TP 3 : Hasting - Metropolis (and Gibbs) samplers

Exercise 1.

1.A. A population model for longitudinal data

$\boxed{1}$ On veut exprimer la log-vraisemblance $\log q(y, z, \theta)$ du modèle et montrer ensuite que le modèle appartient à la famille des curved exponentielle.

Tout d'abord on a:

$$\log(q(y, z, \theta)) = \log(q(y|z, \theta)\, q(z|\theta)\, q(\theta))$$

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

On rappelle que le vecteur paramètres est donnée par:

$$\theta = (\overline{t_0}, \overline{v_0}, \sigma_\xi^2, \sigma_\tau^2, \sigma^2)$$

Afin de simplifier l'expression de $\log q(y, z, \theta)$, nous enlèverons les termes constants qui n'interviennent pas dans l'optimisation de $\log q(y, z, \theta)$.

Les termes qui composent $\log q(y, z, \theta)$ sont donnée par:

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

(En effet on rappelle que $y_{ij} = d_i(t_{ij}) + \varepsilon_{ij}$ où $\varepsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)$)

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

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

$$\log q(z|\theta) = -\frac{n}{2}\log(\sigma_\xi^2) - \frac{1}{2\sigma_\xi^2}\sum_i \|\xi_i\|^2$$
$$-\frac{n}{2}\log(\sigma_\tau^2) - \frac{1}{2\sigma_\tau^2}\sum_i \|\tau_i\|^2$$
$$-\frac{1}{2}\log(\sigma_{t_0}^2) - \frac{1}{2\sigma_{t_0}^2}\|t_0 - \overline{t_0}\|^2$$
$$-\frac{1}{2}\log(\sigma_{v_0}^2) - \frac{1}{2\sigma_{v_0}^2}\|v_0 - \overline{v_0}\|^2$$

On rappelle que d'après l'énoncé on a :

$$\overline{t_0} \rightsquigarrow \mathcal{N}(\overline{\overline{t_0}}, s_{t_0}^2)$$
$$\overline{v_0} \rightsquigarrow \mathcal{N}(\overline{\overline{v_0}}, s_{v_0}^2)$$
$$\sigma_\xi^2 \rightsquigarrow W^{-1}(v_\xi, m_\xi)$$
$$\sigma_\tau^2 \rightsquigarrow W^{-1}(v_\tau, m_\tau)$$
$$\sigma^2 \rightsquigarrow W^{-1}(v, m)$$

On a donc :

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

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

Maintenant, on veut montrer que le modèle appartient à la famille exponentielle. Un modèle de la famille exponentielle s'écrit sous la forme :

$$\log p(y,z,\theta) = \langle S(y,z), \varphi(\theta)\rangle + \psi(\theta).$$

Ici $\theta = (\overline{t_0}, \overline{v_0}, \sigma_\xi^2, \sigma_\tau^2, \sigma^2)$

Si l'on pose :

Si on pose :

$$S_1 = \frac{1}{N} \sum_{i,j} \| y_{i,j} - d_i(t_{i,j}) \|^2 \qquad \varphi_1 = -\frac{N}{2\sigma^2}$$

$$S_2 = \frac{1}{m} \sum_i \| \tau_i \|^2 \qquad \varphi_2 = -\frac{m}{2\sigma_\tau^2}$$

$$S_3 = \frac{1}{m} \sum_i \| \xi_i \|^2 \qquad \varphi_3 = -\frac{m}{2\sigma_\xi^2}$$

$$S_4 = t_0 \qquad \varphi_4 = \frac{1}{\sigma_{t_0}^2} \overline{t_0}$$

$$S_5 = v_0 \qquad \varphi_5 = \frac{1}{\sigma_{v_0}^2} \overline{v_0}$$

$$S(y,z) = (S_1, S_2, S_3, S_4, S_5) \qquad \varphi(\theta) = (\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5)$$

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

On a bien que

$$\log q(y, z, \theta) = \langle S(y,z), \varphi(\theta) \rangle + \Psi(\theta) .$$

On en conclut que le modèle appartient à la famille exponentielle.

1.B. HM-SAEM – Hasting-Metropolis sampler.

3 Les variables latente sont données par $z = (t_0, v_0, \xi_i, \tau_i)_{1 \leq i \leq N}$.
Dans l'algorithme SAEM, pour l'échantillonage de la variable z^{t+1} (au temps $t+1$ de SAEM), nous allons utiliser le Hasting-Metropolis sampler, nous allons en effet effectuer quelques sauts de Hasting-Metropolis à partir de la dernière valeur z^t.

La distribution de proposition considérée pour l'algorithme est une Gaussienne multivariée $\mathcal{N}(z, \sigma_{prop})$, σ_{prop} étant à déterminer.

