Modèle probabiliste : stats $\rightarrow p_\theta(x)$

Bayesien $\rightarrow$ $p(x,\theta) = \underbrace{p(x|\theta)}_{\text{vraisemblance}} \underbrace{p(\theta)}_{\text{prior}}$

x fixé

$$\max_\theta \log p(x,\theta) = \max_\theta \log p(x|\theta) + \log p(\theta)$$

Ex $p(\theta) \sim N(\underline{0}, \underline{\sigma^2})$ $\rightarrow$ $\log p(\theta) = -\frac{1}{2\sigma^2}\|\theta\|^2$

θ_0 prior

$$\max \log p(x,\theta) = \min_\theta \underbrace{-\log p(x|\theta)}_{NLL} + \frac{1}{2\sigma^2}\|\theta\|^2$$

$\theta - \theta_0$

Paramètres (θ)
- $\bar{t_0}$
- $\bar{v_0}$
- σ_τ^2
- σ_ξ^2
- σ^2

$\Rightarrow$ 5 params

Variable latente (z)
- t_0
- v_0

z_{pop}

- $(\tau_i)_{i\in[1,N]}$
- $(\xi_i)_{i\in[1,N]}$

$(z_i)_{i\in[1,N]}$

$z \in \mathbb{R}^{2+2N}$

Observations
$(y_{i,j})$ $i\in[1,N]$, $j\in[1,k_i]$

$(k_i = k)$

$\Rightarrow Y \in \mathbb{R}^{N\times k}$

Ex $t_0 \sim N(\bar{t_0}, \sigma_{t_0}^2)$ $\Leftrightarrow$ loi de variable latente | paramètre

z

__Modèle__ $\theta \longrightarrow z \longrightarrow y$

__Modèle probabiliste__ $p(y, z, \theta) \rightsquigarrow p(\{y_{ij}\}, \{z_i\}_i, z_{pop}, \theta) = P$

variable latente

paramètre

$$\log P = \log p(\{y_{ij}\} \mid \{z_i\}_i, z_{pop}, \theta) \quad (1)$$
$$+ \log p(\{z_i\}_i, z_{pop} \mid \theta) \quad (2)$$
$$+ \log p(\theta) \quad (3)$$

$$\Big[\; y_{ij} \sim \mathcal{N}(d_i(t_{ij}), \sigma^2) \;\Big]$$

(z) ; θ

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

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

$\hookrightarrow d_i(t_{ij}) = \underset{\text{fixé}}{p_0} + \alpha_i v_0 (t_{ij} - t_0 - \tau_i)$

$$(2) \overset{cte}{=} \log p(\{z_i\}_i \mid \theta_{ind}) + \log p(z_{pop} \mid \theta_{pop})$$

$$= \sum_{i=1}^{N} \underbrace{\log p(z_i \mid \theta_{ind})}_{\log p(\tau_i \mid \theta) + \log p(\alpha_i \mid \theta)} + \log p(z_{pop} \mid \theta_{pop})$$

t_0, v_0 ; $\langle \cdot, \cdot \rangle$

$\alpha_i = \exp(\xi_i)$

$\xi_i \sim \mathcal{N}(0, \sigma_\xi^2)$

$$= -\frac{N}{2}\log(\sigma_\tau^2) - \frac{1}{2}\sum \frac{\tau_i^2}{2\sigma_\tau^2} - \frac{N}{2}\log(\sigma_\xi^2) - \frac{1}{2}\sum_{i=1}^{N} \frac{\log(\alpha_i)^2}{2\sigma_\xi^2}$$

$\downarrow \varphi(\theta)$; $\varphi(\theta)$

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

$$(v_0 - \bar{v}_0)^2 = v_0^2 + \bar{v}_0^2 - 2\langle v_0, \bar{v}_0 \rangle$$

$\varphi(\theta)$; $\langle \cdot, \cdot \rangle$

prior $\;\; cte \leftarrow \sigma_{t_0}^2, \sigma_{v_0}^2 \dots$

$\log p(\bar{t}_0) + \log p(\bar{v}_0) + \log p(\sigma_\xi^2) + \log p(\sigma_\tau^2)$

