Mathias Marciano

TP3 : Hasting-Metropolis (and Gibbs) samplers

Exercice 1 :

. Log-vraisemblance du modèle

Étant donné que : $q(y,z,\theta) = q(y|z,\theta)\,q(z,\theta) = q(y|z,\theta)\,q(z|\theta)\,q(\theta)$

On a : $\log(q(y|z,\theta)) = \log(q(y|z,\theta)) + \log(q(z|\theta)) + \log(q(\theta))$

On suppose par la suite que $k_i = K \quad \forall i \in [\![1;N]\!]$

- $q(y|z,\theta) = \prod_{i=1}^{N}\prod_{j=1}^{K} q(y_{ij}|z,\theta)$ où $y_{ij}|z,\theta \sim \mathcal{N}(d_i(t_{ij}),\sigma^2)$

$\Rightarrow \log(q(y|z,\theta)) = \sum_{i=1}^{N}\sum_{j=1}^{K} \log(q(y_{ij}|z,\theta))$

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

- $q(z|\theta) = q(z_{pop}|\theta)\prod_{i=1}^{N} q(z_i|\theta)$; $z_{pop} = (t_0, v_0)$ avec $t_0 \sim \mathcal{N}(\bar{t_0}, \sigma_{t_0}^2)$, $v_0 \sim \mathcal{N}(\bar{v_0}, \sigma_{v_0}^2)$

$\Rightarrow \log(q(z|\theta)) = \log(q(z_{pop}|\theta)) + \sum_{i=1}^{N}\log(q(z_i|\theta))$

. $\log(q(z_{pop}|\theta)) = -\frac{1}{2}\log(\sigma_{t_0}^2) - \frac{1}{2\sigma_{t_0}^2}(t_0-\bar{t_0})^2 - \frac{1}{2}\log(\sigma_{v_0}^2) - \frac{1}{2\sigma_{v_0}^2}(v_0-\bar{v_0})^2 + c$

avec c une constante

. $z_i = (\alpha_i, \tau_i)$ avec $\alpha_i = \exp(\xi_i)$ et $\xi_i \sim \mathcal{N}(0, \sigma_\xi^2)$

$\tau_i \sim \mathcal{N}(0, \sigma_\tau^2)$

$\Rightarrow \sum_{i=1}^{N}\log(q(z_i|\theta)) = \sum_{i=1}^{N}\left(-\frac{1}{2}\frac{\xi_i^2}{\sigma_\xi^2} - \frac{1}{2}\log(\sigma_\xi^2) - \frac{1}{2\sigma_\tau^2}\tau_i^2 - \frac{1}{2}\log(\sigma_\tau^2)\right) + c$

avec c une constante

. $\theta = (\bar{t_0}, \bar{v_0}, \sigma_\xi, \sigma_\tau, \sigma)$ avec $\bar{t_0} \sim \mathcal{N}(\bar{\bar{t_0}}, s_{t_0}^2)$

$\bar{v_0} \sim \mathcal{N}(\bar{\bar{v_0}}, s_{v_0}^2)$

$\sigma_\xi^2 \sim W^{-1}(v_\xi, m_\xi)$; $\sigma_\tau^2 \sim W^{-1}(v_\tau, m_\tau)$ et $\sigma^2 \sim W^{-1}(v, m)$

où W^{-1} est la loi inverse de Wishart

$\Rightarrow \log(q(\theta)) = -\frac{1}{2s_{t_0}^2}(\bar{t_0} - \bar{\bar{t_0}})^2 - \frac{1}{2s_{v_0}^2}(\bar{v_0} - \bar{\bar{v_0}})^2 + m_\xi\log(v_\xi/\sigma_\xi) - \frac{v_\xi^2}{2\sigma_\xi^2} - \log(\sigma_\xi^2)$

$+ m_\tau\log(v_\tau/\sigma_\tau) - \log(\sigma_\tau^2) - \frac{v_\tau^2}{2\sigma_\tau^2}$

$+ m\log(v/\sigma) - \log(\sigma^2) - \frac{v^2}{2\sigma^2} + K_1$ avec K_1 une constante

