Dorian Gaillard

Homework 3:

Exercise 1)

1) $q(y, z, \theta) = p(y|z, \theta)\, p(z|\theta)\, p(\theta)$

$$= p(\theta)\, p(z_{pop}|\theta) \prod_{i=1}^{N} \left(\prod_{j=1}^{k_i} p(y_{ij}|z_i, \theta) \right) p(z_i|\theta)$$

by independance.

Then $p(\theta) = p(\theta_{pop}) \times p(\theta_i)$

$$= \phi(\overline{t_0}; \overline{\overline{t_0}}, \Delta_{t_0}^2)\, \phi(\overline{v_0}; \overline{\overline{v_0}}, \Delta_{v_0}^2)$$

$$\times f_{W_\xi^{-1}}(\sigma_\xi^2)\, f_{W_\tau^{-1}}(\sigma_\tau^2)\, f_{W^{-1}}(\sigma^2)$$

$$p(z_{pop}|\theta) = \phi(t_0; \overline{t_0}, \sigma_{t_0}^2)\, \phi(v_0; \overline{v_0}, \sigma_{v_0}^2)$$

$$p(z_i|\theta) = \phi(\log(\alpha_i); 0, \sigma_\xi^2)\, \phi(\tau_i; 0, \sigma_\tau^2)$$

and $p(y_{ij}|z_i, \theta) = \phi(y_{ij} - d_i(t_{ij}); 0, \sigma^2)$

where $\phi(\cdot; \mu, \sigma^2)$ is the density of gaussian distribution with mean μ and variance σ^2.

$$\log q(y, z|\theta) = \log\left(p(z_{pop}|\theta) \prod_{i=1}^{N} \left(\prod_{j=1}^{k_i} p(y_{ij}|z_i, \theta) \right) p(z_i|\theta) \right)$$

$$= C(y, z) - \frac{\overline{t_0}^2 - 2 t_0 \overline{t_0}}{2\sigma_{t_0}^2} - \frac{\overline{v_0}^2 - 2 v_0 \overline{v_0}}{2\sigma_{v_0}^2}$$

$$- \sum_{i=1}^{N} \left[\frac{\log(\alpha_i)^2}{2\sigma_\xi^2} + \frac{\tau_i^2}{2\sigma_\tau^2} + \sum_{j=1}^{k_i} \frac{(y_{ij} - d_i(t_{ij}))^2}{2\sigma^2} \right]$$

$$- \frac{N}{2} \left[\log(\sigma_\xi^2) + \log(\sigma_\tau^2) \right] - \left(\frac{\sum_{i=1}^{N} k_i}{2} \right) \log(\sigma^2)$$

which can be rewritten as:

$$\log q(y, z | \theta) = C(y, z) - \phi(\theta) + \langle S(y, z), \psi(\theta) \rangle$$

where $C(y, z)$ does not depend on θ,

$$S(y, z) = \begin{pmatrix} \sum_{i=1}^{N} \sum_{j=1}^{k_i} (y_{ij} - d_i(t_{ij}))^2 \\ \sum_{i=1}^{N} \log(\alpha_i)^2 \\ \sum_{i=1}^{N} \tau_i^2 \\ v_0 \\ t_0 \end{pmatrix} \text{ and } \psi(\theta) = \begin{pmatrix} -\frac{1}{2\sigma^2} \\ -\frac{1}{2\sigma_\xi^2} \\ -\frac{1}{2\sigma_\tau^2} \\ \frac{\overline{v_0}}{\sigma_{v_0}^2} \\ \frac{\overline{t_0}}{\sigma_{t_0}^2} \end{pmatrix}$$

$$\text{and } \phi(\theta) = \frac{\overline{t_0}^2}{2\sigma_{t_0}^2} + \frac{\overline{v_0}^2}{2\sigma_{v_0}^2} + \frac{N}{2}\log(\sigma_\xi^2) + \frac{N}{2}\log(\sigma_\tau^2) + \frac{1}{2}\left(\sum_{i=1}^{N} k_i\right)\log(\sigma^2)$$

3) We want to sample from $\Pi(z) = p(z | y, \theta)$, the problem lies in the fact that this distribution is too complicated to compute, rather

but

$$\frac{p(z^* | y, \theta)}{p(z^{(k)} | y, \theta)} = \frac{p(y | z^*, \theta)\, p(z^*, \theta)}{p(y, \theta)} \times \frac{p(y, \theta)}{p(y | z^{(k)}, \theta)\, p(z^{(k)}, \theta)}$$

$$= \frac{p(y | z^*, \theta)}{p(y | z^{(k)}, \theta)} \times \frac{p(z^* | \theta)\, p(\theta)}{p(z^{(k)} | \theta)\, p(\theta)}$$

$$= \frac{p(y | z^*, \theta)\, p(z^* | \theta)}{p(y | z^{(k)}, \theta)\, p(z^{(k)} | \theta)}$$

which is easy to compute.