(3) ii) $\log p(\theta) = \log p(t_0) + \log p(w_0) + \log p(\sigma_\varepsilon^2) + \log p(\sigma_c^2) + \log p(\sigma^2)$

$$\stackrel{cte}{=} -\frac{(\bar{t}_0 - \bar{\bar{t}}_0)^2}{2 s_{t_0}^2} - \frac{(\bar{w}_0 - \bar{\bar{w}}_0)^2}{2 s_{w_0}^2}$$

n_ε, n_c, n

$$-\left(\frac{n}{2}+1\right)\left[\log(\sigma_\varepsilon^2) + \log(\sigma_c^2) + \log(\sigma^2)\right]$$

$$-\frac{v_\varepsilon^2}{2\sigma_\varepsilon^2} - \frac{v_c^2}{2\sigma_c^2} - \frac{v^2}{2\sigma^2} \rightarrow \varphi(\theta)$$

$$\log f_{W^{-1}}(\sigma^2) \stackrel{cte}{=} -b\log(\sigma^2) - \frac{n}{2}\log(\sigma^2) - \frac{v^2}{2\sigma^2}$$

Modèle exponentiel $\log p(x,\theta) = -\varphi(\theta) + \langle S(x), \psi(\theta)\rangle$

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

$z \in$ variables latentes

$$\varphi(\theta) = \underbrace{\frac{\bar{t}_0^2}{2\sigma_{t_0}^2}}_{z|\theta} + \underbrace{\frac{\bar{t}_0^2}{2 s_{t_0}^2}}_{\text{prior}} - \bar{t}_0 \cdot \frac{\bar{\bar{t}}_0}{s_{t_0}^2} + \left(\frac{N}{2} + \frac{n_\varepsilon}{2} + 1\right)\log(\sigma_\varepsilon^2)$$

$$+\frac{\bar{w}_0^2}{2\sigma_{w_0}^2} + \frac{\bar{w}_0^2}{2 s_{w_0}^2} - \bar{w}_0 \frac{\bar{\bar{w}}_0}{s_{w_0}^2} + \left(\frac{N}{2} + \frac{n_c}{2} + 1\right)\log(\sigma_c^2)$$

$$+\left(\frac{1}{2}\sum_{i=1}^N k_i + \frac{n}{2} + 1\right)\log(\sigma^2)$$

$$S(y,z) = \left[\begin{array}{l} \bullet\ \frac{1}{\sum_{i=1} k_i} \cdot \sum_{i=1}^N \sum_{j=1}^{k_i} (y_{ij} - d_i(t_{ij}))^2 \quad \sigma^2 \\ \bullet\ \frac{1}{N}\sum_{i=1}^N \log(\alpha_i)^2 \quad / \sigma_\varepsilon^2 \\ \bullet\ \frac{1}{N}\sum_{i=1}^N z_i^2 \quad / \sigma_c^2 \\ \bullet\ t_0 \quad ; \quad \bar{t}_0 \\ \bullet\ \ldots \end{array}\right] \in \mathbb{R}^J$$

$$\Psi(\Theta) = \begin{bmatrix} \cdot & \cdot & (\sum_i a_i) / \alpha\, \sigma^2 \\ \cdot & - & N / \alpha\, \sigma\varepsilon^2 \\ \cdot & - & N / \alpha\, \sigma\tau^2 \\ \cdot & + & \bar{t_e} / \sigma t_e^2 \\ \cdot & + & \bar{w} / \sigma w^2 \end{bmatrix}$$

Antrice ~~inplea~~

inverse gamma

↳ suff. stats

↳ inverse Wishart

WISHART

$$\Theta^{t+1} = \underset{\Theta}{\text{argmax}}\ \mathbb{E}_{\boxed{z \sim p(\cdot \mid y, \Theta^t)}} \left[\log p(y, z, \Theta) \right] \quad \text{EM}$$

