Bastien LE CHENADEC
bastien.le-chenadec@eleves.enpc.fr

TP 3
Computational statistics

8 décembre 2023

Exercice 1 Hasting-Metropolis within Gibbs

1. Les paramètres du modèle sont $\theta = (\bar{t}_0, \bar{v}_0, \sigma_\xi, \sigma_\tau, \sigma)$. On s'est donné les variables latentes $z_{\text{pop}} = (t_0, v_0)$ et $\forall i \in [\![1, N]\!], z_i = (\alpha_i, \tau_i)$. En utilisant la formule de Bayes et par indépendance des z_i et de z_{pop} :

$$\begin{aligned}
\log p(y, z, \theta) &= \log p(y|z, \theta) + \log p(z_{\text{pop}}|\theta) + \sum_{i=1}^{N} \log p(z_i|\theta) + \log p(\theta) \\
&= ① + ② + ③ + ④
\end{aligned}$$

Pour simplifier les notations on prend $\forall i \in [\![1, N]\!], k_i = k$. En prenant le logarithme du produit des densités gaussiennes on obtient :

$$\begin{aligned}
① &= -\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{k} \left(\frac{y_{i,j} - d_i(t_{i,j})}{\sigma} \right)^2 - kN \log \sigma + \text{cst} \\
② &= -\frac{1}{2} \left(\frac{t_0 - \bar{t}_0}{\sigma_{t_0}} \right)^2 - \log(\sigma_{t_0}) - \frac{1}{2} \left(\frac{v_0 - \bar{v}_0}{\sigma_{v_0}} \right)^2 - \log(\sigma_{v_0}) + \text{cst} \\
③ &= -\frac{1}{2} \left[\sum_{i=1}^{N} \left(\frac{\xi_i}{\sigma_\xi} \right)^2 + \left(\frac{\tau_i}{\sigma_\tau} \right)^2 \right] - N \log \sigma_\xi - N \log \sigma_\tau + \text{cst}
\end{aligned}$$

Pour ④ il nous faut calculer le logarithme de la densité gamma inverse :

$$\begin{aligned}
\log f_{\mathcal{W}^{-1}(v,m)}(\sigma^2) &= -2 \log \sigma - m \log \sigma - \frac{v^2}{2\sigma^2} \\
&= -\frac{v^2}{2\sigma^2} - (m + 2) \log \sigma
\end{aligned}$$

Ce qui nous permet d'obtenir, sachant que $\bar{t}_0 \sim \mathcal{N}(\bar{\bar{t}}_0, s_{t_0}^2)$, $\bar{v}_0 \sim \mathcal{N}(\bar{\bar{v}}_0, s_{v_0}^2)$, $\sigma_\xi^2 \sim \mathcal{W}^{-1}(v_\xi, m_\xi)$, $\sigma_\tau^2 \sim \mathcal{W}^{-1}(v_\tau, m_\tau)$ et $\sigma^2 \sim \mathcal{W}^{-1}(v, m)$:

$$\begin{aligned}
④ = &-\frac{1}{2} \left(\frac{\bar{t}_0 - \bar{\bar{t}}_0}{s_{t_0}} \right)^2 - \log s_{t_0} - \frac{1}{2} \left(\frac{\bar{v}_0 - \bar{\bar{v}}_0}{s_{v_0}} \right)^2 - \log s_{v_0} \\
&- \frac{v_\xi^2}{2\sigma_\xi^2} - (m_\xi + 2) \log \sigma_\xi - \frac{v_\tau^2}{2\sigma_\tau^2} - (m_\tau + 2) \log \sigma_\tau - \frac{v^2}{2\sigma^2} - (m + 2) \log \sigma + \text{cst}
\end{aligned}$$

On souhaite maintenant écrire cette log-vraisemblance sous la forme $-\phi(\theta) + \langle S(y, z) | \psi(\theta) \rangle$. On sépare bien les termes dépendant de y et z, ceux dépendant de θ et ceux constants. Par rapport à ce qu'on a fait en TD, on a rajouté une sixième coordonnée pour les termes qui ne dépendent que de y et z, et on a corrigé l'expression de $\phi(\theta)$.