4) Recalling that $\theta = (\overline{t_0}, \overline{v_0}, \sigma_\xi^2, \sigma_\tau^2, \sigma^2)$, $\nabla_\theta \log(q(y, z | \theta)) = 0$ is equivalent to:

$$\begin{cases} \frac{-\overline{t_0} + t_0}{\sigma_{t_0}^2} = 0 \\ \frac{-\overline{v_0} + v_0}{\sigma_{v_0}^2} = 0 \\ \sum_{i=1}^{N} \frac{\log(\alpha_i)^2}{2\sigma_\xi^4} - \frac{N}{2\sigma_\xi^2} = 0 \\ \sum_{i=1}^{N} \frac{\tau_i^2}{2\sigma_\tau^2} - \frac{N}{2\sigma_\tau^2} = 0 \end{cases}$$

and $\sum_{i=1}^{N}\sum_{j=1}^{k_i} \frac{(y_{ij} - d_i(t_{ij}))^2}{2\sigma^4} - \left(\sum_{i=1}^{N} k_i\right) \times \frac{1}{\sigma^2} = 0$

so $\overline{t_0}^{(k+1)} = t_0 = S_5$

$\overline{v_0}^{(k+1)} = v_0 = S_4$

$\sigma_\xi^{2\,(k+1)} = \frac{1}{N}\sum_{i=1}^{N} \log(\alpha_i)^2 = \frac{1}{N} S_2$

$\sigma_\tau^{2\,(k+1)} = \frac{1}{N}\sum_{i=1}^{N} \tau_i^2 = \frac{1}{N} S_3$

$\sigma^{2\,(k+1)} = \frac{1}{\sum_{i=1}^{N} k_i}\sum_{i=1}^{N}\sum_{j=1}^{k_i} (y_{ij} - d_i(t_{ij}))^2 = \frac{S_1}{\sum_{i=1}^{N} k_i}$

5) Given $z_i^{(k)} = (\xi_i^{(k)}, \tau_i^{(k)})$,

$\xi_i^{(k+1)} \sim p(\xi_i \mid \tau_i^{(k)}, z_{pop}, y, \theta)$

$\tau_i^{(k+1)} \sim p(\tau_i \mid \xi_i^{(k+1)}, z_{pop}, y, \theta)$

where $p(\xi_i \mid \tau_i^{(k)}, z_{pop}, y, \theta)$ can be sampled using Hasting - Metropolis method with

$$\frac{p(\xi_i^* \mid \tau_i, z_{pop}, y, \theta)}{p(\xi_i^{(k)} \mid \tau_i, z_{pop}, y, \theta)} = \frac{p(y \mid z_i^*, z_{pop}, \theta)\, p(z_i^* \mid z_{pop}, \theta)}{p(\tau_i, z_{pop}, y, \theta)} \times \frac{p(\tau_i, z_{pop}, y, \theta)}{p(y \mid z_i^{(k)}, z_{pop}, \theta)\, p(z_i \mid z_{pop}, \theta)}$$

(the densities on y_j, $j \neq i$ disappear in the ratio thanks to independance up and down) and because they do not depend on z_i

$$= \frac{p(y_i \mid z_i^*, z_{pop}, \theta)\, p(\xi_i^* \mid \theta)}{p(y_i \mid z_i^{(k)}, z_{pop}, \theta)\, p(\xi_i^{(k)} \mid \theta)}$$

which can be easily computed.

