Computational statistics
TP 3 : Hasting-Metropolis (and Gibbs) samplers

Thibault de SURREL

Exercice 1 : Hasting-Metropolis within Gibbs – Stochastic Approximation EM

1.

Les paramètres de notre modèle sont $\theta = (\overline{t_0}, \overline{v_0}, \sigma_\xi, \sigma_\tau, \sigma)$. Pour tout $i \in [\![1, N]\!]$, on peut définir la variable aléatoire $z_i = (\alpha_i, \tau_i)$ dont les paramètres associés sont $\theta_i = (\sigma_\xi, \sigma_\tau, \sigma)$ et on peut définir $z_{pop} = (t_0, v_0)$ dont les paramètres associés sont $\theta_{pop} = (\overline{t_0}, \overline{v_0})$. On cherche alors la log-vraisemblance de notre modèle $\log q(y, z, \theta)$. En utilisant deux fois la règle de Bayes, on peut déjà commencer par réécrire la log vraisemblance de la façon suivante :

$$\log q(y, z, \theta) = \log q(y \mid z, \theta) + \log q(z \mid \theta) + \log q(\theta).$$

On va alors calculer ces trois termes séparément :

— Pour $\log q(y \mid z, \theta)$, on commence par utiliser le fait que les $y_{i,j}$ sont tous indépendants :

$$\log q(y \mid z, \theta) = \sum_{i=1}^{N} \sum_{j=1}^{k_i} \log q(y_{i,j} \mid z, \theta)$$

Ensuite, sachant z et θ, on a que $y_{i,j} \sim \mathcal{N}(d_i(t_{i,j}), \sigma^2)$ donc

$$\log q(y \mid z, \theta) = \sum_{i=1}^{N} \sum_{j=1}^{k_i} \left[-\frac{1}{2} \left(\frac{y_{i,j} - d_i(t_{i,j})}{\sigma} \right)^2 - \log \sigma - \frac{1}{2} \log 2\pi \right]$$

— Pour $\log q(z \mid \theta)$, on commence par séparer par indépendance les $(z_i)_i$ et z_{pop} :

$$\log q(z \mid \theta) = \log q(z_{pop} \mid \theta_{pop}) + \log q(z_i \mid \theta_i)$$

Ensuite, on a d'une part que $t_0 \sim \mathcal{N}(\overline{t_0}, \sigma_{t_0}^2)$ et $v_0 \sim \mathcal{N}(\overline{v_0}, \sigma_{v_0}^2)$ et d'autre part que $\alpha_i \sim \log \mathcal{N}(0, \sigma_\xi^2)$ et $\tau_i \sim \mathcal{N}(0, \sigma_\tau^2)$ ainsi,

$$\begin{aligned} \log q(z \mid \theta) = &-\frac{1}{2} \left[\left(\frac{t_0 - \overline{t_0}}{\sigma_{t_0}} \right)^2 + \left(\frac{v_0 - \overline{v_0}}{\sigma_{v_0}} \right)^2 \right] - \log \sigma_{t_0} \sigma_{v_0} - \log 2\pi \\ &+ \sum_{i=1}^{N} \left\{ -\frac{1}{2} \left[\left(\frac{\tau_i}{\sigma_\tau} \right)^2 + \left(\frac{\xi_i}{\sigma_\xi} \right)^2 \right] - \log \sigma_\tau \sigma_\xi - \log 2\pi \right\} \end{aligned} \tag{1}$$

— Enfin, pour $\log q(\theta)$ on sait que $\overline{t_0} \sim \mathcal{N}(\overline{\overline{t_0}}, s_{t_0}^2)$, $\overline{v_0} \sim \mathcal{N}(\overline{\overline{v_0}}, s_{v_0}^2)$ et que $\sigma_\xi^2 \sim \mathcal{W}^{-1}(v_\xi, m_\xi)$, $\sigma_\tau^2 \sim \mathcal{W}^{-1}(v_\tau, m_\tau)$ et $\sigma^2 \sim \mathcal{W}^{-1}(v, m)$ donc on a

$$\begin{aligned} \log q(\theta) = &-\frac{1}{2} \left(\frac{\overline{t_0} - \overline{\overline{t_0}}}{s_{t_0}} \right)^2 - \frac{1}{2} \left(\frac{\overline{v_0} - \overline{\overline{v_0}}}{s_{v_0}} \right)^2 - \log s_{t_0} s_{v_0} + \left(-(m_\xi + 2) \log \sigma_\xi - \frac{v_\xi^2}{2\sigma_\xi^2} \right) \\ &+ \left(-(m_\tau + 2) \log \sigma_\tau - \frac{v_\tau^2}{2\sigma_\tau^2} \right) + \left(-(m + 2) \log \sigma - \frac{v^2}{2\sigma^2} \right) \end{aligned} \tag{2}$$