Finalement :

$$\boxed{\begin{aligned}\log(q(y,z,\theta)) = &\sum_{i=1}^{N}\sum_{j=1}^{k}\left(-\frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y_{ij} - d_i(t_{ij}))^2\right)\\ &- \frac{1}{2}\left(\frac{t_0 - \bar{t_0}}{\sigma_{t_0}}\right)^2 - \log(\sigma_{t_0}) - \frac{1}{2}\left(\frac{v_0 - \bar{v_0}}{\sigma_{v_0}}\right)^2 - \log(\sigma_{v_0})\\ &+ \sum_{i=1}^{N}\left(-\frac{\xi_i^2}{2\sigma_\xi^2} - \log(\sigma_\xi) - \frac{\tau_i^2}{2\sigma_\tau^2} - \log(\sigma_\tau)\right)\\ &- \frac{1}{2}\left(\frac{\bar{t_0} - \bar{\bar{t_0}}}{s_{t_0}}\right)^2 - \frac{1}{2}\left(\frac{\bar{v_0} - \bar{\bar{v_0}}}{s_{v_0}}\right)^2\\ &+ m_\xi\log(v_\xi/\sigma_\xi) - \log(\sigma_\xi^2) - \frac{1}{2}\left(\frac{v_\xi}{\sigma_\xi}\right)^2\\ &+ m_\tau\log(v_\tau/\sigma_\tau) - \log(\sigma_\tau^2) - \frac{1}{2}\left(\frac{v_\tau}{\sigma_\tau}\right)^2\\ &+ m\log(v/\sigma) - \log(\sigma^2) - \frac{1}{2}\left(\frac{v}{\sigma}\right)^2 + C\end{aligned}}$$

avec C une certaine constante

Montrons que ce modèle appartient à la "curved exponential family"

i.e que : $\log(q(y,z,\theta)) = \phi_1(\theta) + \phi_2(y,z) + \langle S(y,z), \Psi(\theta)\rangle_{\mathbb{R}^5}$

avec $z = (t_0, v_0, \xi_{i\in[\![1,N]\!]}, \tau_{i\in[\![1,N]\!]})$

et $\theta = (\sigma, \sigma_\xi, \sigma_\tau, \bar{t_0}, \bar{v_0})$

Par identification, on a :

$S_1 = \frac{1}{kN}\sum_{i=1}^{N}\sum_{j=1}^{k}(y_{ij} - d_i(t_{ij}))^2$; $\Psi_1 = -\frac{kN}{2\sigma^2}$

$S_2 = \frac{1}{N}\sum_{i=1}^{N}\xi_i^2$; $\Psi_2 = -\frac{N}{2\sigma_\xi^2}$

$S_3 = \frac{1}{N}\sum_{i=1}^{N}\tau_i^2$; $\Psi_3 = -\frac{N}{2\sigma_\tau^2}$

$S_4 = t_0$; $\Psi_4 = \frac{\bar{t_0}}{\sigma_{t_0}^2}$

$S_5 = v_0$; $\Psi_5 = \frac{\bar{v_0}}{\sigma_{v_0}^2}$

$$\begin{aligned}\phi_1(\theta) = &-(Nk + m + 2)\log(\sigma) - (N + m_\xi + 2)\log(\sigma_\xi)\\ &-(N + m_\tau + 2)\log(\sigma_\tau) - \frac{\bar{t_0}^2}{2}\left(\frac{1}{\sigma_{t_0}^2} + \frac{1}{s_{t_0}^2}\right) - \frac{\bar{v_0}^2}{2}\left(\frac{1}{\sigma_{v_0}^2} + \frac{1}{s_{v_0}^2}\right)\\ &+ \frac{\bar{t_0}\bar{\bar{t_0}}}{s_{t_0}^2} + \frac{\bar{v_0}\bar{\bar{v_0}}}{s_{v_0}^2} - \frac{1}{2}\left((v_\xi/\sigma_\xi)^2 + (v_\tau/\sigma_\tau)^2 + (v/\sigma)^2\right)\end{aligned}$$

$$\phi_2(y,z) = -\frac{1}{2}\left(\frac{r_0}{\sigma_{r_0}}\right)^2 - \frac{1}{2}\left(\frac{v_0}{\sigma_{v_0}}\right)^2 - \frac{1}{2}\left(\frac{\bar{\bar{r}}_0}{s_{r_0}}\right)^2 - \frac{1}{2}\left(\frac{\bar{\bar{v}}_0}{s_{v_0}}\right)^2$$

$$+ m_\xi \log(v_\xi) + m_\tau \log(v_\tau) + m\log(v) + c$$

→ Donc ce modèle appartient à la "curved exponential family)

3) On a : $\pi(z) = p(z \mid y, \theta) \underset{\text{Bayes}}{=} p(y \mid z, \theta)\, p(z \mid \theta)\, p(\theta) \times \frac{1}{p(y,\theta)}$

donc $\dfrac{\pi(z^*)}{\pi(z^{(a)})} = \dfrac{p(y \mid z^*, \theta)\, p(z^* \mid \theta)}{p(y \mid z^{(a)}, \theta)\, p(z^{(a)} \mid \theta)}$

→ Ainsi il ne reste que des termes qui dépendent de z, ce qui nous sera utile lors du calcul du alpha.