In a similar way,

$p(\tau_i \mid \xi_i^{(k+1)}, z_{pop}, y, \theta)$ can be sampled with the ratio:

$$\frac{p(\tau_i^* \mid \xi_i, z_{pop}, y, \theta)}{p(\tau_i^{(k)} \mid \xi_i, z_{pop}, y, \theta)} = \frac{p(y_i \mid z_i^*, z_{pop}, \theta)\, p(\tau_i^* \mid \theta)}{p(y_i \mid z_i^{(k)}, z_{pop}, \theta)\, p(\tau_i^{(k)} \mid \theta)}$$

6) The same way:
Given $z_{pop}^{(k)}$:

$$t_0^{(k+1)} \sim p(t_0 \mid v_0^{(k)}, z_i, y, \theta)$$

$$v_0^{(k+1)} \sim p(v_0 \mid t_0^{(k+1)}, z_i, y, \theta)$$

where the first can be sampled with the ratio

$$\frac{p(t_0^* \mid v_0, z_i, y, \theta)}{p(t_0^{(k)} \mid v_0, z_i, y, \theta)} = \frac{p(y \mid z_{pop}^*, z_i, \theta)\, p(t_0^* \mid \theta)}{p(y \mid z_{pop}^{(k)}, z_i, \theta)\, p(t_0^{(k)} \mid \theta)}$$

and the second with

$$\frac{p(v_0^* \mid t_0, z_i, y, \theta)}{p(v_0^{(k)} \mid t_0, z_i, y, \theta)} = \frac{p(y \mid z_{pop}^*, z_i, \theta)\, p(v_0^* \mid \theta)}{p(y \mid z_{pop}^{(k)}, z_i, \theta)\, p(v_0^{(k)} \mid \theta)}$$

8) A Block HMwG is less costly: sampling one d-multivariate gaussian is less costly than sampling d univariate and independant gaussians for example (and the same applies to whatever the proposal distribution chosen).

Exercise 7)

1) Let $t \in \mathbb{R}$.

$$P(Y \leq t) = P(\varepsilon X \leq t)\, P(B=1) + P\left(\frac{X}{\varepsilon} \leq t\right) P(B=0)$$

$$= \frac{1}{2}\left[P(\varepsilon X \leq t) + P\left(\frac{X}{\varepsilon} \leq t\right)\right]$$

If $X \geq 0$;

* if $t \geq 0$;

$$P(Y \leq t) = \frac{1}{2}\left[P\left(\varepsilon \leq \frac{t}{X}\right) + P\left(\frac{X}{\varepsilon} \leq t \,\middle|\, \varepsilon \geq 0\right) P(\varepsilon \geq 0)\right.$$

$$\left. + P\left(\frac{X}{\varepsilon} \leq t \,\middle|\, \varepsilon \leq 0\right) P(\varepsilon \leq 0)\right]$$

$$= \frac{1}{2}\left[P\left(\varepsilon \leq \frac{t}{X}\right) + P\left(\varepsilon \geq \frac{X}{t} \,\middle|\, \varepsilon \geq 0\right) P(\varepsilon \geq 0)\right.$$

$$\left. + P(\varepsilon \leq 0)\right]$$

- If $\frac{t}{X} \leq 1$;

$$P(Y \leq t) = \frac{1}{2}\left[P\left(\varepsilon \leq \frac{t}{X}\right) + 0 + P(\varepsilon \leq 0)\right]$$

and $f_Y(t) = \dfrac{d P(Y \leq t)}{dt} = \dfrac{1}{2X} f\left(\dfrac{t}{X}\right)$

- If $\frac{t}{X} \geq 1$;

$$P(Y \leq t) = \frac{1}{2}\left[1 + P\left(\varepsilon \geq \frac{X}{t} \,\middle|\, \varepsilon \geq 0\right) P(\varepsilon \geq 0)\right.$$

$$\left. + P(\varepsilon \leq 0)\right]$$

and $\dfrac{d P(Y \leq t)}{dt} = \dfrac{X}{2t^2} f\left(\dfrac{X}{t}\right)$

* If $t \leq 0$:

$$\mathbb{P}(Y \leq t) = \frac{1}{2}\left[\mathbb{P}\left(\varepsilon \leq \frac{t}{x}\right) + \underbrace{\mathbb{P}\left(\varepsilon \leq \frac{x}{t} \,\middle|\, \varepsilon \geq 0\right)\mathbb{P}(\varepsilon \geq 0)}_{=0}\right.$$

$$\left. + \mathbb{P}\left(\varepsilon \geq \frac{x}{t} \,\middle|\, \varepsilon \leq 0\right)\mathbb{P}(\varepsilon \leq 0)\right]$$

$$= \frac{1}{2}\left[\mathbb{P}\left(\varepsilon \leq \frac{t}{x}\right) + \mathbb{P}\left(\varepsilon \geq \frac{x}{t} \,\middle|\, \varepsilon \leq 0\right)\mathbb{P}(\varepsilon \leq 0)\right]$$

