Correction du TD 12

Exercice 1 (Algorithme de Metropolis : un exemple jouet)
Rappel : Soit π une densité par rapport à une mesure de référence μ (ici on prendra μ la mesure de Lebesgue sur $\mathbb{R}$ et π la densité de $\mathcal{N}(0,1)$). On souhaite approcher une intégrale de la forme $I = \int \phi(x)\pi(x)\,\mathrm{d}\mu(x)$ pour une fonction ϕ π-intégrable.
L'objectif de l'algorithme de Métropolis-Hastings est de construire une chaîne de Markov $(X_n)_{n\geq 0}$ de densité stationnaire π. Le théorème ergodique assure alors que

$$\frac{1}{n}\sum_{i=0}^{n-1}\phi(X_i) \xrightarrow[n\to\infty]{\text{p.s.}} I.$$

Pour ce faire, on construit séquentiellement X_{n+1} à partir de X_n, de sorte que $X_{n+1} \,\big|\, X_n = x \sim p(x,\cdot)\cdot\mu$. Le noyau p est appelé noyau de transition de la chaîne de Markov. Afin que cette dernière soit de densité stationnaire π, le noyau de transition p est défini à travers X_{n+1} et X_n de la manière suivante : soient $\{q(x,\cdot), x\in\mathbb{R}\}$ une famille de densités par rapport à μ, le rapport de Metropolis-Hastings

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

et $U \sim \mathcal{U}_{[0,1]} \perp\!\!\!\perp X_n$. Alors en prenant $Y \,\big|\, X_n = x \sim q(x,\cdot)\cdot\mu$, on choisit $X_{n+1} = Y\mathbb{1}_{U\leq r(X_n,Y)} + X_n\mathbb{1}_{U>r(X_n,Y)}$. Le noyau de transition p défini de sorte que $p(x,\cdot)$ soit la densité de $X_{n+1} \,\big|\, X_n = x$ est alors bien de densité stationnaire π. Lorsque q est symétrique, c'est-à-dire que pour tout couple (x,y) on a $q(xy) = q(y,x)$, le rapport se simplifie en $r(x,y) = \min(1, \pi(y)/\pi(x))$ et on parle d'algorithme de Metropolis.

1. On rappelle que le noyau de proposition q définit une famille de densités $\{q(x,\cdot), x\in\mathbb{R}\}$ par rapport à la mesure de référence μ, qui est ici la mesure de Lebesgue sur $\mathbb{R}$. Dans le cadre de l'algorithme de Metropolis-Hastings, pour tout $x\in\mathbb{R}$, $q(x,\cdot)$ correspond à la densité de probabilité de la variable aléatoire Y candidate pour effectuer une transition (ou un pas) depuis l'état courant x.

   Sachant qu'on part du point x, la variable aléatoire Y suit une loi uniforme sur $[x-1/2, x+1/2]$, donc, pour tout $y\in\mathbb{R}$, $q(x,y) = \mathbb{1}_{[x-1/2,x+1/2]}(y) = \mathbb{1}_{|y-x|\leq 1/2}$. On en déduit que $q(y,x) = \mathbb{1}_{|x-y|\leq 1/2} = q(x,y)$. Ainsi le noyau de proposition q est symétrique et nous sommes donc dans le cadre d'application de l'algorithme de Metropolis, cas particulier de Metropolis-Hastings.

2. Dans le cadre de l'algorithme de Metropolis (q est symétrique), on part de X_0 quelconque puis, sachant $X_n = x$, on simule $W \sim \mathcal{U}_{[0,1]}$ et on définit $Y = W - \frac{1}{2} + x$. En notant φ la densité de la gaussienne centrée réduite, on calcule le rapport d'acceptation

   $$r(x,y) = \frac{\varphi(y)}{\varphi(x)} = e^{\frac{x^2-y^2}{2}},$$

   on tire une loi uniforme $U \sim \mathcal{U}([0,1])$. On pose enfin $X_{n+1} = Y$ si $U \leq r(x,y)$, sinon on reste sur place : $X_{n+1} = X_n$.