On pose : $\begin{cases} S_1 = t_0 \\ S_2 = v_0 \\ S_3 = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \xi_i^2 \\ S_4 = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \tau_i^2 \\ S_5 = \dfrac{1}{kN}\displaystyle\sum_{i=1}^{N}\sum_{j=1}^{k} (y_{i,j} - d_i(t_{i,j}))^2 \\ S_6 = \dfrac{t_0^2}{\sigma_{t_0}^2} + \dfrac{v_0^2}{\sigma_{v_0}^2} \end{cases}$ et : $\begin{cases} \psi_1 = \dfrac{\bar{t_0}}{\sigma_{t_0}^2} \\ \psi_2 = \dfrac{\bar{v_0}}{\sigma_{v_0}^2} \\ \psi_3 = -\dfrac{N}{2\sigma_\xi^2} \\ \psi_4 = -\dfrac{N}{2\sigma_\tau^2} \\ \psi_5 = -\dfrac{kN}{2\sigma^2} \\ \psi_6 = -\dfrac{1}{2} \end{cases}$

On regroupe les termes constants en θ :

$$-\phi(\theta) = -\frac{1}{2}\left[\frac{\bar{t_0}^2}{\sigma_{t_0}^2} + \frac{\bar{v_0}^2}{\sigma_{v_0}^2} + \frac{\bar{t_0}^2}{s_{t_0}^2} + \frac{\bar{v_0}^2}{s_{v_0}^2} + \frac{v_\xi^2}{\sigma_\xi^2} + \frac{v_\tau^2}{\sigma_\tau^2} + \frac{v^2}{\sigma^2}\right] + \frac{\bar{t_0}\bar{\bar{t_0}}}{s_{t_0}^2} + \frac{\bar{v_0}\bar{\bar{v_0}}}{s_{v_0}^2}$$
$$- (N + m_\xi + 2)\log\sigma_\xi - (N + m_\tau + 2)\log\sigma_\tau - (kN + m + 2)\log\sigma$$

On pose $S(y,z) = \begin{pmatrix} S_1 \\ \vdots \\ S_6 \end{pmatrix}$ et $\psi(\theta) = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_6 \end{pmatrix}$. On vérifie bien que :

$$\log p(y,z,\theta) = -\phi(\theta) + \langle S(y,z) | \psi(\theta) \rangle + \text{cst}$$

2. Sur le notebook.

3. Un choix possible est d'utiliser le Symetric Random Walk Metropolis-Hastings. On choisit une loi de proposition q gaussienne centrée en z et de variance σ_{prop}^2 [1]. Le cours nous apprend deux choses :

- Cette procédure est très dépendante de l'initialisation z_0.
- Il faut que le support de la loi cible π soit compacte (ce qui est le cas ici).

La loi cible est $p(z|y,\theta) = \frac{p(y,z,\theta)}{p(y,\theta)}$. On ne connaît pas $p(y,\theta)$ mais ce terme va se simplifier dans le ratio d'acceptation, ce qui nous permet d'utiliser simplement la vraisemblance calculée à la question 1. comme loi cible.

4. Algorithme MCMC-SAEM :

On ne sait pas résoudre explicitement l'étape E de l'EM. On va donc utiliser une méthode de Monte-Carlo pour approximer l'espérance. On va utiliser une méthode de Metropolis-Hastings pour simuler $z^k \sim p(\cdot|y,\theta)$.