Exercice 2 : Multiplicative Hastings-Metropolis

1) On a pour $B \sim \mathcal{B}er(1/2)$:

$$Y = \mathbb{1}_{\{B=1\}} \varepsilon X + \mathbb{1}_{\{B=0\}} \frac{X}{\varepsilon}$$

Déterminons la densité de Y à l'aide de la fonction test :

Soit $g : \mathbb{R} \to \mathbb{R}$ une fonction continue, bornée :

$$\mathbb{E}[g(Y)] = \frac{1}{2}\mathbb{E}[g(\varepsilon X)] + \frac{1}{2}\mathbb{E}[g(X/\varepsilon)] \quad (\text{car } B \perp\!\!\!\perp \varepsilon \text{ et } B \perp\!\!\!\perp X)$$

• $\mathbb{E}[g(\varepsilon X)] = \int_{-1}^{1} g(Xu) f(u)\,du$

On pose le changement de variable $v = Xu$ de $]-|X|;|X|[$ dans $]-1;1[$:

$$\mathbb{E}[g(\varepsilon X)] = \int_{-|X|}^{|X|} g(v) f(v/X) \frac{1}{|X|}\,dv$$

• $\mathbb{E}[g(X/\varepsilon)] = \int_{-1}^{1} g(X/u) f(u)\,du$

On pose $v = X/u$ de $]-\infty; -|X|[$ dans $]0; -\frac{X}{|X|}[$ et de $]|X|; +\infty[$ dans $]0; \frac{X}{|X|}[$ pour avoir une bijection

$$\Rightarrow \mathbb{E}[g(X/\varepsilon)] = \int_{-\infty}^{-|X|} g(v) f(X/v) |X|/v^2\,dv + \int_{|X|}^{+\infty} g(v) f(X/v) \frac{|X|}{v^2}\,dv$$

D'où on a :

$$\mathbb{E}[g(Y)] = \frac{1}{2}\Big[\int_{-\infty}^{-|X|} g(v) f(X/v) |X|/v^2\,dv + \int_{-|X|}^{|X|} g(v) f(v/X) \frac{1}{|X|}\,dv + \int_{|X|}^{+\infty} g(v) f(X/v) \frac{|X|}{v^2}\,dv\Big]$$

Par identification :

$$\boxed{q(X,Y) = \mathbb{1}_{\{|Y|>|X|\}} f(X/Y) \frac{|X|}{2Y^2} + \mathbb{1}_{\{|Y|<|X|\}} f\left(\frac{Y}{X}\right) \frac{1}{2|X|}}$$

2) Calculons le taux d'acceptation α :

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

Or : $$\frac{\pi(Y)\, q(Y,X)}{\pi(X)\, q(X,Y)} = \frac{\pi(Y)}{\pi(X)} \, \frac{1\!\!1_{\{|X|>|Y|\}} f(Y/X)\, |Y|/2X^2 + 1\!\!1_{\{|X|<|Y|\}} f(X/Y) \frac{1}{2|Y|}}{1\!\!1_{\{|Y|>|X|\}} f(X/Y)\, |X|/2Y^2 + 1\!\!1_{\{|Y|<|X|\}} f(Y/X) \frac{1}{2|X|}}$$

$$= \frac{\pi(Y)}{\pi(X)} \left[1\!\!1_{\{|X|>|Y|\}} \frac{|Y|}{|X|} + 1\!\!1_{\{|X|<|Y|\}} \frac{|Y|}{|X|} \right]$$

$$= \frac{\pi(Y)}{\pi(X)} \, \frac{|Y|}{|X|}$$

$$\Rightarrow \boxed{\alpha(X,Y) = \min\left(1, \frac{\pi(Y)\,|Y|}{\pi(X)\,|X|}\right)}$$

Exercice 3 : Data augmentation

1) Soit le processus bivarié $\{(X_n, Y_n), n \geq 0\}$ tel que $\forall n \geq 1$,
$X_n \sim f_{X|Y}(\cdot, Y_{n-1})$ et $Y_n \sim f_{Y|X}(X_n, \cdot)$

On remarque que X_{n+1} est Y_n-mesurable et Y_{n+1} est X_{n+1}-mesurable

donc Y_{n+1} est Y_n-mesurable.