Exercice 2 (Metropolis vs Rejet)

1. Puisque $Y = x + \sigma W$, $q(x, \cdot)$ est la densité d'une gaussienne centrée en x et de matrice de covariance $\sigma^2 I_2$, c'est-à-dire en notant $\| \cdot \|$ la norme euclidienne :

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

   Cette formule étant symétrique en x et y, il est clair que $q(x, y) = q(y, x)$, donc on est dans le cadre de l'algorithme de Metropolis.

2. Appelons h la densité de la loi normale instrumentale. Il est alors nécessaire que $\frac{f(u,v)}{h(u,v)}$ soit majoré par une constante m pour tout (u, v). Or

$$\frac{f(u, v)}{h(u, v)} = \frac{c(\cos u)^2 (\sin v)^2 \, \mathrm{e}^{-0.05(u^2+v^2)}}{f(u, v)} \leq \frac{c \exp\left(-\frac{1}{20}(u^2 + v^2)\right)}{f(u, v)},$$

   où on a noté c la constante telle que cf intègre à 1. En remarquant que le numérateur de la borne correspond au terme générale de la loi $\mathcal{N}(0, 10I_2)$, il est judicieux de prendre pour h la densité de $\mathcal{N}(0, 10I_2)$, i.e.

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

   On a alors

$$\frac{f(u, v)}{h(u, v)} \leq 20\pi c.$$

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

$$r(u, v) = \frac{cf(u, v)}{mh(u, v)} = (\cos u)^2 (\sin v)^2.$$

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

Exercice 3 (Algorithme de Gibbs)

1. En intégrant par rapport à y la densité jointe, on voit que X suit une loi exponentielle de paramètre 1.

2. Sachant $X = x \geq 0$, on a

$$f_{Y \,|\, X=x}(y) \propto_y f(x, y) = \mathrm{e}^{-y} \mathbb{1}_{0 \leq x \leq y} \propto \mathrm{e}^{-(y-x)} \mathbb{1}_{y \geq x}.$$

   En d'autres termes, $Y \sim \mathcal{E}(1) + x$, loi exponentielle de paramètre 1 translatée de x.

3. Puisque l'on sait simuler facilement selon la loi de X et selon celle de $Y \,|\, X = x$, il suffit de construire d'abord une réaliation x de X, puis une réalisation y de Y sachant que $X = x$. Le couple (x, y) est alors une réalisation de (X, Y).

4. Sachant $Y = y \geq 0$,

$$f_{X \,|\, Y=y}(x) \propto_x f(x, y) \propto \mathbb{1}_{0 \leq x \leq y},$$

   donc $X \,|\, Y = y \sim \mathcal{U}([0, y])$.

5. Partant de (X_0, Y_0) vérifiant $f(X_0, Y_0) \geq 0$ p.s., l'algorithme de Gibbs (avec balayage déterministe) consiste à itérer des simulations suivant les schémas :
$$\begin{cases} X_{n+1} \,\middle|\, Y_n = y \sim \mathcal{U}([0, y]) \\ Y_{n+1} \,\middle|\, X_{n+1} = x \sim \mathcal{E}(1) + x. \end{cases}$$

6. La première méthode fournit directement un couple (X, Y) distribué selon la densité jointe f. Pour l'échantillonneur de Gibbs, à n fixé, la densité du couple (X_n, Y_n) n'est pas exactement f : cette propriété n'est qu'asymptotiquement vraie, lorsque $n \to \infty$. De plus, la première méthode fournit une suite indépendante, tandis qu'il y a une dépendance markovienne dans la suite de couples (X_n, Y_n) simulée par l'échantillonneur de Gibbs. Bref il est clair qu'il faut choisir la première méthode, dite de simulation exacte.
Remarque : un autre méthode de simulation exacte est possible, puisque le calcul de la loi marginale de Y montre que $Y \sim \Gamma(2, 1)$. Comme on a vu par ailleurs $X|Y = y \sim \mathcal{U}_{[0,y]}$, il suffit de simuler une loi Gamma puis une loi uniforme pour obtenir une réalisation de (X, Y).