Ainsi, pour conclure, on a que

$$\begin{aligned}\log q(y,z,\theta) = \sum_{i=1}^{N}\sum_{j=1}^{k_i}\left[-\frac{1}{2}\left(\frac{y_{i,j}-d_i(t_{i,j})}{\sigma}\right)^2 - \log\sigma - \frac{1}{2}\log 2\pi\right] - \frac{1}{2}\left[\left(\frac{t_0-\overline{t_0}}{\sigma_{t_0}}\right)^2 + \left(\frac{v_0-\overline{v_0}}{\sigma_{v_0}}\right)^2\right] - \log\sigma_{t_0}\sigma_{v_0} \\ - \log 2\pi + \sum_{i=1}^{N} -\frac{1}{2}\left[\left(\frac{\tau_i}{\sigma_t}\right)^2 + \left(\frac{\xi_i}{\sigma_\xi}\right)^2\right] - \log\sigma_\tau\sigma_\xi - \log 2\pi - \frac{1}{2}\left(\frac{\overline{t_0}-\overline{\overline{t_0}}}{s_{t_0}}\right)^2 - \frac{1}{2}\left(\frac{\overline{v_0}-\overline{\overline{v_0}}}{s_{v_0}}\right)^2 \\ - \log s_{t_0}s_{v_0} + \left(-(m_\xi+2)\log\sigma_\xi - \frac{v_\xi^2}{2{\sigma_\xi}^2}\right) + \left(-(m_\tau+2)\log\sigma_\tau - \frac{v_\tau^2}{2{\sigma_\tau}^2}\right) + \left(-(m+2)\log\sigma - \frac{v^2}{2\sigma^2}\right).\end{aligned} \tag{3}$$

où j'ai mis en rouge les termes correspondant à la variable (y,z) et en bleu les termes correspondant aux paramètres θ. En effet, on veut maintenant montrer que le modèle appartient à une famille exponentielle courbe. Pour cela, on doit montrer qu'il existe $(y,z)\mapsto\phi_1(y,z), \theta\mapsto\phi_2(\theta), (y,z)\mapsto S(y,z)$ et $\theta\mapsto\psi(\theta)$ des fonctions telles que

$$\log q(y,z,\theta) = \phi_1(y,z) + \phi_2(\theta) + \langle S(y,z),\psi(\theta)\rangle + C.$$

où C est une constante indépendante de y,z et de θ.

Pour cela, on regarde chaque terme de $\log q$ comme ci-dessus :

— Pour $\log q(y\mid z,\theta)$, on voit que poser

$$S_1(y,z) = \frac{1}{NK}\sum_{i=1}^{N}\sum_{j=1}^{k_i}(y_{i,j}-d_i(t_{i,j}))^2 \text{ et } \psi_1(\theta) = -\frac{NK}{2\sigma^2}$$

convient où $K=\sum_{i=1}^{N}k_i$.

— Pour $\log q(z\mid\theta)$, on voit que l'on peut réécrire les différences au carré comme ci-dessous :

$$\left(\frac{t_0-\overline{t_0}}{\sigma_{t_0}}\right)^2 = \frac{{t_0}^2 - 2t_0\overline{t_0} + \overline{t_0}^2}{\sigma_{t_0}^2}$$

Ainsi, on pose

$$S_2(y,z) = t_0 \text{ et } \psi_2(\theta) = \frac{\overline{t_0}}{\sigma_{t_0}^2}$$

De même, on pose

$$S_3(y,z) = v_0 \text{ et } \psi_3(\theta) = \frac{\overline{v_0}}{\sigma_{v_0}^2}$$