↳ Ceci implique que le ratio d'acceptation est donnée par

$$\alpha(z^{(k)}, z^*) = \min\left(1, \frac{\pi(z^*)}{\pi(z^{(k)})}\right)$$

Or $\forall t$, $$\frac{\pi(z^{t+1})}{\pi(z^t)} = \frac{q(z^{t+1} \mid y, \theta)}{q(z^t \mid y, \theta)} = \frac{q(z^{t+1}, y, \theta)}{q(z^t, y, \theta)}$$

càd $\forall t$, $$\alpha(z^t, z^{t+1}) = \frac{q(z^{t+1}, y, \theta)}{q(z^t, y, \theta)}$$

$\Rightarrow \forall t,\ \log(\alpha(z^t, z^{t+1})) = \log q(z^{t+1}, y, \theta) - \log q(z^t, y, \theta)$

②

4 Nous allons calculer le paramètre optimal $\theta^{(k)}$ donnée par :

$$\theta^{(k)} = \underset{\theta \in \Theta}{\operatorname{argmax}} \left\{ \langle S(y, z_k), \varphi(\theta) \rangle + \psi(\theta) \right\}$$

$$= \underset{\theta \in \Theta}{\operatorname{argmax}} \log q(y, z_k, \theta)$$

$$\theta^{(k)} = \underset{\theta \in \Theta}{\operatorname{argmax}} \; q_{s^k}(\theta) .$$

On connaît l'expression de $q_{s^k}(\theta)$, qui a été déterminé à la question 1.
Afin de trouver $\theta^{(k)}$, nous allons calculer les paramètres $\bar{t}_0, \bar{v}_0, \sigma^2, \sigma_\xi^2, \sigma_\tau^2$ qui annulent les dérivées partielles de $q_{s^k}(\theta)$, comme suit :

$$\nabla_{\bar{t}_0} q_{s^k}(\theta) = \frac{1}{\sigma_{t_0}^2}(t_0 - \bar{t}_0) - \frac{1}{s_{t_0}^2}(\bar{t}_0 - \bar{\bar{t}}_0) = 0$$

$$\Longleftrightarrow \frac{t_0}{\sigma_{t_0}^2} + \frac{\bar{\bar{t}}_0}{s_{t_0}^2} = \left(\frac{1}{\sigma_{t_0}^2} + \frac{1}{s_{t_0}^2}\right) \bar{t}_0$$

$$\Longleftrightarrow \boxed{\bar{t}_0 = \frac{\frac{t_0}{\sigma_{t_0}^2} + \frac{\bar{\bar{t}}_0}{s_{t_0}^2}}{\frac{1}{\sigma_{t_0}^2} + \frac{1}{s_{t_0}^2}}}$$

$$\nabla_{\bar{v}_0} q_{s^k}(\theta) = \frac{v_0 - \bar{v}_0}{\sigma_{v_0}^2} - \frac{\bar{v}_0 - \bar{\bar{v}}_0}{s_{v_0}^2} = 0$$

$$\Longleftrightarrow \boxed{\bar{v}_0 = \frac{\frac{v_0}{\sigma_{v_0}^2} + \frac{\bar{\bar{v}}_0}{s_{v_0}^2}}{\frac{1}{\sigma_{v_0}^2} + \frac{1}{s_{v_0}^2}}}$$

$$\nabla_{\sigma^2} q_{s^k}(\theta) = -\frac{N}{2\sigma^2} + \frac{S_1 N}{2\sigma^4} - \left(1 + \frac{m}{2}\right) \times \frac{1}{\sigma^2} + \frac{v^2}{2\sigma^4} = 0$$

$$\Longleftrightarrow \boxed{\sigma^2 = \frac{S_1 N + v^2}{N + 2 + m}}$$

$$\nabla_{\sigma_\xi^2} q_{s^k}(\theta) = -\frac{n}{2\sigma_\xi^2} + \frac{n S_2}{2\sigma_\xi^4} - \left(1 + \frac{m_\xi}{2}\right) \times \frac{1}{\sigma_\xi^2} + \frac{v_\xi^2}{2\sigma_\xi^2} = 0$$

$$\Longleftrightarrow \boxed{\sigma_\xi^2 = \frac{n S_3 + v_\xi^2}{n + 2 + m_\xi}}$$

$$\nabla_{\sigma_\tau^2} q_{S^k}(\theta) = 0 \iff \boxed{\sigma_\tau^2 = \frac{n S_2 + v_\tau^2}{n + 2 + m_\tau}}$$

Ainsi, nous avons trouvé les paramètres optimaux $\theta^{(k)}$

1.C HMwG - SAEM - Hastings-Metropolis within Gibbs sampler

$\boxed{5}$ HMwG : On connaît la densité postérieure $q(z_{pop}, z_1, \dots, z_n \mid y, \theta)$

Pour chaque i, on fait un saut de Metropolis-Hastings sur la loi $q(z_i \mid y, z_{pop}, z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n, \theta)$

Le ratio d'acceptation est donnée par :

$$\alpha_i = \frac{q(z_i^* \mid z_{pop}, z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n, y, \theta)}{q(z_i \mid z_{pop}, z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n, y, \theta)}$$

$$\alpha_i = \frac{q(z^*, y, \theta)}{q(z, y, \theta)}$$

$\boxed{6}$ Maintenant, on souhaite échantillonner sur la loi $q(z_{pop} \mid y, z_1, \dots, z_n, \theta)$
Pour cela on effectue un saut de H-M sur cette loi. et on calcule le ratio d'acceptation, donnée par :

$$\alpha_{pop} = \frac{q(z_{pop}^* \mid y, z_1, \dots, z_n, \theta)}{q(z_{pop} \mid y, z_1, \dots, z_n, \theta)}$$

$$\alpha_{pop} = \frac{q(z_{pop}^*, y, z_1, \dots, z_n, \theta)}{q(z_{pop}, y, z_1, \dots, z_n, \theta)}$$

8

Un block Gibbs sampler permet logiquement d'échantillonner plus rapidement qu'un "one-at-a-time" Gibbs sampler.