'_2|x'_1).$$

- (b) Soient  $x = (x_1, x_2) \in \mathcal{S}(\pi) \cap (\mathbb{R}_+ \times [s, \infty])$  et  $B$  un borélien de  $\mathbb{R}^2$  tel que  $\pi(B) > 0$ . Montrer que  $P(x, B) > 0$ .

**Correction :** Soit  $x = (x_1, x_2) \in \mathcal{S}(\pi) \cap (\mathbb{R}_+ \times [s, \infty])$ . La densité du noyau de transition  $P$  vérifie ici :

$$\begin{aligned} \forall x' \in (\mathbb{R}_+)^2, \quad p(x, x') &= \pi(x'_1|x_2)\pi(x'_2|x'_1) \propto e^{-\lambda_1 x'_1} \mathbf{1}_{x'_1 \geq 0} e^{-\lambda_2 x'_2} \mathbf{1}_{x'_2 \geq (s-x'_1)_+} \\ &\propto e^{-\lambda_1 x'_1} e^{-\lambda_2 x'_2} \mathbf{1}_{x' \in \mathcal{S}(\pi)}. \end{aligned}$$

Ainsi, le support de la loi de probabilité  $P(x, \cdot)$  est  $\mathcal{S}(\pi)$ . Puisque, par hypothèse  $\pi(B) > 0$ , on en déduit que  $\lambda_2(B \cap \mathcal{S}(\pi)) > 0$ , d'où  $P(x, B) > 0$ .

Un dessin permet de comprendre ce qui se passe. Partant de  $x = (x_1, x_2) \in \mathcal{S}(\pi) \cap (\mathbb{R}_+ \times [s, \infty])$ , on commence par simuler  $X'_1 \sim \text{Exp}(\lambda_1)$ , de support  $\mathbb{R}_+$ . Puis, sachant  $X'_1 = x'_1$ , on simule  $X'_2 \sim (s - x'_1)_+ + \text{Exp}(\lambda_2)$ , de support  $(s - x'_1)_+ + \mathbb{R}_+$ . Par conséquent le support de la loi de probabilité  $P(x, \cdot)$  est  $\mathcal{S}(\pi)$ .

- (c) En déduire que, pour tout  $x = (x_1, x_2) \in \mathcal{S}(\pi)$  et tout borélien  $B$  de  $\mathbb{R}_+^2$  tel que  $\pi(B) > 0$ , on a  $P^2(x, B) > 0$ .

**Correction :** De nouveau, un dessin permet d'éclairer les choses. Partant de  $x = (x_1, x_2) \in \mathcal{S}(\pi)$ , on commence par simuler  $X'_1 \sim (s - x_2)_+ + \text{Exp}(\lambda_1)$ , de support  $(s - x_2)_+ + \mathbb{R}_+$ . Puis, sachant  $X'_1 = x'_1$ , on simule  $X'_2 \sim (s - x'_1)_+ + \text{Exp}(\lambda_2)$ , de support  $(s - x'_1)_+ + \mathbb{R}_+$ . Par conséquent le support de la loi de probabilité  $P(x, \cdot)$  est  $\mathcal{S}(\pi) \cap ((s - x_2)_+, \infty \times \mathbb{R}_+)$ . Ainsi, partant de  $x$ , on peut aller n'importe où dans  $\mathcal{S}(\pi)$  sauf éventuellement dans une étroite bande verticale à gauche. En notant  $\mathcal{A} = [(s - x_2)_+, \infty \times [s, \infty])$ , qui est inclus à la fois dans  $\mathcal{S}(\pi) \cap ((s - x_2)_+, \infty \times \mathbb{R}_+)$  et dans  $\mathcal{S}(\pi) \cap (\mathbb{R}_+ \times [s, \infty])$ , on comprend donc que l'on a une probabilité strictement positive d'aller dans  $\mathcal{A}$  depuis  $x$ , puis de  $\mathcal{A}$ , d'aller dans  $B$  (d'après la question précédente).

Plus formellement, nous avons, par définition :

$$P^2(x, dx'') = \int P(x, dx') P(x', dx'') = \left( \int p(x, x') p(x', x'') dx' \right) dx''.$$

Ainsi, en appliquant Fubini-Tonelli, on obtient :

$$\begin{aligned}
 P^2(x, B) &= \int \mathbf{1}_B(x'') P^2(x, dx'') = \int \mathbf{1}_B(x'') \left( \int p(x, x') p(x', x'') dx' \right) dx'' \\
 &= \int p(x, x') \left( \int \mathbf{1}_B(x'') p(x', x'') dx'' \right) dx' \\
 &= \int p(x, x') P(x', B) dx' \\
 &\geq \int_{S(\pi) \cap (\mathbb{R}_+ \times [s, \infty])} p(x, x') P(x', B) dx' \\
 &\geq \int_{\mathcal{A}} p(x, x') P(x', B) dx'.
 \end{aligned}$$

Or, pour tout  $x' \in \mathcal{A}$ ,  $p(x, x') > 0$  et puisque  $x' \in S(\pi) \cap (\mathbb{R}_+ \times [s, \infty])$ , la question précédente assure que  $P(x', B) > 0$ . Ainsi  $P^2(x, B) > 0$ .

6. La chaîne  $(X_n)$  ainsi construite vérifie-t-elle la convergence vers  $\pi$  en variation totale et le Théorème ergodique ?

**Correction :** Puisque  $P^2(x, B) > 0$  pour tout  $x = (x_1, x_2) \in \mathcal{S}(\pi)$  et tout  $B$  tel que  $\pi(B) > 0$ , la chaîne est  $\pi$ -irréductible. On peut montrer par l'absurde qu'elle est aussi apériodique. La chaîne vit dans l'espace  $E = \mathcal{S}(\pi)$ . Supposons alors qu'il existe  $d \geq 2$  et une partition de  $\mathcal{S}(\pi) = \cup_{j=1}^d E_j$  tels que

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

Supposons pour commencer  $d = 2$ , alors pour tout  $x \in E_1$ , le support de  $P(x, \cdot)$  est  $E_2$  donc

$$P^2(x, E_1) = \int_{E_2} P(x, dx') P(x', E_1) = \int_{E_2} P(x, dx') = 0.$$

Mais par la propriété  $P^2(x, B) > 0$  pour tout  $B$  tel que  $\pi(B) > 0$ , ceci implique  $\pi(E_2) = 0$ . Prenons alors  $x \in E_2$  : pour les mêmes raisons que ci-dessus, on a  $P^2(x, E_2) = 1$ , ce qui implique  $\pi(E_1) = 0$  et contredit le fait que  $\pi(E_1) + \pi(E_2) = 1$ . Le cas  $d > 2$  se traite de la même façon. Par conséquent, la chaîne est apériodique. Par construction de l'échantillonneur de Gibbs, elle est bien entendu également  $\pi$ -invariante. Ainsi, pour  $\pi$ -presque toute condition initiale, c'est-à-dire pour presque tout  $x$  dans  $\mathcal{S}(\pi)$ , la convergence vers  $\pi$  en variation totale et le Théorème ergodique sont satisfaits.