Enfin, on pose aussi

$$S_4(y,z) = \frac{1}{N}\sum_{i=1}^{N}\tau_i^2 \text{ et } \psi_4(\theta) = -\frac{N}{2\sigma_\tau^2}$$

ainsi que

$$S_5(y,z) = \frac{1}{N}\sum_{i=1}^{N}\xi_i^2 \text{ et } \psi_5(\theta) = -\frac{N}{2\sigma_\xi^2}$$

Finalement, on pose :

$$S(y,z) = \begin{pmatrix}S_1(y,z)\\S_2(y,z)\\S_3(y,z)\\S_4(y,z)\\S_5(y,z)\end{pmatrix} \text{ et } \psi(\theta) = \begin{pmatrix}\psi_1(\theta)\\\psi_2(\theta)\\\psi_3(\theta)\\\psi_4(\theta)\\\psi_5(\theta)\end{pmatrix}$$

ainsi que

$$\begin{cases} \phi_1(y,z) = -\dfrac{t_0^2}{2\sigma_{t_0}^2} - \dfrac{v_0^2}{2\sigma_{v_0}^2} \\ \phi_2(\theta) = -NK\log\sigma - N\log\sigma_\tau\sigma_\xi - \dfrac{\overline{t_0}^2}{2\sigma_{t_0}^2} - \dfrac{\overline{v_0}^2}{2\sigma_{v_0}^2} + \log q(\theta) \end{cases} \tag{4}$$

on a le résultat. De plus, cette famille exponentielle est bien courbe car la dimension du paramètre θ est 5 ce qui est plus petit que la dimension de (y,z) qui vaut $2N+2$.

3.

On cherche ici à générer des points selon la loi $\pi(z) = q(z \mid y, \theta)$ où $z = (z_{pop}, z_i)_{i\in[\![1,N]\!]} = (t_0, v_0, \xi_i, \tau_i)_{i\in[\![1,N]\!]}$. Commençons par noter la relation suivante, qui nous servira pour la suite, et qui utilise la relation de Bayes :

$$\pi(z) = q(z \mid y, \theta) = \frac{q(y \mid z, \theta)q(z \mid \theta)}{q(y \mid \theta)} \tag{5}$$

On choisit comme loi de proposition une gaussienne multivariée $\mathcal{N}(z, \sigma_{prop})$. Alors, cette loi étant symétrique, le ratio d'acceptation se réécrit $\alpha(z^{(k)}, z^*) = \min(1, \frac{\pi(z^*)}{\pi(z^{(k)})})$. Grâce à la relation 5, on peut réécrire le quotient de la façon suivante :

$$\frac{\pi(z^*)}{\pi(z^{(k)})} = \frac{q(z^* \mid y,\theta)}{q(z^{(k)} \mid y,\theta)} = \frac{q(y \mid z^*,\theta)q(z^* \mid \theta)}{q(y \mid z^{(k)},\theta)q(z^{(k)} \mid \theta)}.$$

Afin de simplifier les calcul, on va s'intéresser au log de ce taux d'acception, la condition dans l'algorithme de Metropolis-Hasting se réécrivant alors : Si $U \sim \mathcal{U}([0,1])$ alors on accepte z^* si $\log U \leq \log\alpha(z^{(k)}, z^*)$ et on garde $z^{(k)}$ sinon. Alors, on a

$$\log\alpha(z^{(k)}, z^*) = \min\left\{0, \log q(y \mid z^*,\theta) + \log q(z^* \mid \theta) - \log q(y \mid z^{(k)},\theta) - \log q(z^{(k)} \mid \theta)\right\}.$$

De plus, on connaît déjà la quantité $\log q(y \mid z, \theta)$ d'après la question 1 :

$$\log q(y \mid z,\theta) = \sum_{i=1}^{N}\sum_{j=1}^{k_i}\left[-\frac{1}{2}\left(\frac{y_{i,j} - d_i(t_{i,j})}{\sigma}\right)^2 - \log\sigma - \frac{1}{2}\log 2\pi\right]$$

Dans cette formule, seule $d_{i,j}$ dépend de z, donc c'est le seul terme qui changera entre $\log q(y \mid z^*, \theta)$ et $\log q(y \mid z^{(k)}, \theta)$.