- If $\frac{t}{x} \leq -1$:

$$\mathbb{P}(Y \leq t) = \frac{1}{2}\left[0 + \mathbb{P}\left(\varepsilon \geq \frac{x}{t} \,\middle|\, \varepsilon \leq 0\right)\mathbb{P}(\varepsilon \leq 0)\right]$$

and $f_Y(t) = \dfrac{d\mathbb{P}(Y \leq t)}{dt} = \dfrac{x}{2t^2} f\left(\dfrac{x}{t}\right)$

- If $\frac{t}{x} \geq -1$:

$$\mathbb{P}(Y \leq t) = \frac{1}{2}\left[\mathbb{P}\left(\varepsilon \leq \frac{t}{x}\right) + \mathbb{P}(\varepsilon \leq 0)\right]$$

and $f_Y(t) = \dfrac{1}{2x} f\left(\dfrac{t}{x}\right)$

So if $x \geq 0$,

$$f_Y(t) = \begin{cases} \dfrac{1}{2x} f\left(\dfrac{t}{x}\right) & \text{if } \left|\dfrac{t}{x}\right| \leq 1 \\ \dfrac{x}{2t^2} f\left(\dfrac{x}{t}\right) & \text{if } \left|\dfrac{t}{x}\right| \geq 1 \end{cases}$$

Then if $x \leq 0$,

$$\mathbb{P}(Y \leq t) = \mathbb{P}(-Y \geq -t)$$

$$= 1 - \mathbb{P}(-Y \leq -t)$$

and $-Y$ follows the same distribution as if we had $|X|$ instead of $X \leq 0$,

and $f_Y(t) = \dfrac{d\,P(Y \leq t)}{dt}$

$$= \begin{cases} \frac{1}{2|X|} f\left(\frac{-t}{|X|}\right) & \text{if } \left|\frac{t}{x}\right| \leq 1 \\ \frac{|X|}{2t^2} f\left(\frac{|X|}{-t}\right) & \text{if } \left|\frac{t}{x}\right| \geq 1 \end{cases}$$

$$= \begin{cases} \frac{1}{2|X|} f\left(\frac{t}{x}\right) & \text{if } \left|\frac{t}{x}\right| \leq 1 \\ \frac{|X|}{2t^2} f\left(\frac{X}{t}\right) & \text{if } \left|\frac{t}{x}\right| \geq 1 \end{cases}$$

2) The acceptation ratio becomes:

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

but if $\left|\frac{X}{Y}\right| \leq 1$,

$$\frac{q(Y,X)}{q(X,Y)} = \frac{\frac{1}{2|Y|} f\left(\frac{X}{Y}\right)}{\frac{|X|}{2|Y|^2} f\left(\frac{X}{Y}\right)} = \left|\frac{X}{Y}\right|$$

and if $\left|\frac{X}{Y}\right| \geq 1$,

$$\frac{q(Y,X)}{q(X,Y)} = \frac{\frac{|Y|}{2|X|^2} f\left(\frac{Y}{X}\right)}{\frac{1}{2|X|} f\left(\frac{Y}{X}\right)} = \left|\frac{X}{Y}\right|$$

Exercise 3)

1) $\mathbb{P}((X_n, Y_n) \in A \mid (X_k, Y_k) \forall k \leq n-1)$

$= \mathbb{P}(Y_n \in A_y \mid X_n \in A_x, (X_k, Y_k) \forall k \leq n-1) \mathbb{P}(X_n \in A_x \mid (X_k, Y_k) \forall k)$

where $A = A_x \times A_y$

$= \mathbb{P}(Y_n \in A_y \mid X_n \in A_x) \mathbb{P}(X_n \in A_x \mid Y_{n-1})$

because $Y_n \sim f_{Y|X}(X_n, \cdot)$ and $X_n \sim f_{X|Y}(\cdot, Y_{n-1})$

$= \mathbb{P}(Y_n \in A_y \mid X_n \in A_x, Y_{n-1}) \mathbb{P}(X_n \in A_x \mid Y_{n-1})$

$= \mathbb{P}((X_n, Y_n) \in A \mid Y_{n-1})$

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

This is a Markov chain.
Its transition kernel is:

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

$$= \iint_A f_{X|Y}(x, y_n) f_{Y|X}(x, y) \, dx \, dy$$

2) $\mathbb{P}(Y_n \in A \mid Y_k \forall k \leq n-1)$

$= \int_{\mathbb{R}^r} \mathbb{P}(Y_n \in A \mid X_n = x, Y_k \forall k \leq n-1) \, p(X_n = x \mid Y_k \forall k)$