Sampling de la posterior $p((z_i), z_{pop} \mid y_i, \Theta)$

De façon générale : $p(z \mid y, \Theta) \propto p(y, z \mid \Theta)$

$$= \frac{p(y \mid z, \Theta)\, p(z \mid \Theta)}{\int p(y \mid \tilde{z}, \Theta)\, p(\tilde{z} \mid \Theta)\, d\tilde{z}}$$

cte Weibo

$$p(y \mid x) = \frac{p(x \mid y)\, p(y)}{p(x)}$$ ne dépend pas de y

$$\propto p(x \mid y)\, p(y)$$

cte de normalis

$$\log p(z \mid y, \Theta) \overset{cte(y,\Theta)}{=} \log p(y \mid z, \Theta) + \log p(z \mid \Theta)$$

↓

ICI

$$\log p((z_i)_i, z_{pop} \mid y_{ij}, \Theta) = \log p(y_{ij} \mid (z_i), z_{pop}) + \log p((z_i), z_{pop} \mid \Theta)$$

$$\overset{cte}{=} \sum \frac{(y_{ij} - a_i(t_i))^2}{2(\sigma^2)} - \sum \frac{\log(a_i)^2}{(\cdot)} + \frac{z_i^2}{(\cdot)}$$

$$\frac{(t_0 - \bar{t_0})^2}{(\cdot)} - \frac{(w - \bar{w})^2}{(\cdot)}$$

Metropolis Hastings

on veut π : on crée une chaîne de Markov

$(X_n)_n$

On se donne X_n $\longrightarrow$ X_{n+1} ?

Ex $q(\cdot|x) \sim N(x, \sigma^2)$

proposal

① proposition $\tilde{X}_{n+1} \sim q(\cdot \mid X_n)$

suit

② Acceptation Rejet $\alpha_{n+1} = 1 \wedge \dfrac{\pi(\tilde{X}_{n+1})\, q(X_n \mid \tilde{X}_{n+1})}{\pi(X_n)\, q(\tilde{X}_{n+1} \mid X_n)}$

état courant — proposition

$\pi(x) P(x,y) = \pi(y) P(y,x)$

detailed balance

Pourquoi le ratio ?

quand on calcul detailed balance,

$P(x,y) \to \alpha_{n+1}$ $\qquad \dfrac{\pi(y)\, q(x|y)}{\pi(x)\, q(y|x)}$ $\qquad \Rightarrow \pi$ invariant pour P

③ Accepter $\tilde{X}_{n+1} \to X_{n+1}$

avec probabilité α_{n+1}

Intérêts

- ratio α $\to$ pas besoin de constante de normalisation

$\pi(\tilde{X}_{n+1}) / \pi(X_n)$ $\qquad \pi = \dfrac{e^{-U(x)}}{Z}$

- q symétrique $\qquad q(y|x) = q(x|y)$

$\qquad N(y; x, \sigma^2) \qquad N(x; y, \sigma^2)$

$\longrightarrow \alpha = \dfrac{\pi(\tilde{X}_{n+1})}{\pi(X_n)}$

Ici on veut échantillonner de la posterior

on a $\log p((z_i), z_{pop} \mid (y_{ij}), \theta)$ à cte près

ne dépend de Z

→ MH avec proposition gaussienne   hyperpar $= \|\sigma^2_{prop}\|$

$$\tilde{\alpha}_{n+1} = \min\left(0, \log \pi(z^*) - \log \pi(z^{i})\right)$$

prop

π : $p(z \mid y, \theta)$ qu'apparaît

$$\alpha_{n+1} = \exp\left(\tilde{\alpha}_{n+1}\right)$$

Sampling dans $\mathbb{R}^{2N+2}$ $\sim \mathbb{R}^{200}$

- Astuce : faire varier σ^2 → moyenne acceptation 20% - 40%

Intuition $\sigma^2 \nearrow$ ⇒ beaucoup rejeté

$\sigma^2 \searrow$ ⇒ beaucoup accepté

- vraie donnée y_i, vrais paramètres θ ⟶ est ce qu'on retrouve la bonne distribution des Z

- $\Sigma = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_N^2 \end{pmatrix}$ ⊕ corrélation aux z_i

En théorie ⇒ 1 sample de la posterior

$N \times$ (• N_{MH} itérations MH

↳ Espérance (EM)

Nouvelle idée EM !