1. S Simuler $z^k \sim p(\cdot|y,\theta)$
2. A $Q_{k+1} = Q_k + \varepsilon_k(\log(y,z^k) - Q_k) = (1-\varepsilon_k)Q_k + \varepsilon_k \log(y,z^k)$
3. M $\theta_{k+1} = \arg\max Q_{k+1}(\theta)$

1. Pour une proposition gaussienne le meilleur $\sigma_{\text{prop}} = 0.234$.

Si p est un modèle exponentiel on peut ré-écrirer l'étape (A) :

$$S_{k+1} = S_k + \varepsilon_k (S(y, z^{k+1}) - S_k)$$

Comme $\log p(y, z, \theta)$ est concave, on veut annuler $\nabla_\theta(-\phi(\theta) + \langle S(y,z) | \psi(\theta) \rangle)$. On dérive successivement par rapport à chaque coordonnée de θ.

$$\begin{aligned}
&\nabla_{\bar{t_0}} \left(-\phi(\theta) + \langle S(y,z) | \psi(\theta) \rangle\right) = 0 \\
&\Leftrightarrow -\nabla_{\bar{t_0}}\phi(\theta) + S_1 \nabla_{\bar{t_0}} \psi_1(\theta) = 0 \\
&\Leftrightarrow -\frac{\bar{t_0}}{\sigma_{t_0}^2} - \frac{\bar{t_0}}{s_{t_0}^2} + \frac{\bar{\bar{t_0}}}{s_{t_0}^2} + \frac{S_1}{\sigma_{t_0}^2} = 0 \\
&\Leftrightarrow \boxed{\bar{t_0} = \frac{\sigma_{t_0}^2 \bar{\bar{t_0}} + s_{t_0}^2 S_1}{\sigma_{t_0}^2 + s_{t_0}^2}}
\end{aligned}$$

$$\begin{aligned}
&\nabla_{\bar{v_0}} \left(-\phi(\theta) + \langle S(y,z) | \psi(\theta) \rangle\right) = 0 \\
&\Leftrightarrow -\nabla_{\bar{v_0}}\phi(\theta) + S_2 \nabla_{\bar{v_0}} \psi_2(\theta) = 0 \\
&\Leftrightarrow -\frac{\bar{v_0}}{\sigma_{v_0}^2} - \frac{\bar{v_0}}{s_{v_0}^2} + \frac{\bar{\bar{v_0}}}{s_{v_0}^2} + \frac{S_2}{\sigma_{v_0}^2} = 0 \\
&\Leftrightarrow \boxed{\bar{v_0} = \frac{\sigma_{v_0}^2 \bar{\bar{v_0}} + s_{v_0}^2 S_2}{\sigma_{v_0}^2 + s_{v_0}^2}}
\end{aligned}$$

$$\begin{aligned}
&\nabla_{\sigma_\xi} \left(-\phi(\theta) + \langle S(y,z) | \psi(\theta) \rangle\right) = 0 \\
&\Leftrightarrow -\nabla_{\sigma_\xi}\phi(\theta) + S_3 \nabla_{\sigma_\xi} \psi_3(\theta) = 0 \\
&\Leftrightarrow \frac{v_\xi^2}{\sigma_\xi^3} - \frac{N + m_\xi + 2}{\sigma_\xi} + \frac{S_3 N}{\sigma_\xi^3} = 0 \\
&\Leftrightarrow \boxed{\sigma_\xi^2 = \frac{v_\xi^2 + N S_3}{N + m_\xi + 2}}
\end{aligned}$$

$$\begin{aligned}
&\nabla_{\sigma_\tau} \left(-\phi(\theta) + \langle S(y,z) | \psi(\theta) \rangle\right) = 0 \\
&\Leftrightarrow -\nabla_{\sigma_\tau}\phi(\theta) + S_4 \nabla_{\sigma_\tau} \psi_4(\theta) = 0 \\
&\Leftrightarrow \frac{v_\tau^2}{\sigma_\tau^3} - \frac{N + m_\tau + 2}{\sigma_\tau} + \frac{S_4 N}{\sigma_\tau^3} = 0 \\
&\Leftrightarrow \boxed{\sigma_\tau^2 = \frac{v_\tau^2 + N S_4}{N + m_\tau + 2}}
\end{aligned}$$

$$\begin{aligned}
&\nabla_{\sigma} \left(-\phi(\theta) + \langle S(y,z) | \psi(\theta) \rangle\right) = 0 \\
&\Leftrightarrow -\nabla_{\sigma}\phi(\theta) + S_5 \nabla_{\sigma} \psi_5(\theta) = 0 \\
&\Leftrightarrow \frac{v^2}{\sigma^3} - \frac{kN + m + 2}{\sigma} + \frac{S_5 N}{\sigma^3} = 0 \\
&\Leftrightarrow \boxed{\sigma^2 = \frac{v^2 + N S_5}{kN + m + 2}}
\end{aligned}$$