Finalement, (X_{n+1}, Y_{n+1}) est (X_n, Y_n)-mesurable.

Soient $A \in \mathcal{B}(\mathbb{R}^p)$ et $B \in \mathcal{B}(\mathbb{R}^q)$ deux boréliens et $\mathcal{F}_n = \sigma((X_0, Y_0), \dots (X_n, Y_n))$

$$\mathbb{P}((X_{n+1}, Y_{n+1}) \in A \times B \mid \mathcal{F}_n] = \mathbb{E}[\mathbb{1}_{\{(X_{n+1}, Y_{n+1}) \in A\times B\}} \mid \mathcal{F}_n]$$

$\sigma(X_n, Y_n) \subset \mathcal{F}_n$

$$= \mathbb{E}[\mathbb{E}[\mathbb{1}_{\{(X_{n+1}, Y_{n+1}) \in A\times B\}} \mid \mathcal{F}_n] \mid X_n, Y_n]$$

$$= \mathbb{E}[\mathbb{1}_{\{(X_{n+1}, Y_{n+1}) \in A\times B\}} \mid X_n, Y_n]$$

$$= \mathbb{P}((X_{n+1}, Y_{n+1}) \in A \times B \mid X_n, Y_n)$$

On a bien la propriété de Markov donc $\{(X_n, Y_n), n \geq 0\}$ est une chaîne de Markov.

Calculons son noyau de transition :

$$\boxed{\mathbb{P}((X_{n+1}, Y_{n+1}) \in A \times B \mid X_n, Y_n) = \int_{A\times B} f_{X|Y}(x, Y_n) f_{Y|X}(x, y)\, dx\, dy}$$

2) Montrons que $\{Y_n, n \geq 0\}$ est une chaîne de Markov :

Y_n est Y_{n-1}-mesurable d'après précédemment.

Soit $A \in \mathcal{B}(\mathbb{R}^q)$ et $\mathcal{G}_n = \sigma(Y_0, \dots Y_n)$

$$\mathbb{P}(Y_{n+1} \in A \mid \mathcal{G}_n) = \mathbb{E}[\mathbb{1}_{\{Y_{n+1} \in A\}} \mid \mathcal{G}_n]$$

$$= \mathbb{E}[\mathbb{E}[\mathbb{1}_{\{Y_{n+1} \in A\}} \mid \mathcal{G}_n] \mid Y_n] \quad \text{car } \sigma(Y_n) \subset \mathcal{G}_n$$

$$= \mathbb{E}[\mathbb{1}_{\{Y_{n+1} \in A\}} \mid Y_n]$$

$$\boxed{\mathbb{P}(Y_{n+1} \in A \mid \mathcal{G}_n) = \mathbb{P}(Y_{n+1} \in A \mid Y_n)}$$

Donc $\{Y_n, n \geq 0\}$ est une chaîne de Markov.

Son noyau de transition est donné par :

$$\mathbb{P}(Y_{n+1} \in A \mid Y_n) = \mathbb{P}(Y_{n+1} \in A,\ X_{n+1} \in \mathbb{R}^p \mid Y_n)$$

$$\boxed{\mathbb{P}(Y_{n+1} \in A \mid Y_n) = \int_{A \times \mathbb{R}^p} f_{X|Y}(x, Y_n)\, f_{Y|X}(x, y)\, dy\, dx}$$

Montrons que $f_Y(y)dy$ est invariant pour son noyau de transition :

$$\mu P_Y(\cdot, A) = \int_{\mathbb{R}^d} P_Y(y, A) f_Y(y) dy$$

$$= \int_{\mathbb{R}^d} \int_{A \times \mathbb{R}^p} f_{X|Y}(x, y)\, f_{Y|X}(x, \tilde{y})\, f_Y(y)\, dy\, d\tilde{y}\, dx$$

$$= \int_{\mathbb{R}^d \times A \times \mathbb{R}^p} f(x, y)\, f_{Y|X}(x, \tilde{y})\, dy\, d\tilde{y}\, dx$$

car $f_{X|Y}(\cdot, \cdot) = \dfrac{f(\cdot, \cdot)}{f_Y(\cdot)}$

} loi "marginale" de X

$$\Rightarrow \mu P_Y(\cdot, A) = \int_{A \times \mathbb{R}^p} f_X(x) \frac{f(x, \tilde{y})}{f_X(x)}\, d\tilde{y}\, dx$$

$$= \int_{A \times \mathbb{R}^p} f(x, \tilde{y})\, d\tilde{y}\, dx$$

$$= \int_A f_Y(\tilde{y})\, d\tilde{y}$$

$$\boxed{\mu P_Y(\cdot, A) = \mu(A)}$$

$\rightarrow$ $f_Y(y)dy$ est invariante pour le noyau de transition.