⇒ SAEM

$$\mathbb{E}_{p(z|y,\theta_t)}\left[\log p(z,y,\theta)\right] \approx \log p(\tilde{z},y,\theta)$$

1 sample de la posterior $= \tilde{z}$

SAEM

$$Q_{t+1}(\theta) = \mathbb{E}_{p(z|y,\theta_t)}\left[\log p(z,y,\theta)\right]$$

tiré de

hyper

$$Q_{t+1}(\theta) = Q_t(\theta) + \gamma_{t+1}\left(\log p(y,z^{t+1},\theta) - Q_t(\theta)\right)$$

$$= (1-\gamma_{t+1})\,Q_t(\theta) + \gamma_{t+1}\log p(z^{t+1})$$

1 sample

$$\sum \gamma_t \to +\infty$$

$$\sum \gamma_t^2 \to CV$$

Modèle exponentiel $\quad \log p(y,z,\theta) = -\varphi(\theta) + \langle S(y,z), \psi(\theta)\rangle$

Update SAEM

$$Q_{t+1}(\theta) = Q_t(\theta) + \gamma_{t+1}\left[-\varphi(\theta) + \langle S(y,z^{t+1}), \psi(\theta)\rangle - Q_t(\theta)\right]$$

Récurrence revient à updater S^t

$$S^{t+1} = S^t + \gamma_{t+1}\left(S(y,z^{t+1}) - S^t\right)$$

modèle exponentiel

$S^0 = 0$

$$Q_{t+1}(\theta) = -\varphi(\theta) + \langle S^{t+1}, \psi(\theta)\rangle$$

↳ dépend du style

$E \xrightarrow{SAEM}$ calcul de S^k

$Z u + \varepsilon \sim p(\cdot \mid y, \theta^k)$

Maximi $\xrightarrow{SAEM}$ argmax$_\theta$ $\underbrace{-\psi(\theta) + \langle S^k, \Psi(\theta) \rangle}_{\text{Convexe en } \theta}$

maximum θ^* est unique et vérifie $\nabla_\theta \left(-\psi(\theta) + \langle S^k, \Psi(\theta) \rangle \right)_{\theta = \theta^*} = 0$

TO DO

- calcul $\nabla_\theta \psi(\theta)$

$\theta = \begin{bmatrix} \cdot\ \bar{t_0} \\ \cdot\ \bar{v_0} \\ \cdot\ \sigma_\tau^2 \\ \cdot\ \sigma_\xi^2 \\ \cdot\ \sigma^2 \end{bmatrix}$

$\begin{bmatrix} \circ \text{ affine en } \bar{t_0} \\ \circ \text{ affine en } \bar{v_0} \\ \circ \ \ldots / \sigma_\tau^2 \\ \circ \ \ldots / \sigma_\xi^2 \\ \circ \ \ldots / \sigma^2 \end{bmatrix}$

$\Psi : \mathbb{R}^5 \to \mathbb{R}^5$

$\partial \Psi : \mathbb{R}^5 \to \underbrace{\mathbb{R}^{5 \times 5}}_{\text{diagonale}}$

- calcule de $\nabla_\theta \langle S_k, \Psi(\theta) \rangle$

$\Psi(\theta) = \begin{bmatrix} f(\bar{t_0}) \\ f(\bar{v_0}) \\ f(\sigma_\tau^2) \\ \vdots \end{bmatrix}$

$\begin{bmatrix} \cdot\ S_k(1) \times \partial_{(1)} \Psi(\theta) \\ \cdot\ S_k(2) \times \partial_{(2)} \Psi(\theta) \\ \vdots \end{bmatrix}$