Exercice 4 (Algorithme de Metropolis pour l'échantillonnage d'une loi a posteriori)

1. Soit $h : \mathbb{R} \to \mathbb{R}$ une fonction borélienne bornée. Alors, pour tout $\theta \in [0, 1]$, sachant $\boldsymbol{\theta} = \theta$,
$$\begin{aligned} \mathbb{E}[h(ZY_1 + (1 - Z)Y_{-1})] &= \mathbb{E}\left[\mathbb{E}\left[h(ZY_1 + (1 - Z)Y_{-1}) \,\middle|\, Z\right]\right] \\ &= \mathbb{P}(Z = 1)\mathbb{E}\left[h(ZY_1 + (1 - Z)Y_{-1}) \,\middle|\, Z = 1\right] + \\ &\quad\ \mathbb{P}(Z = 0)\mathbb{E}\left[h(ZY_1 + (1 - Z)Y_{-1}) \,\middle|\, Z = 0\right] \\ &= \theta\mathbb{E}\left[h(Y_1) \,\middle|\, Z = 1\right] + (1 - \theta)\mathbb{E}\left[h(Y_{-1}) \,\middle|\, Z = 0\right] \\ &= \theta\mathbb{E}\left[h(Y_1)\right] + (1 - \theta)\mathbb{E}\left[h(Y_{-1})\right] && \text{(par indépendance)} \\ &= \theta \int_{\mathbb{R}} h(y)\varphi(y - 1)\,\mathrm{d}y + (1 - \theta) \int_{\mathbb{R}} h(y)\varphi(y + 1)\,\mathrm{d}y \\ &= \int_{\mathbb{R}} h(y)\left(\theta\varphi(y - 1) + (1 - \theta)\varphi(y + 1)\right)\,\mathrm{d}y \\ &= \mathbb{E}\left[h(X_1) \,\middle|\, \boldsymbol{\theta} = \theta\right]. \end{aligned}$$
Par la méthode de la fonction muette, on a donc, sachant $\boldsymbol{\theta} = \theta$, $X_1 \stackrel{\mathcal{L}}{=} ZY_1 + (1 - Z)Y_{-1}$.
Pour simuler $\mathbf{X}$, il suffit donc de tirer une réalisation θ de la loi $\mathcal{U}([0, 1])$ puis n réalisations i.i.d. $z_1, \ldots, z_n$ de $\mathcal{B}(\theta)$. En prenant n réalisations i.i.d. $w_1, \ldots, w_n$ de $\mathcal{N}(0, 1)$, on choisit alors, pour tout $i \in \{1, \ldots, n\}$, $x_i = w_i + 1$ si $z_i = 1$ et $x_i = w_i - 1$ si $z_i = 0$. $(x_1, \ldots, x_n)$ est bien une réalisation de $\mathbf{X}$.