### EXERCICE 3 (Modèle de mélange et variables latentes)

Soient  $(a, b) \in (\mathbb{R}_+^*)^2$ ,  $(m_0, m_1) \in \mathbb{R}^2$ ,  $(\sigma_0, \sigma_1) \in (\mathbb{R}_+^*)^2$  connus et fixés. On considère le schéma bayésien suivant : la loi a priori sur  $\theta$  est  $\Pi = \text{Beta}(a, b)$ . Sachant  $\theta = \theta$ , les observations  $\mathbf{X} = (X_1, \dots, X_N)$  sont i.i.d. selon le mélange de gaussiennes de densité

$$p_\theta(x) = (1 - \theta)\phi_0(x) + \theta\phi_1(x) = (1 - \theta) \times \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{(x-m_0)^2}{2\sigma_0^2}} + \theta \times \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(x-m_1)^2}{2\sigma_1^2}}.$$

1. Déterminer la loi a posteriori, c'est-à-dire  $\text{Loi}(\theta|\mathbf{X})$ . Est-ce une loi classique ?

**Correction :** En notant comme d'habitude  $\pi(\theta|\mathbf{X})$  la densité de la loi a posteriori, on a

$$\pi(\theta|\mathbf{X}) \propto \theta^{a-1}(1-\theta)^{b-1} \mathbf{1}_{]0,1[}(\theta) \prod_{j=1}^N p_{\theta}(X_j) = \theta^{a-1}(1-\theta)^{b-1} \mathbf{1}_{]0,1[}(\theta) \prod_{j=1}^N \{(1-\theta)\phi_0(X_j) + \theta\phi_1(X_j)\},$$

qui n'est pas une loi classique. En particulier, ce n'est pas une loi Beta. En d'autres termes, la famille des lois Beta n'est pas conjuguée pour le modèle de mélange gaussien.

2. On considère le schéma bayésien hiérarchique suivant : la loi a priori sur  $\boldsymbol{\theta}$  est encore  $\Pi = \text{Beta}(a, b)$ . Puis, sachant  $\boldsymbol{\theta} = \theta$ , les variables  $\mathbf{Z} = (Z_1, \dots, Z_N) \in \{0, 1\}^N$  sont i.i.d. selon la loi de Bernoulli  $\text{Ber}(\theta)$ . Enfin, sachant  $\mathbf{Z} = \mathbf{z}$  et  $\boldsymbol{\theta} = \theta$ , les observations  $\mathbf{X} = (X_1, \dots, X_N)$  sont indépendantes avec

$$\text{Loi}(X_j|\mathbf{Z} = \mathbf{z}, \boldsymbol{\theta} = \theta) = \text{Loi}(X_j|Z_j = z_j, \boldsymbol{\theta} = \theta) = \text{Loi}(X_j|Z_j = z_j) = \mathcal{N}(m_{z_j}, \sigma_{z_j}^2).$$

Les variables aléatoires  $Z_j$  sont appelées variables latentes (car non observées).

- (a) Soit  $\varphi : \mathbb{R} \rightarrow \mathbb{R}_+$  une fonction test, montrer que

$$\mathbb{E}[\varphi(X_1)|\boldsymbol{\theta}] = \int_{\mathbb{R}} \varphi(x) p_{\boldsymbol{\theta}}(x) dx.$$

Qu'en déduisez-vous pour la loi a posteriori  $\text{Loi}(\boldsymbol{\theta}|\mathbf{X})$  de ce schéma bayésien hiérarchique ?

**Correction :** On commence par appliquer la formule des espérances conditionnelles emboîtées :

$$\mathbb{E}[\varphi(X_1)|\boldsymbol{\theta}] = \mathbb{E}[\mathbb{E}[\varphi(X_1)|Z_1, \boldsymbol{\theta}]|\boldsymbol{\theta}],$$

or par hypothèse  $\text{Loi}(X_1|Z_1, \boldsymbol{\theta}) = \text{Loi}(X_1|Z_1) = \mathcal{N}(m_{Z_1}, \sigma_{Z_1}^2)$  donc

$$\mathbb{E}[\varphi(X_1)|Z_1, \boldsymbol{\theta}] = \mathbb{E}[\varphi(X_1)|Z_1] = \int_{\mathbb{R}} \varphi(x) \phi_{Z_1}(x) dx.$$

On note par ailleurs que

$$\phi_{Z_1}(x) = \phi_0(x) \mathbf{1}_{Z_1=0} + \phi_1(x) \mathbf{1}_{Z_1=1},$$

si bien que

$$\mathbb{E}[\varphi(X_1)|Z_1] = \left( \int_{\mathbb{R}} \varphi(x) \phi_0(x) dx \right) \mathbf{1}_{Z_1=0} + \left( \int_{\mathbb{R}} \varphi(x) \phi_1(x) dx \right) \mathbf{1}_{Z_1=1} =: \alpha_0 \mathbf{1}_{Z_1=0} + \alpha_1 \mathbf{1}_{Z_1=1},$$

donc, tenant compte du fait que  $\text{Loi}(Z_1|\boldsymbol{\theta}) = \text{Ber}(\boldsymbol{\theta})$ , il vient

$$\mathbb{E}[\varphi(X_1)|\boldsymbol{\theta}] = \mathbb{E}[\alpha_0 \mathbf{1}_{Z_1=0} + \alpha_1 \mathbf{1}_{Z_1=1}|\boldsymbol{\theta}] = \alpha_0 \mathbb{P}(Z_1 = 0|\boldsymbol{\theta}) + \alpha_1 \mathbb{P}(Z_1 = 1|\boldsymbol{\theta}) = (1 - \boldsymbol{\theta})\alpha_0 + \boldsymbol{\theta}\alpha_1,$$

ce qui donne bien

$$\mathbb{E}[\varphi(X_1)|\boldsymbol{\theta}] = \int_{\mathbb{R}} \varphi(x) ((1 - \boldsymbol{\theta})\phi_0(x) + \boldsymbol{\theta}\phi_1(x)) dx = \int_{\mathbb{R}} \varphi(x) p_{\boldsymbol{\theta}}(x) dx.$$

Ainsi, sachant  $\boldsymbol{\theta}$ , les observations  $X_j$  ont toutes la même loi et celle-ci correspond à celle du modèle bayésien de la Question 1, c'est-à-dire un mélange de gaussiennes. De plus, sachant  $\boldsymbol{\theta}$ , il est intuitivement clair que les  $X_j$  sont indépendantes, ce que l'on peut le vérifier par les fonctions tests :

$$\mathbb{E}[\varphi_1(X_1) \dots \varphi_N(X_N)|\boldsymbol{\theta}] = \mathbb{E}[\mathbb{E}[\varphi_1(X_1) \dots \varphi_N(X_N)|\mathbf{Z}, \boldsymbol{\theta}]|\boldsymbol{\theta}] = \mathbb{E}[\mathbb{E}[\varphi_1(X_1) \dots \varphi_N(X_N)|\mathbf{Z}|\boldsymbol{\theta}],$$

avec, par indépendance conditionnelle des  $X_j$  sachant les  $Z_j$ ,

$$\mathbb{E}[\varphi_1(X_1) \dots \varphi_N(X_N) | \mathbf{Z}] = \prod_{j=1}^N \mathbb{E}[\varphi_j(X_j) | Z_j] =: \prod_{j=1}^N \psi_j(Z_j).$$