Pour le terme $\log q(z \mid \theta)$, on commence par utiliser l'indépendance de nos variables aléatoires selon i pour écrire :

$$\log q(z \mid \theta) = \log q(t_0 \mid \theta) + \log(v_0 \mid \theta) + \sum_{i=1}^{N}\log q(\xi_i \mid \theta) + \log q(\tau_i \mid \theta)$$

Toutes ces quantités sont connues d'après la question 1. On remarque qu'une partie des termes intervenant dans le calcul de $\log q(z \mid \theta)$ ne dépendent pas de z, (par exemple $\log\sigma_{t_0}\sigma_{v_0}$) et donc, d'après la formule de $\log\alpha$, seront éliminés lors de l'opération $\log q(z^* \mid \theta) - \log q(z^{(k)} \mid \theta)$.

On a alors tous les éléments pour implémenter l'algorithme de Symmetric Random Walk Hasting-Metropolis. La suite est sur le notebook.

4.

On cherche à calculer $\theta^{(k)} \in \text{argmax}_{\theta\in\Theta}\{\phi_2(\theta) + \langle S_k, \psi(\theta)\rangle\}$ avec les notations de la question 1. Posons alors g la fonction telle que

$$\forall\theta\in\Theta,\ g(\theta) = \phi_2(\theta) + \langle S_k, \psi(\theta)\rangle = \phi_2(\theta) + \sum_{i=1}^{5} S_k^{(i)}\psi_i(\theta).$$

où les $S^{(i)}$ sont les S_i de la question 1.

— **Pour $\overline{t_0}$** : On a dans un premier temps,

$$\frac{\partial \phi_2}{\partial \overline{t_0}}(\theta) = -\frac{\overline{t_0}}{\sigma_{t_0}^2} - \frac{\overline{t_0} - \overline{\overline{t_0}}}{s_{t_0}^2}$$

De plus, d'après les expressions des ψ_i, $\overline{t_0}$ n'intervient que dans ψ_2 donc :

$$\frac{\partial g}{\partial \overline{t_0}}(\theta) = -\frac{\overline{t_0}}{\sigma_{t_0}^2} - \frac{\overline{t_0} - \overline{\overline{t_0}}}{s_{t_0}^2} + S_k^{(2)} \frac{1}{\sigma_{t_0}^2}$$

Donc,

$$\frac{\partial g}{\partial \overline{t_0}}(\theta) = 0 \iff \overline{t_0}^{(k)} = \left(\frac{S_k^{(2)}}{\sigma_{t_0}^2} + \frac{\overline{\overline{t_0}}}{s_{t_0}^2} \right) \left(\frac{1}{\sigma_{t_0}^2} + \frac{1}{s_{t_0}^2} \right)^{-1}$$

Le calcul est le même pour $\overline{v_0}^{(k)}$.

— **Pour σ_ξ** : On a

$$\frac{\partial \phi_2}{\partial \sigma_\xi}(\theta) = -\frac{N}{\sigma_\xi} - \frac{(m_\xi + 2)}{\sigma_\xi} + \frac{v_\xi^2}{\sigma_\xi^3}$$

et σ_ξ n'intervient que dans l'expression de ψ_5, ainsi :

$$\frac{\partial g}{\partial \sigma_\xi}(\theta) = -\frac{N}{\sigma_\xi} - \frac{(m_\xi + 2)}{\sigma_\xi} + \frac{v_\xi^2}{\sigma_\xi^3} + S_k^{(5)} \frac{N}{\sigma_\xi^3}$$

Ce qui permet de dire que

$$\frac{\partial g}{\partial \sigma_\xi}(\theta) = 0 \iff \sigma_\xi^{(k)} = \sqrt{\frac{S_k^{(5)} N + v_\xi^2}{N + m_\xi + 2}}$$