5. Pour i une dimension quelconque de z, on a :

$$p(z_i | y, z_{(-i)}, \theta) = \frac{q(y, z, \theta)}{q(y, z_{(-i)}, \theta)}$$

Si on note $z^{(k,i)} = (z_1^{(k+1)}, \dots, z_i^{(k+1)}, z_{i+1}^{(k)}, \dots, z_{2N+1}^{(k)})$ la valeur de z à l'itération k où les i premières dimensions ont déjà été mises à jour. On veut calculer le ratio d'acceptation de la proposition $z_i^{(k+1)}$

depuis $z_i^{(k)}$ pour un RW-MH. On a :

$$\begin{aligned}
\frac{\pi_i(z_i^{(k+1)})}{\pi_i(z_i^{(k)})} &= \frac{p(z_i^{(k+1)}|y, z_{(-i)}^{(k,i)}, \theta)}{p(z_i^{(k)}|y, z_{(-i)}^{(k,i)}, \theta)} \\
&= \frac{q(y, z^{(k,i)}, \theta)}{q(y, z^{(k,i-1)}, \theta)} && \text{(en utilisant la formule du dessus).}
\end{aligned}$$

On peut donc simplement utiliser le ratio des vraisemblances calculées à la question 1. pour calculer le ratio d'acceptation. Ce n'est pas optimal dans le sens où on calcule des termes qui vont s'annuler mutuellement, mais ça simplifie l'implémentation.

6. La question 5 nous fournit le résultat souhaité.

7. Sur le notebook.

8. En utilisant un Gibbs par blocs, nous pouvons réduire de moitié le nombre d'itérations nécessaires pour mettre à jour toutes les dimensions de z. L'avantage est donc un gain de temps de calcul. On fait des tirages en dimension seulement 2 donc on n'a pas pour autant les problèmes du Metropolis-Hastings.

9. Les résultats sont très similaires à ceux obtenus avec l'échantillonneur de Gibbs, avec un temps de calcul deux fois plus faible. Dans les deux cas on a du mal à converger pour tous les paramètres. Un problème peut être qu'on utilise toujours la même variance pour les lois de propositions, alors que les distributions des paramètres sont différentes.

Exercice 2 Multiplicative Hasting-Metropolis

1. On considère la transition de x à y. On a deux cas :

- On est passé par une multiplication avec probabilité $\frac{1}{2}$. Dans ce cas $\varepsilon = \frac{y}{x}$.
- On est passé par une division avec probabilité $\frac{1}{2}$. Dans ce cas $\varepsilon = \frac{x}{y}$.

On a donc :

$$q(x, y) = \frac{1}{2} f\left(\frac{y}{x}\right) + \frac{1}{2}\left(\frac{x}{y}\right)$$

2.

$$\begin{aligned}
\alpha(x, y) &= \min\left(1, \frac{\pi(y) q(y, x)}{\pi(x) q(x, y)}\right) \\
&= \min\left(1, \frac{\pi(y)\left[\frac{1}{2}\left(\frac{x}{y}\right) + \frac{1}{2} f\left(\frac{y}{x}\right)\right]}{\pi(x)\left[\frac{1}{2} f\left(\frac{y}{x}\right) + \frac{1}{2}\left(\frac{x}{y}\right)\right]}\right) \\
&= \min\left(1, \frac{\pi(y)}{\pi(x)}\right)
\end{aligned}$$

Le noyau de proposition a le bon goût d'être symétrique, ce qui nous permet de simplifier le ratio d'acceptation.

3. et 4. Sur le notebook.