TP 3 _ Computational statistics

Exercice 1

1 _ $\log p(\{y_{ij}\}_{ij}, \{z_i\}_i, z_{pop}, \theta)$

$$= \underbrace{\log p(\{y_{ij}\}_{ij} \mid \{z_i\}, z_{pop}, \theta)}_{(i)} + \underbrace{\log p(\{z_i\}_i, z_{pop} \mid \theta)}_{(ii)} + \underbrace{\log p(\theta)}_{(iii)}$$

$$(i) = \sum_{i=1}^{N} \sum_{j=1}^{k_i} -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{(y_{ij} - d_i(t_{ij}))^2}{2\sigma^2}$$

$$= \underbrace{\sum_{i=1}^{N} -\frac{k_i}{2}\log(2\pi)}_{c^{ste}} + \sum_{i=1}^{N} \sum_{j=1}^{k_i} -\frac{1}{2}\log(\sigma^2) - \frac{(y_{ij} - d_i(t_{ij}))^2}{2\sigma^2}$$

with $d_i(t_{ij}) = p_0 + \alpha_i v_0 (t_{ij} - t_0 - \tau_i)$

$$(ii) = \log p(\{z_i\}, z_{pop} \mid \theta) = \log p(\{z_i\} \mid \theta) + \log p(z_{pop} \mid \theta)$$

$$= \sum_{i=1}^{N} \left(-\log(\alpha_i) - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma_\xi^2) - \frac{\ln \alpha_i^2}{2\sigma_\xi^2} - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma_\tau^2 - \frac{\tau_i^2}{2\sigma_\tau^2} \right) - \log(2\pi) - \frac{1}{2}\left[\log\sigma_{t_0}^2 + \log\sigma_{v_0}^2\right]$$

$$- \frac{(t_0 - \bar{t}_0)^2}{2\sigma_{t_0}^2} - \frac{(v_0 - \bar{v}_0)^2}{2\sigma_{v_0}^2}$$

$$= \underbrace{\sum_{i=1}^{N} \cdot \log(2\pi) - \log(2\pi)}_{cste} + \sum_{i=1}^{N}\Big[-\log(d_i) - \frac{1}{2}\log(\sigma_g^2) - \frac{\log(d_i)^2}{2\sigma_g^2}$$

$$-\frac{1}{2}\log\sigma_\tau^2 - \frac{\tau_i^2}{2\sigma_\tau^2}\Big] - \frac{1}{2}\log\sigma_{v_0}^2 - \frac{1}{2}\log\sigma_{t_0}^2 - \frac{(t_0-\bar{t}_0)^2}{2\sigma_{t_0}^2} - \frac{(v_0-\bar{v}_0)^2}{2\sigma_{v_0}^2}$$

(iii) $= \log p(\theta) = \log p(\bar{t}_0, \bar{v}_0, \sigma_g, \sigma_\tau, \theta)$

$$= \underbrace{-\log(2\pi) - \frac{1}{2}\log(s_{t_0}^2) - \frac{1}{2}\log(s_{v_0}^2)}_{cste} - \frac{(\bar{t}_0-\bar{\bar{t}}_0)^2}{2s_{t_0}^2} - \frac{(\bar{v}_0-\bar{\bar{v}}_0)^2}{2s_{v_0}^2}$$

$$+ \sum_{x\in\{g,\tau,\phi\}} \Big[\underbrace{-\log\big(\Gamma(\tfrac{m_x}{2})\big)}_{} - \log(\sigma_x^2) + \underbrace{m_x\log(v_x)}_{} - \frac{m_x}{2}\log(\sigma_x^2)$$

$$\underbrace{-\frac{1}{2}m_x\log(2)}_{cste} - \frac{v_x^2}{2\sigma_x^2}\Big]$$

By summing (i), (ii) and (iii), we obtain the log-likelihood.

Let's write it as : $\log p(y,z,\theta) = -\phi(\theta) + \langle S(y,z) | \psi(\theta)\rangle$

$\theta = (\bar{t}_0, \bar{v}_0, \sigma_g, \sigma_\tau, \theta)$

- $$\phi(\theta) = \frac{\bar{t}_0^2}{2\sigma_{t_0}^2} + \frac{\bar{\bar{t}}_0^2}{2s_{t_0}^2} - \bar{t}_0\frac{\bar{\bar{t}}_0}{s_{t_0}^2} + \left(\frac{N}{2} + \frac{m_g}{2} + 1\right)\log(\sigma_g^2)$$

$$+ \frac{\bar{v}_0^2}{2s_{v_0}^2} + \frac{\bar{v}_0^2}{2\sigma_{v_0}^2} - \bar{v}_0\frac{\bar{\bar{v}}_0}{s_{v_0}^2} + \left(1 + \frac{m_\tau}{2} + \frac{N}{2}\right)\log(\sigma_\tau^2)$$

$$+ \left(1 + \frac{m}{2} + \frac{1}{2}\sum_{i=1}^{N} k_i\right)\log\sigma^2 - \frac{v_g^2}{2\sigma_g^2} - \frac{v_\tau^2}{2\sigma_\tau^2} - \frac{v^2}{2\sigma^2}$$