Les calculs sont les mêmes (à un facteur K près pour σ) pour trouver $\sigma^{(k)}$ et $\sigma_\tau^{(k)}$
Ainsi, on peut résumer ces calculs en donnant l'expression complète de $\theta^{(k)}$:

$$\theta^{(k)} = \begin{pmatrix} \overline{t_0}^{(k)} \\ \overline{v_0}^{(k)} \\ \sigma_\xi^{(k)} \\ \sigma_\tau^{(k)} \\ \sigma^{(k)} \end{pmatrix} = \begin{pmatrix} \left(\frac{S_k^{(2)}}{\sigma_{t_0}^2} + \frac{\overline{\overline{t_0}}}{s_{t_0}^2} \right) \left(\frac{1}{\sigma_{t_0}^2} + \frac{1}{s_{t_0}^2} \right)^{-1} \\ \left(\frac{S_k^{(3)}}{\sigma_{v_0}^2} + \frac{\overline{\overline{v_0}}}{s_{v_0}^2} \right) \left(\frac{1}{\sigma_{v_0}^2} + \frac{1}{s_{v_0}^2} \right)^{-1} \\ \sqrt{\frac{S_k^{(5)} N + v_\xi^2}{N + m_\xi + 2}} \\ \sqrt{\frac{S_k^{(4)} N + v_\tau^2}{N + m_\tau + 2}} \\ \sqrt{\frac{S_k^{(1)} KN + v^2}{KN + m + 2}} \end{pmatrix}$$

5./6./7.

Sur le notebook.

8.

L'avantage d'échantillonner des blocs de variables au lieu de les échantillonner une par une est un gain de temps en diminuant le nombre de tirages à faire. Cette méthode est utile lorsque les éléments à échantilloner comportent des variables qui sont corrélées entre elles, comme ici c'est le cas pour les z_{pop} et les z_i qui forment $N + 1$ blocs de variables corrélées.

9.

Sur le notebook.

Exercice 2 : Multiplicative Hasting-Metropolis

1.

On cherche à déterminer la distribution de $Y \sim q(X, Y)$. Pour cela, on se munit d'une fonction h continue bornée. On a alors,

$$\begin{aligned}\mathbb{E}[h(Y)] &= \mathbb{E}[h(\varepsilon X) \mid \mathcal{B} = 0]\mathbb{P}(\mathcal{B} = 0) + \mathbb{E}\left[h\left(\frac{X}{\varepsilon}\right) \mid \mathcal{B} = 1\right]\mathbb{P}(\mathcal{B} = 1) \\ &= \frac{1}{2}\int_{-1}^{1} h(\varepsilon X)f(\varepsilon)\mathrm{d}\varepsilon + \frac{1}{2}\int_{-1}^{1} h\left(\frac{X}{\varepsilon}\right) f(\varepsilon)\mathrm{d}\varepsilon\end{aligned} \tag{6}$$

Car $\mathcal{B}$ suit une loi de Bernoulli de paramètre $\frac{1}{2}$.

— Cas $X > 0$, alors, on effectue le changement de variable affine $y = \varepsilon X$ dans la première intégrale de 6 qui nous donne :

$$\int_{-1}^{1} h(\varepsilon X)f(\varepsilon)\mathrm{d}\varepsilon = \int_{-X}^{X} h(y)\frac{f(\frac{y}{X})}{X}\mathrm{d}y$$

De plus, on effectue le changement de variable $y = \frac{X}{\varepsilon}$ dans la deuxième intégrale qui nous donne :

$$\begin{aligned}\int_{-1}^{1} h\left(\frac{X}{\varepsilon}\right) f(\varepsilon)\mathrm{d}\varepsilon &= \int_{-1}^{0} h\left(\frac{X}{\varepsilon}\right) f(\varepsilon)\mathrm{d}\varepsilon + \int_{0}^{1} h\left(\frac{X}{\varepsilon}\right) f(\varepsilon)\mathrm{d}\varepsilon \\ &= \int_{-X}^{-\infty} h(y)f\left(\frac{X}{y}\right)\left(-\frac{X}{y^2}\right)\mathrm{d}y + \int_{+\infty}^{X} h(y)f\left(\frac{X}{y}\right)\left(-\frac{X}{y^2}\right)\mathrm{d}y \\ &= \int_{-\infty}^{-X} h(y)f\left(\frac{X}{y}\right)\frac{X}{y^2}\mathrm{d}y + \int_{X}^{+\infty} h(y)f\left(\frac{X}{y}\right)\frac{X}{y^2}\mathrm{d}y\end{aligned} \tag{7}$$