De même, par indépendance conditionnelle des  $Z_j$  sachant  $\boldsymbol{\theta}$ ,

$$\mathbb{E}[\psi_1(Z_1) \dots \psi_N(Z_N) | \boldsymbol{\theta}] = \prod_{j=1}^N \mathbb{E}[\psi_j(Z_j) | \boldsymbol{\theta}] = \prod_{j=1}^N \mathbb{E}[\mathbb{E}[\varphi_j(X_j) | Z_j] | \boldsymbol{\theta}],$$

et on refait le film à l'envers :

$$\prod_{j=1}^N \mathbb{E}[\mathbb{E}[\varphi_j(X_j) | Z_j] | \boldsymbol{\theta}] = \prod_{j=1}^N \mathbb{E}[\mathbb{E}[\varphi_j(X_j) | Z_j, \boldsymbol{\theta}] | \boldsymbol{\theta}] = \prod_{j=1}^N \mathbb{E}[\varphi_j(X_j) | \boldsymbol{\theta}]$$

pour conclure que

$$\mathbb{E}[\varphi_1(X_1) \dots \varphi_N(X_N) | \boldsymbol{\theta}] = \prod_{j=1}^N \mathbb{E}[\varphi_j(X_j) | \boldsymbol{\theta}],$$

ce qui prouve que, sachant  $\boldsymbol{\theta}$ , les observations  $X_j$  sont indépendantes. Bref, tout ça pour dire que, sachant  $\boldsymbol{\theta}$ , les observations  $X_j$  sont i.i.d. et de densité  $p_{\boldsymbol{\theta}}(x)$ , le mélange gaussien introduit en début d'exercice. Puisque la loi a priori sur  $\boldsymbol{\theta}$  est également la même qu'en Question 1, on en déduit que la loi a posteriori  $\text{Loi}(\boldsymbol{\theta} | \mathbf{X})$  est la même qu'en Question 1.

- (b) Quelle est la loi de  $\boldsymbol{\theta}$  sachant  $\mathbf{Z}$ ? Quelle est la loi de  $\boldsymbol{\theta}$  sachant  $(\mathbf{Z}, \mathbf{X})$ ?

**Correction :** On a affaire à un schéma bayésien classique où la loi a priori sur  $\boldsymbol{\theta}$  est une loi Beta et, sachant  $\boldsymbol{\theta}$ , les  $Z_j$  sont i.i.d. selon une Bernoulli de paramètre  $\boldsymbol{\theta}$ , donc la loi de  $\boldsymbol{\theta}$  sachant  $\mathbf{Z}$  admet une densité par rapport à la mesure de Lebesgue, laquelle se calcul comme d'habitude :

$$\pi(\boldsymbol{\theta} | \mathbf{Z}) \propto \theta^{a-1} (1-\theta)^{b-1} \mathbf{1}_{]0,1[}(\theta) \prod_{j=1}^N \theta^{Z_j} (1-\theta)^{1-Z_j} = \theta^{a+N\bar{Z}_N-1} (1-\theta)^{b+N-N\bar{Z}_N-1} \mathbf{1}_{]0,1[}(\theta),$$

ce qui signifie que

$$\text{Loi}(\boldsymbol{\theta} | \mathbf{Z}) = \text{Beta}(a + N\bar{Z}_N, b + N - N\bar{Z}_N).$$

Pour la loi de  $\boldsymbol{\theta}$  sachant  $(\mathbf{Z}, \mathbf{X})$ , notons  $f(\boldsymbol{\theta}, \mathbf{z}, \mathbf{x})$  la densité du triplet  $(\boldsymbol{\theta}, \mathbf{Z}, \mathbf{X})$  par rapport à la mesure dominante  $\lambda \otimes \mu^{\otimes N} \otimes \lambda^{\otimes N}$ , où  $\lambda$  et  $\mu$  désignent respectivement les mesures de Lebesgue sur  $\mathbb{R}$  et de comptage sur  $\{0, 1\}$ . Alors, avec des notations évidentes,

$$\pi(\boldsymbol{\theta} | \mathbf{z}, \mathbf{x}) = f(\boldsymbol{\theta} | \mathbf{z}, \mathbf{x}) \propto f(\boldsymbol{\theta}, \mathbf{z}, \mathbf{x}) = f(\boldsymbol{\theta}) f(\mathbf{z} | \boldsymbol{\theta}) f(\mathbf{x} | \mathbf{z}, \boldsymbol{\theta}),$$

or par construction  $f(\mathbf{x} | \mathbf{z}, \boldsymbol{\theta}) = f(\mathbf{x} | \mathbf{z})$ , qui ne dépend pas de  $\boldsymbol{\theta}$  donc passe à la trappe du " $\propto$ ", d'où

$$\pi(\boldsymbol{\theta} | \mathbf{z}, \mathbf{x}) \propto f(\boldsymbol{\theta}) f(\mathbf{z} | \boldsymbol{\theta}) \propto \pi(\boldsymbol{\theta} | \mathbf{z}),$$

c'est-à-dire que  $\text{Loi}(\boldsymbol{\theta} | \mathbf{Z}, \mathbf{X}) = \text{Loi}(\boldsymbol{\theta} | \mathbf{Z})$ , ce qui était intuitivement clair : si l'on connaissait les variables latentes  $Z_j$ , les observations  $X_j$  n'apporteraient aucune information de plus sur le paramètre  $\boldsymbol{\theta}$ .

- (c) Quelle est la loi de  $Z_1$  sachant  $(\theta, \mathbf{X})$  ?

**Correction :** Les observations  $X_j$  étant conditionnellement indépendantes sachant les  $Z_j$  et les  $Z_j$  étant elles-mêmes conditionnellement indépendantes sachant  $\theta$ , on commence par noter que<sup>1</sup>  $\text{Loi}(Z_1|\theta, \mathbf{X}) = \text{Loi}(Z_1|\theta, X_1)$ . Ensuite, puisque  $Z_1$  ne peut prendre que les valeurs 0 et 1, la loi de  $Z_1$  sachant  $(\theta, \mathbf{X})$  est bien entendu une loi de Bernoulli dont le paramètre dépend du couple  $(\theta, \mathbf{X})$ . Nous modifions un peu les notations de la réponse à la question précédente comme suit :  $f(\theta, z_1, x_1)$  est la densité du triplet  $(\theta, Z_1, X_1)$  par rapport à la mesure dominante  $\lambda \otimes \mu \otimes \lambda$ . Pour  $z_1 \in \{0, 1\}$  on a donc, par les mêmes arguments que ci-dessus,

$$\mathbb{P}(Z_1 = z_1|\theta, x_1) = f(z_1|\theta, x_1) \propto f(\theta, z_1, x_1) \propto f(x_1|z_1, \theta)f(z_1|\theta) = f(x_1|z_1)f(z_1|\theta),$$

ce qui donne

$$\mathbb{P}(Z_1 = z_1|\theta, x_1) \propto \theta^{z_1}(1 - \theta)^{1-z_1}\phi_{z_1}(x_1),$$

et signifie que

$$\text{Loi}(Z_1|\theta, \mathbf{X}) = \text{Ber}\left(\frac{\theta\phi_1(X_1)}{(1 - \theta)\phi_0(X_1) + \theta\phi_1(X_1)}\right).$$

- (d) Quelle est la loi de  $Z_1$  sachant  $(\theta, \mathbf{X}, Z_2, \dots, Z_N)$  ?