2. La densité a posteriori s'obtient par la formule de Bayes :
$$\pi(\theta \,|\, \mathbf{X}) = \frac{\pi(\theta)p_\theta(\mathbf{X})}{f(\mathbf{X})} = C_{\mathbf{X}} \prod_{i=1}^{n} \left\{\theta\phi(X_i - 1) + (1 - \theta)\phi(X_i + 1)\right\} \mathbb{1}_{[0,1]}(\theta),$$
où, comme d'habitude, la constante de normalisation
$$C_{\mathbf{X}} = \left(\int_0^1 \prod_{i=1}^{n} \left\{t\phi(X_i - 1) + (1 - t)\phi(X_i + 1)\right\} dt\right)^{-1}$$
dépend de $\mathbf{X}$ mais pas de θ. Un estimateur de Bayes pour la perte quadratique $\widehat{\theta}_n$ est par définition $\mathbb{E}[\boldsymbol{\theta} \,|\, \mathbf{X}]$, c'est-à-dire
$$\widehat{\theta}_n = \int_{\mathbb{R}} \theta\ \pi(\theta \,|\, \mathbf{X})\,\mathrm{d}\theta = C_{\mathbf{X}} \int_0^1 \theta \prod_{i=1}^{n} \left\{\theta\phi(X_i - 1) + (1 - \theta)\phi(X_i + 1)\right\}\,\mathrm{d}\theta.$$

On ne dispose pas de formule analytique évidente pour calculer cette intégrale, d'où le recours aux méthodes Monte-Carlo dans la suite.

3. Nous avons déjà vu que

$$\widehat{\theta}_n = \mathbb{E}[\boldsymbol{\theta} \,|\, \mathbf{X}] = \int_{\mathbb{R}} \theta\pi(\theta|\mathbf{X})\,\mathrm{d}\theta = \int_{\mathbb{R}} \theta \left\{ \frac{p_\theta(\mathbf{X})\pi(\theta)}{\int_{\mathbb{R}} p_t(\mathbf{X})\pi(t)\,\mathrm{d}t} \right\} \mathrm{d}\theta = \frac{\int_{\mathbb{R}} \theta p_\theta(\mathbf{X})\pi(\theta)\,\mathrm{d}\theta}{\int_{\mathbb{R}} p_t(\mathbf{X})\pi(t)\,\mathrm{d}t},$$

où $p_\theta(\mathbf{X}) = \prod_{i=1}^n [\theta\varphi(x-1) + (1-\theta)\varphi(x+1)]$ est la densité de la loi de $\mathbf{X} \,|\, \boldsymbol{\theta} = \theta$, faisant intervenir la densité φ de la loi $\mathcal{N}(0,1)$. Numérateur et dénominateur peuvent être vus comme des espérances par rapport à la loi a priori sur $\boldsymbol{\theta}$, à savoir :

$$\widehat{\theta}_n = \varphi(\mathbf{X}), \qquad \text{où} \qquad \varphi(x) = \frac{\mathbb{E}\left[\boldsymbol{\theta} p_{\boldsymbol{\theta}}(x)\right]}{\mathbb{E}\left[p_{\boldsymbol{\theta}}(x)\right]}.$$

Cela conduit à l'estimateur Monte-Carlo :

$$\widehat{\theta}_n^N = \frac{\frac{1}{N}\sum_{j=1}^N \boldsymbol{\theta}_j \; p_{\boldsymbol{\theta}_j}(\mathbf{X})}{\frac{1}{N}\sum_{j=1}^N p_{\boldsymbol{\theta}_j}(\mathbf{X})} = \frac{\sum_{j=1}^N \theta_j \prod_{i=1}^n \{\theta_j\varphi(X_i-1) + (1-\theta_j)\varphi(X_i+1)\}}{\sum_{j=1}^N \prod_{i=1}^n \{\theta_j\varphi(X_i-1) + (1-\theta_j)\varphi(X_i+1)\}},$$

où $(\boldsymbol{\theta}_i)_{1\leq i\leq N}$ sont N variables aléatoires réelles i.i.d. selon la loi a priori (et indépendantes de $X_1, \ldots, X_n$).

4. Clairement, la loi de proposition correspondante à q, dont la densité est, pour tout $\theta \in \mathbb{R}$, $q(\theta,\cdot)$, est la loi $\mathcal{U}([0,1])$. Elle est indépendante de θ (en particulier, ce n'est pas une loi translatée comme dans les cas précédents), mais ce n'est qu'une proposition : le prochain point de la chaîne de Markov dépendra, lui, bien du point courant (cf le rapport de Metropolis-Hastings).