On peut effectuer le même calcul dans le cas $X < 0$ ce qui permet de dire, dans le cas où $X \neq 0$, on a :

$$\begin{aligned}q(X, Y) &= \frac{1}{2}\frac{f(\frac{Y}{X})}{|X|}\mathbb{1}_{]-|X|,|X|[}(Y) + \frac{1}{2}f\left(\frac{X}{Y}\right)\frac{|X|}{Y^2}\mathbb{1}_{]-\infty,-|X|[} + \frac{1}{2}f\left(\frac{X}{Y}\right)\frac{|X|}{Y^2}\mathbb{1}_{]|X|,+\infty[} \\ &= \frac{1}{2}\frac{f(\frac{Y}{X})}{|X|}\mathbb{1}_{|Y|\leq|X|} + \frac{1}{2}f\left(\frac{X}{Y}\right)\frac{|X|}{Y^2}\mathbb{1}_{|Y|>|X|}\end{aligned} \tag{8}$$

— Dans le cas où $X = 0$, alors $Y = 0$ donc

$$q(0, Y) = \delta_0$$

On considéra dans la suite que X ne prend jamais la valeur 0.

2.

Le taux d'acceptation α est par définition :

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

Dans notre cas, et compte tenu de la densité q calculée à la question 1, nous avons :

$$\begin{aligned}\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} &= \frac{\frac{1}{2}\frac{f(\frac{x}{y})}{|y|}\mathbb{1}_{|x|\leq|y|} + \frac{1}{2}f(\frac{y}{x})\frac{|y|}{x^2}\mathbb{1}_{|x|>|y|}}{\frac{1}{2}\frac{f(\frac{y}{x})}{|x|}\mathbb{1}_{|y|\leq|x|} + \frac{1}{2}f(\frac{x}{y})\frac{|x|}{y^2}\mathbb{1}_{|y|>|x|}} \\ &= \frac{|y|}{|x|}\frac{\frac{1}{2}\frac{f(\frac{y}{x})}{|x|}\mathbb{1}_{|y|\leq|x|} + \frac{1}{2}f(\frac{x}{y})\frac{|x|}{y^2}\mathbb{1}_{|y|>|x|}}{\frac{1}{2}\frac{f(\frac{y}{x})}{|x|}\mathbb{1}_{|y|\leq|x|} + \frac{1}{2}f(\frac{x}{y})\frac{|x|}{y^2}\mathbb{1}_{|y|>|x|}} \\ &= \frac{|y|}{|x|}\end{aligned} \tag{9}$$

Ainsi, le taux d'acceptation est :

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

3.

Pour la première distribution que l'on souhaite échantillonner, on choisit la loi exponentielle $\mathcal{E}(1)$ car elle est échantillonnable par la méthode de la transformée inverse. On choisit f comme étant la distribution uniforme sur $]-1,1[$: $\mathcal{U}(]-1,1[)$.

Les résultats (comparaison des densités et des quantiles) montrent que l'on arrive bien à estimer cette loi $\mathcal{E}(1)$. On notera tout de même sur le qq-plot que notre méthode a tendance à donner plus souvent des échantillons de grandes valeurs que ce que la vraie loi $\mathcal{E}(1)$ ferait.

Pour la deuxième loi à échantillonner, on prend une $\mathcal{N}(0,1)$. Cette fois aussi, on arrive très bien à l'échantillonner, et les quantiles estimés sont extrêmement proches des quantiles théoriques. On peut donc valider notre méthode.

Exercice 3 : Data augmentation

1.

On commence par montrer que le processus $\{(X_n, Y_n), n \geq 0\}$ est une chaîne de Markov. Pour cela, on considère la filtration canoniquement associée à ce processus : $(\mathcal{F}_n = \sigma(X_0, Y_0, ..., X_n, Y_n)_{n\in\mathbb{N}}$. On note $\mathcal{B}(\mathbb{R}^p \times \mathbb{R}^q)$ l'ensemble des boréliens de $\mathbb{R}^p \times \mathbb{R}^q$, alors on a $\mathcal{B}(\mathbb{R}^p \times \mathbb{R}^q) = \mathcal{B}(\mathbb{R}^p) \otimes \mathcal{B}(\mathbb{R}^q)$, c'est-à-dire que pour $A \in \mathcal{B}(\mathbb{R}^p \times \mathbb{R}^q)$, on a $A = A_1 \times A_2$, avec $A_1 \in \mathcal{B}(\mathbb{R}^p)$ et $A_2 \in \mathcal{B}(\mathbb{R}^q)$.