**Correction :** Pour les mêmes raisons que précédemment et en reprenant les notations du cours, on a

$$\text{Loi}(Z_1|\theta, Z_{-1}, \mathbf{X}) := \text{Loi}(Z_1|\theta, \mathbf{X}, Z_2, \dots, Z_N) = \text{Loi}(Z_1|\theta, \mathbf{X}) = \text{Loi}(Z_1|\theta, X_1).$$

- (e) En déduire un échantillonneur de Gibbs permettant de construire, sachant  $\mathbf{X}$ , une chaîne de Markov  $((\theta_n, \mathbf{Z}_n))_{n \geq 0}$  admettant pour loi stationnaire la loi a posteriori du couple  $(\theta, \mathbf{Z})$ , c'est-à-dire  $\text{Loi}((\theta, \mathbf{Z})|\mathbf{X})$ .

**Correction :** On rappelle que les observations  $X_j$  sont connues et fixées une fois pour toutes. On part d'une condition initiale  $(\theta_0, \mathbf{Z}_0) = (\theta_0, \mathbf{z}_0) \in ]0, 1[ \times \{0, 1\}^N$  arbitraire : on peut par exemple simuler  $\theta_0 \sim \text{Beta}(a, b)$  et une fois ceci fait simuler  $\mathbf{Z}_0 \sim \text{Ber}(\theta_0)^{\otimes N}$ . Sachant  $(\theta_n, \mathbf{Z}_n) = (\theta, \mathbf{z}) = (\theta, (z_1, \dots, z_N))$ , une étape de l'échantillonneur de Gibbs séquentiel fonctionne comme suit :

— On simule

$$\theta' = \theta' \sim \text{Beta}(a + N\bar{z}_N, b + N - N\bar{z}_N).$$

— On simule

$$Z'_1 = z'_1 \sim \text{Ber}\left(\frac{\theta'\phi_1(X_1)}{(1 - \theta')\phi_0(X_1) + \theta'\phi_1(X_1)}\right).$$

— ...

— On simule

$$Z'_N = z'_N \sim \text{Ber}\left(\frac{\theta'\phi_1(X_N)}{(1 - \theta')\phi_0(X_N) + \theta'\phi_1(X_N)}\right).$$

— On pose

$$(\theta_{n+1}, \mathbf{Z}_{n+1}) = (\theta', \mathbf{z}') = (\theta', (z'_1, \dots, z'_N)).$$

1. Pour l'établir rigoureusement, on peut à nouveau passer par les fonctions tests.

- (f) La chaîne  $((\theta_n, \mathbf{Z}_n))_{n \geq 0}$  ainsi construite est-elle irréductible ? apériodique ?

**Correction :** La densité jointe a posteriori vérifie :

$$\pi((\theta, \mathbf{Z})|\mathbf{X}) \propto \theta^{a-1}(1-\theta)^{b-1} \mathbf{1}_{]0,1[}(\theta) \prod_{i=1}^N [\theta^{Z_i}(1-\theta)^{1-Z_i} \phi_{Z_i}(X_i)] .$$

Ainsi, le support de la loi jointe a posteriori  $\text{Loi}((\theta, \mathbf{Z})|\mathbf{X})$  est  $]0,1[ \times \{0,1\}^N$  et coïncide avec le produit des supports des marginales a posteriori  $\text{Loi}(\theta|\mathbf{X})$  et  $\text{Loi}(\mathbf{Z}|\mathbf{X})$ . Par un résultat du cours, ceci assure que la chaîne est apériodique et irréductible pour la mesure correspondant à  $\text{Loi}((\theta, \mathbf{Z})|\mathbf{X})$ , elle-même équivalente à  $\lambda_{]0,1[} \otimes \mu^{\otimes N}$ .

- (g) La convergence en variation totale de  $\text{Loi}(\theta_n|\mathbf{X})$  vers la loi a posteriori  $\text{Loi}(\theta|\mathbf{X})$  et le Théorème ergodique sont-ils vérifiés ?

**Correction :** À une fonction  $\phi : \Theta = [0,1] \rightarrow [0,1]$  on peut associer  $\Phi : \Theta \times \{0,1\}^N \rightarrow [0,1]$  par  $\Phi(\theta, \mathbf{z}) := \phi(\theta)$ . Ainsi

$$|\mathbb{E}[\phi(\theta_n)|\mathbf{X}] - \mathbb{E}[\phi(\theta)|\mathbf{X}]| = |\mathbb{E}[\Phi(\theta_n, \mathbf{Z}_n)|\mathbf{X}] - \mathbb{E}[\Phi(\theta, \mathbf{Z})|\mathbf{X}]| ,$$

or l'une des caractérisations de la distance en variation totale assure que

$$|\mathbb{E}[\Phi(\theta_n, \mathbf{Z}_n)|\mathbf{X}] - \mathbb{E}[\Phi(\theta, \mathbf{Z})|\mathbf{X}]| \leq \|\text{Loi}((\theta_n, \mathbf{Z}_n)|\mathbf{X}) - \text{Loi}((\theta, \mathbf{Z})|\mathbf{X})\| ,$$

donc en prenant le supremum sur les fonctions  $\phi$ , on en déduit que

$$\|\text{Loi}(\theta_n|\mathbf{X}) - \text{Loi}(\theta|\mathbf{X})\| \leq \|\text{Loi}((\theta_n, \mathbf{Z}_n)|\mathbf{X}) - \text{Loi}((\theta, \mathbf{Z})|\mathbf{X})\| .$$

Nous venons de voir que pour  $\lambda_{]0,1[} \otimes \mu^{\otimes N}$ -presque toute condition initiale  $(\theta_0, \mathbf{z}_0)$ , le majorant tend p.s. vers 0. Il s'ensuit que pour presque tout  $\theta_0$  de  $[0,1]$ , on a

$$\|\text{Loi}(\theta_n|\mathbf{X}) - \text{Loi}(\theta|\mathbf{X})\| \xrightarrow{\text{p.s.}} 0 .$$

On montre de même que le Théorème ergodique est vérifié.

- (h) La suite de variables aléatoires  $(\theta_n)$  est-elle une chaîne de Markov ?