   Par ailleurs, le rapport de Metropolis-Hastings vaut, pour tous $\theta \in \mathbb{R}$ et $\theta \in \mathbb{R}$:

$$\begin{aligned} r(\theta,\theta') &= \frac{\pi(\theta' \,|\, \mathbf{X})}{\pi(\theta \,|\, \mathbf{X})} \frac{q(\theta',\theta)}{q(\theta,\theta')} = \frac{\pi(\theta')p_{\theta'}(\mathbf{X})}{\pi(\theta)p_\theta(\mathbf{X})} \frac{q(\theta',\theta)}{q(\theta,\theta')} = \frac{\mathbb{1}_{[0,1]}(\theta')p_{\theta'}(\mathbf{X})}{\mathbb{1}_{[0,1]}(\theta)p_\theta(\mathbf{X})} \frac{\mathbb{1}_{[0,1]}(\theta)}{\mathbb{1}_{[0,1]}(\theta')} \\ &= \frac{p_{\theta'}(\mathbf{X})}{p_\theta(\mathbf{X})} = \frac{\prod_{i=1}^n \{\theta'\varphi(X_i-1) + (1-\theta')\varphi(X_i+1)\}}{\prod_{i=1}^n \{\theta\varphi(X_i-1) + (1-\theta)\varphi(X_i+1)\}}. \end{aligned}$$

Exercice 5 (Slice sampler)

1. On effectue le changement de variable $u = \sqrt{x}$ pour obtenir

$$\int_0^{+\infty} g(x)\,\mathrm{d}x = \int_0^{+\infty} u\,\mathrm{e}^{-u}du = 1,$$

   puisqu'on a reconnu la moyenne d'une loi exponentielle de paramètre 1. Par le théorème de Fubini-Tonelli, puisque $f(u,x)$ est positive, on a donc

$$\iint_{\mathbb{R}^2} f(u,x)\,\mathrm{d}u\,\mathrm{d}x = \int_0^{+\infty} \left(\int_0^{+\infty} f(u,x)\,\mathrm{d}u \right) \mathrm{d}x = \int_0^{+\infty} g(x)\,\mathrm{d}x = 1,$$

   ce qui prouve que f est bien une densité sur $\mathbb{R}^2$.

2. Pour tout $x > 0$ fixé, on a (avec des notations évidentes)

$$f_{U \,|\, X=x}(u) = \frac{f(u,x)}{g(x)} = \frac{\mathbb{1}_{0<u<\frac{1}{2}\exp(-\sqrt{x})}}{\frac{1}{2}\exp(-\sqrt{x})},$$

ce qui montre que, sachant $X = x$, U suit une loi uniforme sur $\left[0, \frac{\mathrm{e}^{-\sqrt{x}}}{2}\right]$, dont on note h la densité. De la même façon, pour tout $u > 0$, puisque

$$0 < u < \frac{1}{2}\exp(-\sqrt{x}) \iff 0 < x < (\log(2u))^2,$$

on en déduit que

$$f_{X \,|\, U=u}(x) = \frac{f(u,x)}{h(u)} = \frac{\mathbb{1}_{0<x<(\log(2u))^2}}{(\log(2u))^2},$$

ce qui montre que, sachant $u > 0$, X suit une loi uniforme sur $[0, (\log(2u))^2]$.

3. Partant de (U_0, X_0) vérifiant $f(U_0, X_0) \geq 0$ p.s., l'algorithme de Gibbs consiste à itérer des simulations suivant les schémas :

$$\begin{cases} U_{n+1} \,|\, X_n = x \sim \mathcal{U}(\left[0, \frac{\mathrm{e}^{-\sqrt{x}}}{2}\right]) \\ X_{n+1} \,|\, U_{n+1} = u \sim \mathcal{U}([0, (\log(2u))^2]). \end{cases}$$