Soit $A \in \mathcal{B}(\mathbb{R}^p \times \mathbb{R}^q)$, alors

$$\begin{aligned}
\mathbb{P}((X_{n+1}, Y_{n+1}) \in A \mid \mathcal{F}_n) &= \mathbb{P}(X_{n+1} \in A_1, Y_{n+1} \in A_2 \mid \mathcal{F}_n) \\
&= \mathbb{P}(X_{n+1} \in A_1 \mid \mathcal{F}_n)\mathbb{P}(Y_{n+1} \in A_2 \mid \mathcal{F}_n, X_{n+1}) \text{ d'après la formule de Bayes,} \\
&= \left(\int_{A_1} f_{X|Y}(x, Y_n)\mathrm{d}x\right)\left(\int_{A_2} f_{Y|X}(X_{n+1}, y)\mathrm{d}y\right) \qquad (10) \\
&= \mathbb{P}(X_{n+1} \in A_1 \mid X_n, Y_n)\mathbb{P}(Y_{n+1} \in A_2 \mid, X_{n+1}, X_n, Y_n) \\
&= \mathbb{P}(X_{n+1} \in A_1, Y_{n+1} \in A_2 \mid X_n, Y_n)
\end{aligned}$$

Ainsi, $\{(X_n, Y_n), n \geq 0\}$ est une chaîne de Markov.

De plus, en reprenant la 3ème ligne de l'équation 10, on a que

$$\begin{aligned}
\mathbb{P}((X_n, Y_n) \in A \mid (X_{n-1}, Y_{n-1})) &= \left(\int_{A_1} f_{X|Y}(x, Y_{n-1})\mathrm{d}x\right)\left(\int_{A_2} f_{Y|X}(X_n, y)\mathrm{d}y\right) \\
&= \int_{A_1 \times A_2} f_{X|Y}(x, Y_{n-1})f_{Y|X}(x, y)\mathrm{d}x\mathrm{d}y
\end{aligned} \qquad (11)$$

d'après le théorème de Fubini-Tonelli car les densités sont positives. On a ainsi l'expression du noyau de transition de la chaîne de Markov.

2.

Montrons que le processus $(Y_n)_{n\in\mathbb{N}}$ est une chaîne de Markov. On considère dans cette question que $(\mathcal{F}_n)_{n\in\mathbb{N}}$ est la filtration canoniquement associée au processus $(Y_n)_{n\in\mathbb{N}}$. Soit $A \in \mathcal{B}(\mathbb{R}^q)$, alors on a

$$\begin{aligned}
\mathbb{P}(Y_{n+1} \in A \mid \mathcal{F}_n) &= \int_A f_{Y|X}(X_{n+1}, y)\mathrm{d}y \\
&= \int_{\mathbb{R}^p} f_{X|Y}(x, Y_n)\left(\int_A f_{Y|X}(x, y)\mathrm{d}y\right)\mathrm{d}x \\
&= \int_{\mathbb{R}^p}\int_A f_{X|Y}(x, Y_n)f_{Y|X}(x, y)\mathrm{d}y\mathrm{d}x \quad \text{par le théorème de Fubini-Tonelli} \\
&= \int_A \left(\int_{\mathbb{R}^p} f_{X|Y}(x, Y_n)f_{Y|X}(x, y)\mathrm{d}x\right)\mathrm{d}y \quad \text{encore par le théorème de Fubini-Tonelli} \\
&= \mathbb{P}(Y_{n+1} \\
&\in A \\
&\mid Y_n)
\end{aligned} \qquad (12)$$

Ainsi, $(Y_n)_{n\in\mathbb{N}}$ est bien une chaîne de Markov et son noyau de transition est :

$$P(y,A) = \int_A \left(\int_{\mathbb{R}^p} f_{X|Y}(x,y) f_{Y|X}(x,z)\mathrm{d}x\right)\mathrm{d}z$$

3.

On considère maintenant la fonction de densité f définie par

$$\forall (x,y)\in\mathbb{R}^2,\ f(x,y) = \frac{4}{\sqrt{2\pi}} y^{3/2}\exp\left[-y\left(\frac{x^2}{2}+2\right)\right]\mathbb{1}_{\mathbb{R}^+}(y)$$

On commence par calculer les densités marginales f_X et f_Y.