**Correction :** La réponse est oui et c'est un résultat général pour l'échantillonneur de Gibbs. Dans notre cas particulier, si l'on se réfère à la construction donnée en réponse à la Question 2e, on voit que  $\theta_{n+1}$  est obtenu par simulation d'une loi Beta dont les paramètres dépendent de  $\mathbf{Z}_n$  et que le vecteur  $\mathbf{Z}_n$  est lui-même obtenu par simulations de variables de Bernoulli dont les paramètres dépendent de  $\theta_n$ . Par conséquent on peut trouver une suite d'éléments aléatoires  $(W_n)_{n \geq 1} = ((U_n, V_n))_{n \geq 1}$  i.i.d. et indépendants de  $(\theta_0, \mathbf{Z}_0)$  tels que, pour tout  $n$ , on puisse écrire

$$\theta_{n+1} = h(\mathbf{Z}_n, V_{n+1}) = h(g(\theta_n, U_{n+1}), V_{n+1}) = f(\theta_n, W_{n+1}) ,$$

ce qui prouve que  $(\theta_n)$  est une chaîne de Markov.

Attention : De façon générale, si  $((X_n, Y_n))_{n \geq 0}$  est une chaîne de Markov,  $(Y_n)_{n \geq 0}$  n'en est pas forcément une, cf. par exemple les modèles de Markov cachés.

#### EXERCICE 4 (Gibbs simultané = Gibbs erroné)

On considère sur  $\mathbb{R}^2$  le couple de variables aléatoires  $(X, Y)$  où  $X$  est gaussienne standard et  $Y = X + W$  avec  $W$  gaussienne standard indépendante de  $X$ .

1. Quelle est la densité de la loi jointe  $\pi$  du couple  $(X, Y)$  ?

**Correction :** Le vecteur  $[X, W]^T$  est gaussien puisque  $X$  et  $W$  sont deux variables gaussiennes indépendantes. Par conséquent le vecteur  $[X, Y]^T = M[X, W]^T$  est gaussien en tant que transformée linéaire du vecteur gaussien  $[X, W]^T$ . Il suffit donc de spécifier le vecteur moyenne et la matrice de covariance. Clairement  $\mathbb{E}[X] = \mathbb{E}[Y] = 0$  donc le vecteur moyenne est nul. Pour la matrice de covariance, on a par construction  $\text{Var}(X) = 1$ ,  $\text{Var}(Y) = \text{Var}(X) + \text{Var}(W) = 2$  et  $\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 1$ , d'où

$$\pi = \text{Loi}(X, Y) = \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}\right) =: \mathcal{N}(m, \Sigma).$$

Sa densité est donc

$$\pi(x, y) = \frac{1}{2\pi\sqrt{\det\Sigma}} \exp\left\{-\frac{1}{2}[x, y]\Sigma^{-1}\begin{bmatrix} x \\ y \end{bmatrix}\right\} = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(2x^2 - 2xy + y^2)\right\}.$$

2. Quelles sont les lois conditionnelles  $\text{Loi}(Y|X)$  et  $\text{Loi}(X|Y)$  ?

**Correction :** Puisque  $Y = X + W$  avec  $W$  indépendante de  $X$ , il est clair que  $\text{Loi}(Y|X) = \mathcal{N}(X, 1)$ . On peut le retrouver par le calcul :

$$\pi(y|x) \propto \pi(x, y) \propto \exp\left\{-\frac{1}{2}(y^2 - 2xy)\right\} \propto \exp\left\{-\frac{1}{2}(y - x)^2\right\}.$$

On procède de la même façon pour déterminer la loi conditionnelle  $\text{Loi}(X|Y)$  avec cette fois

$$\pi(x|y) \propto \pi(x, y) \propto \exp\left\{-(x^2 - xy)\right\} \propto \exp\left\{-\left(x - \frac{y}{2}\right)^2\right\},$$

ce qui prouve que  $\text{Loi}(X|Y) = \mathcal{N}\left(\frac{Y}{2}, \frac{1}{2}\right)$ . Autrement dit, sachant  $Y$ , on a la représentation en loi  $X = Y/2 + Z/\sqrt{2}$  où  $Z$  est gaussienne standard indépendante de  $Y$ .

3. Décrire une étape de l'échantillonneur de Gibbs séquentiel pour passer du couple  $(X_n, Y_n)$  au couple  $(X_{n+1}, Y_{n+1})$  grâce à des simulations de variables gaussiennes. Vérifier que la loi  $\pi$  est bien stationnaire.

**Correction :** Notons  $W_{n+1}$  et  $Z_{n+1}$  deux variables gaussiennes standards indépendantes entre elles et du couple  $(X_n, Y_n)$ . D'après la question précédente, une étape de l'échantillonneur de Gibbs séquentiel consiste tout simplement à poser  $X_{n+1} = Y_n/2 + Z_{n+1}/\sqrt{2}$  puis  $Y_{n+1} = X_{n+1} + W_{n+1}$ . Pour vérifier que  $\pi$  est bien stationnaire, il suffit de montrer que si  $(X_n, Y_n) \sim \pi$ , alors  $(X_{n+1}, Y_{n+1}) \sim \pi$ . Or, dire que  $(X_n, Y_n) \sim \pi$  signifie en particulier que  $[X_n, Y_n]^T \sim \mathcal{N}(m, \Sigma)$  est un vecteur gaussien. Ainsi

$$\begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1/2 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} X_n \\ Y_n \end{bmatrix} + \begin{bmatrix} 0 & 1/\sqrt{2} \\ 1 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} W_{n+1} \\ Z_{n+1} \end{bmatrix}$$

est la somme de deux vecteur gaussiens indépendants, ce qui en fait un vecteur gaussien. Il est clairement de moyenne nulle. Sa matrice de covariance peut se calculer via la représentation précédente : comme

$$\begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix} = A \begin{bmatrix} X_n \\ Y_n \end{bmatrix} + B \begin{bmatrix} W_{n+1} \\ Z_{n+1} \end{bmatrix},$$

on a, par indépendance,

$$\text{Cov} \left( \begin{bmatrix} X_{n+1} \\ Y_{n+1} \end{bmatrix} \right) = A \text{Cov} \left( \begin{bmatrix} X_n \\ Y_n \end{bmatrix} \right) A^T + B \text{Cov} \left( \begin{bmatrix} W_{n+1} \\ Z_{n+1} \end{bmatrix} \right) B^T = A \Sigma A^T + B I_2 B^T = \Sigma,$$

ce qui prouve que, comme vu en cours, la loi  $\pi$  est stationnaire pour l'échantillonneur de Gibbs séquentiel.