3) Il suffit de calculer $f_{X|Y}$ et $f_{Y|X}$ i.e calculer f_X et f_Y car on dispose déjà de f !

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy = \frac{4}{\sqrt{2\pi}} \int_{\mathbb{R}} y^{3/2} \exp\left(-y\left(\frac{x^2}{2}+2\right)\right) \mathbb{1}_{\{y \geq 0\}}\,dy$$

Posons $\alpha = \frac{5}{2}$ et $\beta = \frac{x^2}{2} + 2$ pour faire apparaître la densité d'une loi Gamma !

$$\Rightarrow f_X(x) = \frac{4}{\sqrt{2\pi}} \times \frac{\Gamma(\alpha)}{\beta^\alpha} \underbrace{\int_{\mathbb{R}} y^{\alpha-1} \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta y} \mathbb{1}_{\{y\geq 0\}}\,dy}_{\text{densité loi Gamma}(\alpha,\beta)}$$

$$\boxed{f_X(x) = \frac{4}{\sqrt{2\pi}} \frac{\Gamma(\alpha)}{\beta^\alpha}}$$ (β dépend de x)

$$\Rightarrow f_{Y|X}(x,y) = \frac{f(x,y)}{f_X(x)} = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\beta y} \mathbb{1}_{\{y \geq 0\}}$$

$$\Rightarrow \boxed{Y|X \sim \Gamma\left(\frac{5}{2}, \frac{X^2}{2} + 2\right)}$$

- $$f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx = 4 y^{3/2} e^{-2y} \mathbb{1}_{\{y\geq 0\}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp(-yx^2/2)\,dx$$

$$= 4 y^{3/2} e^{-2y} \mathbb{1}_{\{y \geq 0\}} y^{-1/2} \underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi \times (\frac{1}{\sqrt{y}})^2}} \exp\left(-x^2/2 \times (1/\sqrt{y})^2\right) dx}_{\mathcal{N}(0, \frac{1}{y})}$$

$$\Rightarrow \boxed{f_Y(y) = 4 y e^{-2y} \mathbb{1}_{\{y \geq 0\}}}$$

Donc : $$f_{X|Y}(x,y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{\sqrt{2\pi \times (\frac{1}{\sqrt{y}})^2}} \exp\left(-x^2/2\left(\frac{1}{\sqrt{y}}\right)^2\right)$$

$$\Rightarrow \boxed{X|Y \sim \mathcal{N}\left(0, \frac{1}{Y}\right)}$$

En appliquant l'algorithme 4 on a ainsi le pseudo-code suivant :

Given $(X_0, Y_0) \in \mathbb{R}^p \times \mathbb{R}^d$ and $N \in \mathbb{N}$

for $n = 1$ to N do

$$X_n \sim \mathcal{N}\left(0, \frac{1}{Y_{n-1}}\right)$$

$$Y_n \sim \Gamma\left(5/2, \frac{X_n^2}{2} + 2\right)$$

end

return $\{(X_n, Y_n), 0 \leq n \leq N\}$

4) On souhaite approximer : $I = \int_{\mathbb{R}} \frac{H(x)}{(4+x^2)^{5/2}} dx$

Or, il s'agit de remarquer que :

$$I = \frac{\sqrt{2\pi}}{4\,\Gamma(5/2)} \int_{\mathbb{R}} H(x) f_X(x) dx = K \underset{X \sim f_X}{\mathbb{E}} [H(X)]$$

avec $K = \frac{\sqrt{2\pi}}{4\,\Gamma(5/2)}$

En disposant de l'échantillon de Gibbs $\{(X_n, Y_n), 0 \leq n \leq N\}$, on peut approximer l'espérance avec la quantité :

$\frac{1}{N} \sum_{i=1}^{N} H(X_i)$ qui tend presque sûrement quand $N \to +\infty$ vers $\underset{X \sim f_X}{\mathbb{E}} [H(X)]$. Ainsi, on approxime I par :

$$\boxed{I_N = \frac{K}{N} \sum_{i=1}^{N} H(X_i) \underset{N \to +\infty}{\longrightarrow} I}$$