- $$S(y,z) = \begin{bmatrix} \frac{1}{\sum k_i} + \sum_{i=1}^{N}\sum_{j=1}^{k_i}(y_{ij} - d_i(t_{ij}))^2 \\ \frac{1}{N}\sum_{i=1}^{N}\log(d_i)^2 \\ \frac{1}{N}\sum_{i=1}^{N}\tau_i^2 \\ t_0 \\ v_0 \end{bmatrix}$$

$$\psi(\theta) = \begin{pmatrix} -\left(\sum_{i=1}^{N} h_i\right)/2\sigma^2 \\ -N/(2\sigma_\xi^2) \\ -N/(2\sigma_\tau^2) \\ \bar{t}_0/\sigma_{t_0}^2 \\ \bar{v}_0/\sigma_{v_0}^2 \end{pmatrix}$$

3 _ To use the Metropolis-Hastings algorithm, we need to compute, up to some constant independent of $z = (\{z_i\}, z_{pop})$, the target distribution

$\log(\pi) = \cancel{\log p(\{y_{ij}\}, \{z_i\}, z_{pop}, \theta)}\ \log p(z|y,\theta)$

We have: $\log p(z|y,\theta) = \log p(z,y,\theta) - \log p(y,\theta)$

$$= \log p(z,y,\theta) + C$$

$$= \log p(y|z,\theta) + \log p(z|\theta) + C + \log p(\theta)$$

$$\log(\pi) = \log p(y|z,\theta) + \log p(z|\theta) + C'$$

Hence: $\log(\pi) \overset{cst}{=} -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{k_i}\frac{(y_{ij} - d_i(t_{ij}))^2}{\sigma^2} - \frac{1}{2}\sum_{i=1}^{N}\frac{\log(\alpha_i)^2}{\sigma_\xi^2}$

$$-\frac{1}{2}\sum_{i=1}^{N}\frac{\tau_i^2}{\sigma_\tau^2} - \frac{(t_0-\bar{t}_0)^2}{2\sigma_{t_0}^2} - \frac{(v_0-\bar{v}_0)^2}{2\sigma_{v_0}^2}$$

$$-\sum_{i=1}^{N}\log(\alpha_i)$$

4 - The function $g : \theta \longmapsto -\phi(\theta) + \langle s^{(k)} | \psi(\theta) \rangle$ is concave and differentiable, and we have no constraints explicit for θ.

Hence: $\theta^* = \underset{\theta \in \Theta}{\arg\max} \; -\phi(\theta) + \langle s^{(k)} | \psi(\theta) \rangle$

We have: $\nabla g(\theta^*) = 0$ ($\Leftarrow$ holds)

As suggested, we will differentiate with respect to: $\theta' = (t_0, v_0, \sigma_\xi^2, \sigma_\tau^2, \sigma^2)$.

We have:

$$\nabla_{\theta'}(\phi(\theta')) = \begin{bmatrix} \frac{\bar{t_0}}{\sigma_{t_0}^2} + \frac{\bar{t_0}}{s_{t_0}^2} - \frac{\bar{t_0}}{s_{t_0}^2} \\ \frac{\bar{v_0}}{\sigma_{v_0}^2} + \frac{\bar{v_0}}{s_{v_0}^2} - \frac{\bar{v_0}}{s_{v_0}^2} \\ \left(\frac{N}{2} + \frac{m_\xi}{2} + 1\right)\frac{1}{\sigma_\xi^2} + \frac{v_\xi^2}{2\sigma_\xi^4} \\ \left(\frac{N}{2} + \frac{m_\tau}{2} + 1\right)\frac{1}{\sigma_\tau^2} + \frac{v_\tau^2}{2\sigma_\tau^4} \\ \left(\frac{N}{2} + \frac{m}{2} + 1\right)\frac{1}{\sigma^2} + \frac{\theta}{2\sigma^4} \end{bmatrix}$$

$$\nabla_{\theta'} \langle s^{(k)} | \psi(\theta) \rangle = \begin{bmatrix} s_1^{(k)} \frac{1}{\sigma_{t_0}^2} \\ s_2^{(k)} \frac{1}{\sigma_{v_0}^2} \\ s_3^{(k)} \frac{N}{2\sigma_\xi^4} \\ s_4^{(k)} \frac{N}{2\sigma_\tau^4} \\ s_5^{(k)} \frac{\sum_{i=1} k_i}{2\sigma^4} \end{bmatrix}$$