- $= 0 \Rightarrow$ closed form sur θ

- Implémentation

  - $z^0 = 0$
  - $\theta^0 \ldots$ loin de θ^*

- Θ^c pas trop loin de Θ

- En tout 1000 itérations
  ↳ MH : 100 itérations

- Burn-in : $\epsilon_k = 1$ au début $1 \to N_b$

- MH filter ~ 20-40 % (σ^2_{prop})
  $\in [10^{-5}, 10^{-4}]$

CHECK

- $\Theta^0 = \Theta^*$ $\xrightarrow{\text{vérifier}}$ on converge vers z_i

- $z^0 = z^i$ $\xrightarrow{\text{vérifier}}$ on converge vers Θ^*

Sampler de $p(\overbrace{z_1, z_2 \ldots z_n}^{z_i}, z_{pop} \mid y, \Theta)$

état donné (z_i) (z_{pop})

Gibbs en 2 temps

1 étape

- $\tilde{z}_{pop} \sim p(z_{pop} \mid z_i, y, \Theta)$
- $(\tilde{z}_i) \sim p((z_i) \mid \tilde{z}_{pop}, y, \Theta)$
- $\tilde{z}_{pop} \sim p(z_{pop} \mid \tilde{z}_i, y, \Theta)$

$\textcircled{\sigma}$ $p(z_i \mid z_{pop}, y, \Theta)$ $\boxed{\text{Gibbs}}$ $(\rightarrow) = (\alpha_i, z_i)$

⚠ condition à z_{pop}, (z_i) sont indépendants

→ on peut tout paralléliser (c_i, κ_i) = état courant
on se donne

$(\ldots z_i)$ Gibbs MH

$z_i = (\alpha_i, c_i)$

① $\tilde{c}_i \overset{\text{MH}}{\sim} p(\cdot \mid \alpha_i, y, \theta)$

② $\tilde{\alpha}_i \overset{\text{MH}}{\sim} p(\cdot \mid \tilde{c}_i, y, \theta)$

Requirement

$\log p(\cdot \mid \alpha_i, y, \theta) = ?$ ← • on regarde toute la log densité $p(y, (z_i), \theta)$

$\log p(\alpha_i \mid c_i, y, \theta) = ?$ • on garde uniquement ce qui dépend de la variable aléatoire

$$p(z \mid y) \propto p(y \mid z)\, p(z) \propto p(y, z)$$

• jointe ⇒ ce qui dépend de z

• $z_{pop} \mid z_i$ $z_{pop} = (t_0, w_0)$

(t_0, w_0) état courant

① $\tilde{t}_0 \sim p(\cdot \mid w_0, y, \theta)$

$\tilde{w}_0 \sim p(\cdot \mid \tilde{t}_0, y, \theta)$

Sampler de la posterior

• Avant dans $\mathbb{R}^{2N+2}$

• intérieur $(2N+2)$ étapes dans $\mathbb{R}$

• état courant $(z_i)_i$, (z_{pop})

① $\tilde{z}_i \overset{\text{Gibbs}}{\sim} p(z_i \mid z_{pop}, y, \theta)$ Gibbs → α_i, → c_i

↳ question σ

② $\tilde{z}_{pop} \sim p(z_{pop} \mid \tilde{z}_i, y, \theta)$ Gibbs → t_0

↳ question σ

En pratique

($\Rightarrow$), zpop

- $\tilde{\alpha}_i \sim p(\cdot \mid t_i, zpop, \theta)$
- $\tilde{z}_i \sim p(\cdot \mid \tilde{\alpha}_i, zpop)$

, $i = 1 \dots N$ (indépendant)

[MH → 1 itération ; → plusieurs itérations]

- zpop →

Astuce NUMBA

: précompiler toutes les fonctions numpy-based

```
from numba import jit

@jit(nopython=True)
def moyenne(x):
    return np.mean(x)
```

```
moyenne([0, 1])
```

def fonction ← NUMBA

Application → appeler la fonction

⚠ les fonctions acceptent uniquement numpy

→ log, exp, *