— Pour f_X,

$$f_X(x) = \int_R f(x,y)\mathrm{d}y = \frac{4}{\sqrt{2\pi}}\int_0^{+\infty}\exp\left[-y\left(\frac{x^2}{2}+2\right)\right]$$

On effectue un changement de variable affine $u = y\left(\frac{x^2}{2}+2\right)$ ce qui nous donne,

$$f_X(x) = \frac{4}{\sqrt{2\pi}}\frac{1}{\frac{x^2}{2}+2}\int_0^{+\infty}\left(\frac{1}{\frac{x^2}{2}+2}u\right)^{3/2}\exp(-u)\mathrm{d}u = \frac{4}{\sqrt{2\pi}}\frac{1}{(\frac{x^2}{2}+2)^{5/2}}\int_0^{+\infty}u^{3/2}\exp(-u)$$

Ainsi, on a

$$f_X(x) = \frac{4}{\sqrt{2\pi}}\frac{\Gamma\left(\frac{5}{2}\right)}{(\frac{x^2}{2}+2)^{5/2}}$$

— Pour f_Y,

$$f_Y(y) = \int_{\mathbb{R}} f(x,y)\mathrm{d}x = \frac{4}{\sqrt{2\pi}} y^{3/2}e^{-2y}\mathbb{1}_{\mathbb{R}^+}(y)\int_{\mathbb{R}}\exp(-y\frac{x^2}{2})\mathrm{d}x$$

Par le changement de variable affine $u = \sqrt{y}x$, on retrouve l'intégrale de Gauss :

$$f_Y(y) = \frac{4}{\sqrt{2\pi}} y^{3/2}e^{-2y}\mathbb{1}_{\mathbb{R}^+}(y)\int_{\mathbb{R}}\frac{1}{\sqrt{y}}\exp(-\frac{u^2}{2})\mathrm{d}u = 4ye^{-2y}\mathbb{1}_{\mathbb{R}^+}(y)$$

On peut maintenant donner les densités conditionnelles $f_{X|Y}$ et $f_{Y|X}$:

$$\begin{aligned} f_{X|Y}(x,y) &= \frac{f(x,y)}{f_Y(y)} = \frac{1}{\sqrt{2\pi}}\sqrt{y}\exp\left(-y\frac{x^2}{2}\right)\mathbb{1}_{\mathbb{R}^+}(y) \\ f_{Y|X}(x,y) &= \frac{f(x,y)}{f_Y(x)} = \frac{\left(\frac{x^2}{2}+2\right)^{5/2}}{\Gamma\left(\frac{5}{2}\right)} y^{3/2}\exp\left[-y\left(\frac{x^2}{2}+2\right)\right]\mathbb{1}_{\mathbb{R}^+}(y) \end{aligned} \tag{13}$$

Ainsi, $X_n|Y_{n-1} \sim \mathcal{N}\left(0, \frac{1}{\sqrt{Y_{n-1}}}\right)$ et $Y_n|X_n \sim \Gamma\left(\frac{5}{2}, \frac{X_n^2}{2}+2\right)$.

4.

Soit H une fonction bornée. On cherche à calculer

$$I = \int_{\mathbb{R}}\frac{H(x)}{(4+x^2)^{5/2}}\mathrm{d}x.$$

On peut utiliser l'échantillonneur de Gibbs des questions précédentes pour estimer l'intégrale I. En effet, on peut réécrire

$$I = \frac{1}{2^{5/2}}\int_{\mathbb{R}}\frac{H(x)}{(2+\frac{x^2}{2})^{5/2}}\mathrm{d}x = \frac{\sqrt{2\pi}}{4\Gamma\left(\frac{5}{2}\right)2^{5/2}}\mathbb{E}[H(X)]$$

où $X \sim f_X$. On peut donc simuler une suite $(X_n)_{n\in\mathbb{N}}$ grâce à l'algorithme précédent et ensuite utiliser un estimateur de Monte-Carlo :

$$I \approx \frac{\sqrt{2\pi}}{4\Gamma\left(\frac{5}{2}\right)2^{5/2}}\frac{1}{n}\sum_{k=0}^{n}H(X_k)$$