We then compute θ^* : $\nabla g(\theta^*) = 0 = -\nabla\psi(\theta^*) + \nabla\langle S^{(k)}, \psi(\theta^*)\rangle$

We obtain:

$$\cancel{\overline{\omega_0} = \left(\frac{1}{\sigma_{\omega_0}^2} + \frac{1}{s_{\omega_0}^2}\right)^{-1}\left(\frac{\bar{\bar{\omega}}_0}{s_{\omega_0}^2} + \frac{S_1^{(k)}}{\sigma_{\omega_0}^2}\right)}$$

$$\cancel{\overline{t_0} = \left(\frac{1}{\sigma_{t_0}^2} + \frac{1}{s_{t_0}^2}\right)^{-1}\left(\frac{\bar{\bar{t}}_0}{s_{t_0}^2} + S_1\right)}$$

$$\begin{cases}
\overline{t_0} = \left(\dfrac{1}{\sigma_{t_0}^2} + \dfrac{1}{s_{t_0}^2}\right)^{-1}\left(\dfrac{\bar{\bar{t}}_0}{s_{t_0}^2} + \dfrac{S_1^{(k)}}{\sigma_{t_0}^2}\right) \\[2ex]
\overline{\omega_0} = \left(\dfrac{1}{\sigma_{\omega_0}^2} + \dfrac{1}{s_{\omega_0}^2}\right)^{-1}\left(\dfrac{\bar{\bar{\omega}}_0}{s_{\omega_0}^2} + \dfrac{S_2^{(k)}}{\sigma_{\omega_0}^2}\right) \\[2ex]
\sigma_g^2 = \dfrac{N S_3^{(k)} - v_g^2}{2 + m_g + N} \\[2ex]
\sigma_\tau^2 = \dfrac{N S_4^{(k)} - v_\tau^2}{2 + m_\tau + N} \\[2ex]
\sigma^2 = \dfrac{S_5^{(k)} \sum_{i=1}^{N} k_i - v^2}{2 + m + \sum_{i=1}^{N} k_i}
\end{cases}$$

Exercice 2

Y is : $Y = \frac{X}{\varepsilon} \mathbb{1}(B=0) + \varepsilon X \, \mathbb{1}(B=1)$

Let h be a measurable bounded function :

$$\mathbb{E}[h(Y)] = \frac{1}{2} \mathbb{E}[h(\frac{X}{\varepsilon})] + \frac{1}{2} \mathbb{E}[h(\varepsilon X)]$$

$$= \frac{1}{2}\int_{[-1,1]} h(\frac{X}{e}) f(e) de$$

$$+ \frac{1}{2}\int_{[-1,1]} h(Xe) f(e) de$$

$$= \frac{1}{2}\int_{]-\infty,-|X|[\cup]|X|,+\infty[} h(y) \frac{|X|}{y^2} f(\frac{X}{y}) dy + \frac{1}{2}\int_{-|X|}^{|X|} h(y) f(\frac{y}{X}) \frac{dy}{|X|}$$

We have use the change of variable via f bijective :

$$f : \left| \begin{array}{l}]-\infty,-|X|[\cup]|X|,+\infty[\to]-1,1[\setminus \{0\} \\ \varepsilon \to \frac{X}{\varepsilon} \end{array} \right.$$

with : $|\varphi'(\varepsilon)| = \frac{|X|}{\varepsilon^2} = \frac{y^2}{|X|}$

Therefore, we can conclude that the density of $Y|X$ is :

$$y \mapsto \frac{|X|}{2y^2} f(\frac{X}{y}) \mathbb{1}_{|y|>|X|}(y) + \frac{f(\frac{y}{X})}{2|X|} \mathbb{1}_{|y|<|X|}$$

Hence:

$$q(x,y) = \frac{|x|}{2y^2} f\left(\frac{x}{y}\right) \mathbb{1}_{|y|>x}(y) + \frac{1}{2|x|} f\left(\frac{y}{x}\right) \mathbb{1}_{|y|<|x|}(y)$$

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

We obtain the result with:

$$\frac{q(y,x)\pi(y)}{q(x,y)\pi(x)} = \frac{\pi(y)}{\pi(x)} \frac{|y|}{|x|}$$