Remarque : Sans calcul matriciel, une autre façon de vérifier que la matrice de covariance est bien  $\Sigma$  est de dire que

$$\text{Var}(X_{n+1}) = \text{Var}(Y_n/2 + Z_{n+1}/\sqrt{2}) = \frac{1}{4} \text{Var}(Y_n) + \frac{1}{2} \text{Var}(Z_n) = 1,$$

puis

$$\text{Var}(Y_{n+1}) = \text{Var}(X_{n+1} + W_{n+1}) = \text{Var}(X_{n+1}) + \text{Var}(W_{n+1}) = 2,$$

et enfin

$$\text{Cov}(X_{n+1}, Y_{n+1}) = \text{Cov}(X_{n+1}, X_{n+1} + W_{n+1}) = \text{Var}(X_{n+1}) = 1.$$

4. On considère maintenant une chaîne  $(X_n, Y_n)$  dont la dynamique est la suivante : sachant  $(X_n, Y_n)$ , on simule  $X_{n+1}$  selon la loi conditionnelle  $\text{Loi}(X|Y_n) = \pi(\cdot|Y_n)$  et  $Y_{n+1}$  selon la loi conditionnelle  $\text{Loi}(Y|X_n) = \pi(\cdot|X_n)$ . Cette façon de se déplacer s'apparente donc à un échantillonneur de Gibbs où les deux coordonnées seraient remises à jour simultanément. Montrer que dans ce cas la loi  $\pi$  n'est pas stationnaire.

**Correction :** Avec les mêmes notations qu'en question précédente, on a cette fois  $X_{n+1} = Y_n/2 + Z_{n+1}/\sqrt{2}$  mais  $Y_{n+1} = X_n + W_{n+1}$ . Si  $(X_n, Y_n) \sim \pi = \mathcal{N}(m, \Sigma)$ , le vecteur  $[X_{n+1}, Y_{n+1}]^T$  est encore gaussien centré, les variances de  $X_{n+1}$  et  $Y_{n+1}$  sont les mêmes que celles de  $X_n$  et  $Y_n$ , à savoir respectivement 1 et 2, mais la covariance a changé :

$$\text{Cov}(X_{n+1}, Y_{n+1}) = \text{Cov}(Y_n/2 + Z_{n+1}/\sqrt{2}, X_n + W_{n+1}) = \frac{1}{2} \text{Cov}(X_n, Y_n) = \frac{1}{2} \neq 1 = \text{Cov}(X_n, Y_n),$$

ce qui prouve que  $\pi$  n'est pas stationnaire. Morale de l'histoire : il ne faut pas remettre les deux coordonnées à jour en même temps !

## EXERCICE 5 (Slice Sampling)

On considère la fonction

$$\pi(x, y) = \mathbf{1}_{x>0} \mathbf{1}_{0<y<\frac{1}{2}\exp(-\sqrt{x})}.$$

1. Montrer que la fonction  $\pi(x) = \frac{1}{2} \exp(-\sqrt{x}) \mathbf{1}_{x>0}$  est une densité sur  $\mathbb{R}$  et en déduire que  $\pi(x, y)$  est une densité sur  $\mathbb{R}^2$ . Quelle loi reconnaissez-vous pour le couple  $(X, Y)$  de densité  $\pi(x, y)$  ?

**Correction :** La fonction  $x \mapsto \pi(x)$  est positive et le changement de variable  $t = \sqrt{x}$  montre qu'elle somme à 1 :

$$\int_0^\infty \frac{1}{2} \exp(-\sqrt{x}) dx = \int_0^\infty t e^{-t} dt = \mathbb{E}[\text{Exp}(1)] = 1.$$

La fonction  $(x, y) \mapsto \pi(x, y)$  est positive et, par Fubini-Tonelli,

$$\iint_{\mathbb{R}^2} \pi(x, y) dx dy = \int_0^\infty \left( \int_0^{\frac{1}{2}\exp(-\sqrt{x})} dy \right) dx = \int_0^\infty \frac{1}{2} \exp(-\sqrt{x}) dx = 1.$$

Un couple  $(X, Y)$  de densité  $\pi$  est tout simplement un couple distribué uniformément dans l'hypographe de la densité  $x \mapsto \pi(x)$ . On retrouve le fait que, dans ce cas, la première coordonnée  $X$  a pour densité  $\pi(x)$ .

2. Soient  $x > 0$  et  $y > 0$ . Quelle est la loi de  $Y$  sachant  $X = x$ ? Quelle est la loi de  $X$  sachant  $Y = y$ ?

**Correction :** Pour tout  $x > 0$ , on a

$$\pi(y|x) \propto \pi(x, y) = \mathbf{1}_{0 < y < \frac{1}{2} \exp(-\sqrt{x})},$$

donc  $\text{Loi}(Y|X = x) = \text{Unif}([0, \frac{1}{2} \exp(-\sqrt{x})])$ . Réciproquement, pour tout  $y > 0$ , on a

$$\pi(x|y) \propto \pi(x, y) = \mathbf{1}_{x > 0} \mathbf{1}_{0 < y < \frac{1}{2} \exp(-\sqrt{x})} = \mathbf{1}_{0 < x < (\log(2y))^2},$$

donc  $\text{Loi}(X|Y = y) = \text{Unif}([0, (\log(2y))^2])$ .

3. En déduire un échantillonneur de Gibbs séquentiel pour simuler une chaîne de Markov  $((X_n, Y_n))_{n \geq 0}$  admettant  $\pi(x, y)$  pour loi stationnaire. On pourra également faire un dessin pour comprendre graphiquement comment fonctionne cet algorithme.

**Correction :** On part d'une condition initiale appartenant à l'hypographe de  $x \mapsto \pi(x)$ , c'est-à-dire un couple  $(x_0, y_0)$  tel que  $x_0 > 0$  et  $0 < y_0 < \pi(x_0) = \frac{1}{2} \exp(-\sqrt{x_0})$ . Sachant  $(X_n, Y_n)$ , une étape de l'échantillonneur de Gibbs séquentiel se décompose de la façon suivante :

- (a) Simuler  $X_{n+1} \sim \text{Unif}([0, (\log(2Y_n))^2])$  ;
- (b) Simuler  $Y_{n+1} \sim \text{Unif}([0, \frac{1}{2} \exp(-\sqrt{X_{n+1}})])$ .

On sait que la chaîne  $((X_n, Y_n))_{n \geq 0}$  ainsi construite admet  $\pi(x, y)$  comme loi stationnaire. Graphiquement, cet échantillonneur consiste à alterner des déplacements horizontaux et verticaux de façon uniforme dans les segments délimités par les frontières de l'hypographe de  $\pi(x)$  : c'est donc une marche aléatoire dans cet hypographe.

4. Montrer que  $(X_n)_{n \geq 0}$  est une chaîne de Markov admettant  $\pi(x)$  comme loi stationnaire.

**Correction :** Pour la même raison qu'en Exercice 3, question 2h, la suite  $(X_n)_{n \geq 0}$  est bien une chaîne de Markov en tant que suite définie par une récurrence aléatoire :

$$X_{n+1} = h(Y_n, V_{n+1}) = h(g(X_n, T_{n+1}), V_{n+1}) =: f(X_n, W_{n+1}).$$

Pour montrer qu'elle admet  $\pi(x)$  pour loi stationnaire, il suffit de prouver que si  $X_0 \sim \pi(x)$  alors  $X_1 \sim \pi(x)$ . Or si  $X_0 \sim \pi(x)$ , considérons la variable  $Y_0$  dont la loi conditionnellement à  $X_0$  est définie par  $\text{Loi}(Y_0|X_0) = \text{Unif}([0, \frac{1}{2} \exp(-\sqrt{X_0})])$ . Ainsi construit, le couple  $(X_0, Y_0)$  a les mêmes première marginale et seconde loi conditionnelle que celles de la densité jointe  $\pi(x, y)$  donc  $(X_0, Y_0) \sim \pi(x, y)$ . Du résultat précédent on déduit que  $(X_1, Y_1) \sim \pi(x, y)$ . Par la première question, il en découle que  $X_1 \sim \pi(x)$  et au final  $(X_n)_{n \geq 0}$  est bien une chaîne de Markov admettant  $\pi(x)$  comme loi stationnaire.