$= \int_{\mathbb{R}^r} \mathbb{P}(Y_n \in A \mid X_n = x) \, p(X_n = x \mid Y_{n-1})$

$= P(Y_n \in A \mid Y_{n-1})$

This is a Markov chain.
Its transition kernel is:

$$P(y_n, A) = P(Y_{n+1} \in A \mid Y_n = y)$$

$$= \int_{x \in \mathbb{R}^r} \int_{y \in A} f_{X|Y}(x, y_n)\, f_{Y|X}(x, y)\, dx\, dy$$

* Invariance of $\int f_Y(y)\, dy$:

$$\int_{y \in \mathbb{R}^q} f_Y(y)\, P(y, A)\, dy$$

$$= \int_{y \in \mathbb{R}^q} f_Y(y) \left(\int_{x' \in \mathbb{R}^r} \int_{y' \in A} \underbrace{f_{X|Y}(x', y)}_{\frac{f(x', y)}{f_Y(y)}}\, f_{Y|X}(x', y')\, dx'\, dy' \right) dy$$

$$= \int_{y \in \mathbb{R}^q} \int_{x' \in \mathbb{R}^r} \int_{y' \in A} f(x', y)\, f_{Y|X}(x', y')\, dx'\, dy'\, dy$$

$$= \int_{x' \in \mathbb{R}^r} \int_{y' \in A} \Big(\underbrace{\int_y f(x', y)\, dy}_{= f_X(x')} \Big)\, \underbrace{f_{Y|X}(x', y')}_{\frac{f(x', y')}{f_X(x')}}\, dx'\, dy'$$

$$= \int_{x' \in \mathbb{R}^r} \int_{y' \in A} f(x', y')\, dx'\, dy' = \int_{y' \in A} f_Y(y')\, dy'$$

It is indeed invariant.

3) We can compute $f_X(x)$ and $f_Y(y)$.

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

$$= \int_0^{+\infty} \frac{4}{\sqrt{2\pi}} y^{\frac{3}{2}} \exp\left(-\frac{y}{2}(x^2+4)\right) dy$$

doing a first integration by part:

$$= 0 + \frac{4 \times 2 \times \frac{3}{2}}{\sqrt{2\pi}\,(x^2+4)} \int_0^{+\infty} y^{1/2} \exp\left(-\frac{y}{2}(x^2+4)\right) dy$$

and a second one:

$$= 0 + \frac{12}{\sqrt{2\pi}\,(x^2+4)^2} \int_0^{+\infty} y^{-\frac{1}{2}} \exp\left(-\frac{y}{2}(x^2+4)\right) dy$$

Changing variable $u = \sqrt{y(x^2+4)}$

$$du = \frac{\sqrt{(x^2+4)}}{2\sqrt{y}}\, dy$$

$$= \frac{12}{(x^2+4)^{5/2}}$$

The same way:

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

$$= \frac{4e^{-2y}}{\sqrt{2\pi}} y^{3/2} \mathbb{1}_{\mathbb{R}^+(y)} \int_{\mathbb{R}} \exp\left(-\frac{y x^2}{2}\right) dx$$

Changing variable $u = \sqrt{y}\, x$

$$du = \sqrt{y}\, dx$$

$$f_Y(y) = 4 e^{-2y}\, y\, \mathbb{1}_{\mathbb{R}^+}(y)$$

We can then use the following Gibbs sampler, similar to Algorithm 4,

Given (X_0, Y_0)
for $n=1$ to max iter do
$X_n \sim f_{X|Y}(\cdot, Y_{n-1}) = f(X_n | Y_{n-1})$

$Y_n \sim f_{Y|X}(X_n, \cdot) = f(Y_n | X_n)$
end
return the last couple,

4) $$\int_{\mathbb{R}} \frac{H(x)}{(4+x^2)^{5/2}}\, dx = \frac{1}{12}\, \mathbb{E}\left[H(X) \mid X \sim f_X \right]$$

In question 2 we showed that f_Y was a stationary distribution of $\{Y_n, n \geqslant 0\}$, we can show similarly that f_X is a stationary distribution of $\{X_n, n \geqslant 0\}$, the previous integral can then be approached using a Monte Carlo method using the samples $(X_n)_{n \geqslant 0}$,

ie $$\frac{1}{12} \sum_{n=0}^{N} H(X_n) \xrightarrow[a.s.]{N \to +\infty} \int_{\mathbb{R}} \frac{H(x)}{(4+x^2)^{5/2}}\, dx$$