5. (a) Partant de  $X_0 = x \in \mathbb{R}_+^*$ , préciser  $\text{Loi}(X_1|X_0 = x)$ . En déduire que le noyau de transition de la chaîne  $(X_n)_{n \geq 0}$  est  $P(x, dx') = (\int \pi(y|x) \pi(x'|y) dy) dx'$ .

**Correction :** Pour obtenir  $X_1$  à partir de  $X_0$ , on simule d'abord  $Y_0$  de sorte que, sachant  $X_0$ ,  $Y_0$  a pour densité  $\pi(\cdot|X_0)$ , puis  $X_1$  de sorte que, sachant  $Y_0$  (et  $X_0$ ),  $X_1$  a pour densité

$\pi(\cdot|Y_0)$ . Ainsi, sachant  $X_0 = x$ , le couple  $(X_1, Y_0)$  a pour densité  $(x', y) \mapsto \pi(y|x)\pi(x'|y)$ . Par marginalisation,  $\text{Loi}(X_1|X_0 = x)$  a pour densité  $x \mapsto \int \pi(y|x)\pi(x'|y)dy$ . Ainsi la chaîne  $(X_n)_{n \geq 0}$  a pour noyau de transition  $P(x, dx') = (\int \pi(y|x)\pi(x'|y)dy)dx'$ .

- (b) Soit un borélien  $B$  de  $\mathbb{R}_+$  de mesure de Lebesgue non-nulle. Montrer que pour tout  $x \in \mathbb{R}_+^*$ ,  $P(x, B) > 0$ . En déduire que la chaîne  $(X_n)_{n \geq 0}$  est Lebesgue-irréductible et apériodique.

**Correction :** Puisque  $B$  est de mesure de Lebesgue non-nulle, il existe un segment  $[a, b] \subset \mathbb{R}_+$  tel qu'en notant  $B' := [a, b] \cap B$ ,  $\lambda(B') > 0$ . Par suite, en prenant  $c = \min(\exp(-\sqrt{a})/2, \exp(-\sqrt{b})/2)$ , le borélien  $B' \times [0, c]$  est inclus dans l'hypographe de  $x \mapsto \exp(-\sqrt{x})/2$  (i.e. dans le support de  $\pi(x, y)$ ). Il vient ensuite, en appliquant Fubini-Tonelli :

$$\begin{aligned} P(x, B) &= \int \mathbf{1}_{x' \in B} P(x, dx') = \iint \mathbf{1}_{x' \in B} \pi(y|x)\pi(x'|y)dydx' \\ &\geq \int_0^c \pi(y|x) \left( \int_{B'} \pi(x'|y)dx' \right) dy \\ &= \frac{1}{\frac{1}{2} \exp(-\sqrt{x})} \int_0^c \frac{1}{(\log(2y))^2} \left( \int_{B'} dx' \right) dy \\ &= \frac{\lambda(B')}{\frac{1}{2} \exp(-\sqrt{x})} \int_0^c \frac{1}{(\log(2y))^2} dy > 0. \end{aligned}$$

Un résultat du cours assure enfin que la chaîne  $(X_n)_{n \geq 0}$  est Lebesgue-irréductible et apériodique.

6. Montrer que si  $X$  a pour densité  $\pi(x)$ , alors la variable  $T := \sqrt{X}$  suit une loi classique et en déduire une nouvelle méthode pour simuler  $X$ . Entre cette dernière et l'échantillonneur de Gibbs, laquelle choisissez-vous ?

**Correction :** Il suffit d'appliquer la méthode de la fonction muette : pour toute fonction test  $\varphi$ , on a en effet

$$\mathbb{E}[\varphi(T)] = \mathbb{E}[\varphi(\sqrt{X})] = \int_{\mathbb{R}} \varphi(\sqrt{x}) \frac{1}{2} \exp(-\sqrt{x}) \mathbf{1}_{x>0} dx = \int_{\mathbb{R}} \varphi(t) t e^{-t} \mathbf{1}_{t>0} dt,$$

ce qui prouve que  $T \sim \text{Gamma}(2, 1)$ . Puisque tout logiciel digne de ce nom sait simuler selon une loi Gamma, il suffit pour simuler  $X$  de densité  $\pi(x)$  de commencer par simuler  $T$  suivant cette loi Gamma(2, 1) puis de prendre sa racine carrée. Cette méthode est bien sûr préférable à l'échantillonneur de Gibbs puisqu'elle permet en particulier de simuler autant de variables i.i.d. que l'on veut exactement selon la loi  $\pi$ , tandis que l'échantillonneur de Gibbs construit une suite de variables corrélées dont la loi tend vers  $\pi$ . Néanmoins, ceci est dû au fait que nous sommes sur un exemple très particulier.

7. Généralisation. La méthode par échantillonnage de Gibbs introduite ci-dessus est connue sous le nom de *Slice Sampling*. Soit  $\pi(x)$  une densité sur  $\mathbb{R}$ , expliquer à quelles conditions et de quelle façon cette méthode peut s'appliquer pour construire une chaîne de Markov de loi stationnaire  $\pi$  ?

**Correction :** Pour tout couple  $(x_0, y_0) \in \mathbb{R} \times \mathbb{R}_+^*$ , notons

$$\mathcal{A}(x_0) := \{y : 0 < y < \pi(x_0)\} = ]0, \pi(x_0)[$$

et

$$\mathcal{B}(y_0) := \{x : \pi(x) \geq y_0\}.$$

Le Slice Sampling consiste à effectuer des tirages uniformes dans des ensembles  $\mathcal{A}(x)$  et  $\mathcal{B}(y)$ . Pour tout réel  $x$ , le tirage dans  $\mathcal{A}(x)$  est immédiat dès lors que l'on sait calculer  $\pi(x)$  : il nous faut donc juste une formule pour cette densité. L'autre étape est plus exigeante puisque pour tout  $y > 0$ , il faut pouvoir simuler uniformément dans  $\mathcal{B}(y)$ , donc il faut non seulement avoir accès à cet ensemble, mais également que le tirage uniforme dans celui-ci soit possible. Si ces conditions sont remplies, alors tout ce qui a été fait précédemment dans un cas particulier se généralise de façon immédiate.

### EXERCICE 6 ((†) Modèle d'Ising, échantillonneur de Gibbs et dynamique de Glauber)

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  $\mathcal{C}$ . À une 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.

1. Soit  $(i_0, j_0)$  un site fixé dans  $\mathcal{C}$ . Préciser la loi conditionnelle  $\mu_T(S_{i_0 j_0} | S_{-(i_0 j_0)})$ .

**Correction :** Si  $X = 2B - 1$  avec  $B \sim \text{Ber}(p)$ , on dira que  $X$  suit une loi de Rademacher généralisée de paramètre  $p$ , noté  $X \sim \text{Rad}(p)$ . C'est bien le cas ici puisque le spin  $S_{i_0 j_0}$  ne peut prendre que les valeurs  $\pm 1$ . Pour déterminer le paramètre, on fait comme d'habitude, en éliminant tous les termes parasites grâce au symbole de proportionnalité :

$$\mu_T(S_{i_0 j_0} | S_{-(i_0 j_0)}) \propto \mu_T(S_{i_0 j_0}, S_{-(i_0 j_0)}) \propto \exp\left(\frac{S_{i_0 j_0}}{T} \sum_{(i,j) \sim (i_0, j_0)} S_{ij}\right).$$

Ainsi

$$\mathbb{P}(S_{i_0 j_0} = 1 | S_{-(i_0 j_0)}) \propto \exp\left(\frac{1}{T} \sum_{(i,j) \sim (i_0, j_0)} S_{ij}\right),$$

donc

$$\mathbb{P}(S_{i_0 j_0} = 1 | S_{-(i_0 j_0)}) = \frac{\exp\left(\frac{1}{T} \sum_{(i,j) \sim (i_0, j_0)} S_{ij}\right)}{\exp\left(\frac{1}{T} \sum_{(i,j) \sim (i_0, j_0)} S_{ij}\right) + \exp\left(-\frac{1}{T} \sum_{(i,j) \sim (i_0, j_0)} S_{ij}\right)},$$

c'est-à-dire

$$\text{Loi}(S_{i_0 j_0} | S_{-(i_0 j_0)}) = \text{Rad}\left(\frac{1}{1 + \exp\left(-\frac{2}{T} \sum_{(i,j) \sim (i_0, j_0)} S_{ij}\right)}\right).$$

2. En déduire un échantillonneur de Gibbs aléatoire (donc non séquentiel) pour construire une chaîne de Markov  $(S_n)_{n \geq 0}$  de loi stationnaire  $\mu_T$ .

**Correction :** On part d'une configuration de spins  $S_0$  quelconque, par exemple chaque spin est tiré indépendamment à pile ou face. Si à l'instant  $n$  la configuration est  $S_n = (S_{ij})_{1 \leq i, j \leq N}$  :

- (a) On tire un site  $(i_0, j_0)$  au hasard uniforme.  
(b) On simule

$$S'_{i_0 j_0} \sim \text{Rad} \left( \frac{1}{1 + \exp \left( -\frac{2}{T} \sum_{(i,j) \sim (i_0, j_0)} S_{ij} \right)} \right).$$

- (c) On pose  $S_{n+1} = (S'_{i_0 j_0}, S_{-(i_0 j_0)})$ .

Cet échantillonneur de Gibbs produit une chaîne de Markov  $(S_n)_{n \geq 0}$  pour laquelle la mesure de Gibbs  $\mu_T$  est réversible, donc stationnaire. Cette façon de faire évoluer la configuration de spins est également appelée dynamique de Glauber.

3. On se place dans le cadre d'un espace d'états quelconque. Dans l'algorithme de Metropolis, au lieu de  $\alpha(x, y) = \min(1, \rho(x, y))$ , on considère plus généralement une fonction  $h : \mathbb{R}_+ \rightarrow [0, 1]$  vérifiant  $h(u) = uh(1/u)$  et la fonction d'acceptation  $\alpha(x, y) = h(\rho(x, y))$ . Par exemple  $h(u) = u/(1 + u)$  vérifie cette équation fonctionnelle.

- (a) Montrer que la réversibilité de la chaîne vis-à-vis de la loi cible est encore satisfaite.

**Correction :** Avec les notations du cours, il suffit de vérifier que, pour tout  $x \neq y$ ,

$$\pi(x)\alpha(x, y)q(x, y) = \pi(y)\alpha(y, x)q(y, x).$$

Or, puisque  $h(u) = uh(1/u)$ , on a

$$\pi(x)q(x, y)\alpha(x, y) = \pi(x)q(x, y)h(\rho(x, y)) = \pi(y)q(y, x)h(1/\rho(x, y)) = \pi(y)q(y, x)\alpha(y, x),$$

et le résultat est établi.

- (b) On revient au modèle d'Ising. Montrer que l'échantillonneur de Gibbs ci-dessus correspond à un algorithme de Metropolis où, 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 un seul un spin a changé de signe. On calcule le rapport d'acceptation  $\rho(S, S^{(i_0, j_0)})$ , puis on décide en fonction de

$$\alpha(S, S^{(i_0, j_0)}) = h(\rho(S, S^{(i_0, j_0)})),$$

avec  $h(u) = u/(1 + u)$ .

**Correction :** Ce Metropolis est symétrique et le rapport d'acceptation vaut

$$\rho(S, S^{(i_0, j_0)}) = \frac{\mu_T(S^{(i_0, j_0)})}{\mu_T(S)} = \exp \left( -\frac{1}{T} \left( V(S^{(i_0, j_0)}) - V(S) \right) \right) = \exp \left( -\frac{2S_{i_0 j_0}}{T} \sum_{(i, j) \sim (i_0, j_0)} S_{ij} \right)$$

donc

$$\alpha(S, S^{(i_0, j_0)}) = h(\rho(S, S^{(i_0, j_0)})) = \frac{\rho(S, S^{(i_0, j_0)})}{1 + \rho(S, S^{(i_0, j_0)})},$$

donne

$$\alpha(S, S^{(i_0, j_0)}) = \frac{1}{1 + \exp \left( \frac{2S_{i_0 j_0}}{T} \sum_{(i, j) \sim (i_0, j_0)} S_{ij} \right)}.$$

Par conséquent, si le spin valait  $S_{i_0 j_0} = -1$ , alors la probabilité qu'il devienne 1 vaut

$$\mathbb{P}(\text{Unif}([0, 1]) \leq \alpha(S, S^{(i_0, j_0)})) = \frac{1}{1 + \exp\left(-\frac{2}{T} \sum_{(i, j) \sim (i_0, j_0)} S_{ij}\right)},$$

qui est exactement le paramètre de la Rademacher susmentionnée. Si à l'inverse le spin valait  $S_{i_0 j_0} = 1$ , alors la probabilité qu'il reste à 1 est celle que la proposition soit refusée, c'est-à-dire

$$\mathbb{P}(\text{Unif}([0, 1]) > \alpha(S, S^{(i_0, j_0)})) = 1 - \frac{1}{1 + \exp\left(\frac{2}{T} \sum_{(i, j) \sim (i_0, j_0)} S_{ij}\right)},$$

c'est-à-dire à nouveau

$$\frac{1}{1 + \exp\left(-\frac{2}{T} \sum_{(i, j) \sim (i_0, j_0)} S_{ij}\right)}.$$

Ceci prouve que les algorithmes sont les mêmes.

---