TP6 – Calculs EM

$x = (x_1, \ldots, x_n)$ variables observées $y = (y_1, \ldots, y_n)$ variables cachées

E–step

$$\begin{aligned}
\mathbb{E}_{y|x,\theta^{(m)}}\left[\log p(x, y \mid \theta)\right] &= \mathbb{E}_{y|x,\theta^{(m)}}\left[\log \prod_{i=1}^{n} p(x_i, y_i \mid \theta)\right] \\
&= \mathbb{E}_{y|x,\theta^{(m)}}\left[\sum_{i=1}^{n} \log p(x_i, y_i \mid \theta)\right] \\
&= \sum_{i=1}^{n} \mathbb{E}_{y_i|x_i,\theta^{(m)}}\left[\log p(x_i, y_i \mid \theta)\right]
\end{aligned}$$

À partir de maintenant, on fixe l'indice i. On va expliciter l'expression suivante :

$$\mathbb{E}_{y_i|x_i,\theta^{(m)}}\left[\log p(x_i, y_i \mid \theta)\right] \tag{1}$$

Déterminons tout d'abord $p(x_i, y_i \mid \theta)$:

$$\begin{aligned}
p(x_i, y_i \mid \theta) &= p(x_i \mid y_i, \theta) \cdot p(y_i \mid \theta) \\
&= \begin{cases} p(x_i \mid y_i = 1, \theta) \cdot p(y_i = 1 \mid \theta) & \text{si } y_i = 1, \\ p(x_i \mid y_i = 0, \theta) \cdot p(y_i = 0 \mid \theta) & \text{si } y_i = 0 \end{cases} \\
&= \begin{cases} \phi(x_i, \mu, \sigma) \cdot \pi \quad \text{si } y_i = 1, \\ c \cdot \mathbb{1}_{[-a,a]}(x_i) \cdot (1 - \pi) \quad \text{si } y_i = 0 \end{cases}
\end{aligned}$$

En utilisant la notation condensée,

$$p(x_i, y_i \mid \theta) = (\phi(x_i, \mu, \sigma) \cdot \pi)^{y_i} \left(c \cdot \mathbb{1}_{[-a,a]}(x_i) \cdot (1 - \pi)\right)^{1-y_i}$$

On trouve alors,

$$\log p(x_i, y_i \mid \theta) = y_i \log(\pi \cdot \phi(x_i, \mu, \sigma)) + (1 - y_i) \log(c(1 - \pi)\mathbb{1}_{[-a,a]}(x_i))$$

D'après (1), il faut ensuite prendre l'espérance de l'expression précédente par rapport à y_i, en supposant que y_i suit la distribution $p(y_i \mid x_i, \theta^{(m)})$. On trouve donc,

$$\begin{aligned}
\mathbb{E}_{y_i|x_i,\theta^{(m)}}\left[\log p(x_i, y_i \mid \theta)\right] = \;& \mathbb{E}_{y_i|x_i,\theta^{(m)}}(y_i) \log(\pi \cdot \phi(x_i, \mu, \sigma)) \\
&+ (1 - \mathbb{E}_{y_i|x_i,\theta^{(m)}}(y_i)) \log(c(1 - \pi)\mathbb{1}_{[-a,a]}(x_i))
\end{aligned}$$

Pour calculer la quantité $\mathbb{E}_{y_i|x_i,\theta^{(m)}}(y_i) = \mathbb{E}(y_i \mid x_i, \theta^{(m)})$, on a besoin de la fonction de masse $p(y_i \mid x_i, \theta^{(m)})$. Elle s'écrit,

$$\begin{aligned} p(y_i \mid x_i, \theta^{(m)}) &= \frac{p(x_i, y_i \mid \theta^{(m)})}{p(x_i \mid \theta^{(m)})} \\ &= \begin{cases} \dfrac{\phi(x_i, \mu^{(m)}, \sigma^{(m)})\pi^{(m)}}{\pi^{(m)}\phi(x_i, \mu^{(m)}, \sigma^{(m)}) + (1-\pi^{(m)})c} & \text{si } y_i = 1, \\ \dfrac{c \cdot \mathbb{1}_{[-a,a]}(x_i) \cdot (1-\pi^{(m)})}{\pi^{(m)}\phi(x_i, \mu^{(m)}, \sigma^{(m)}) + (1-\pi^{(m)})c} & \text{si } y_i = 0, \end{cases} \end{aligned}$$

On pose

$$q(x_i, \theta^{(m)}) = \frac{\phi(x_i, \mu^{(m)}, \sigma^{(m)})\pi^{(m)}}{\pi^{(m)}\phi(x_i, \mu^{(m)}, \sigma^{(m)}) + (1-\pi^{(m)})c}$$

et donc

$$1 - q(x_i, \theta^{(m)}) = \frac{c \cdot \mathbb{1}_{[-a,a]}(x_i) \cdot (1-\pi^{(m)})}{\pi^{(m)}\phi(x_i, \mu^{(m)}, \sigma^{(m)}) + (1-\pi^{(m)})c}$$

La variable aléatoire $y_i \mid x_i, \theta^{(m)}$ est un Bernouilli de paramètre $q(x_i, \theta^{(m)})$. On a donc $\mathbb{E}(y_i \mid x_i, \theta^{(m)}) = q(x_i, \theta^{(m)})$. Finalement,

$$\begin{aligned} &\mathbb{E}_{y_i|x_i,\theta^{(m)}}[\log p(x_i, y_i \mid \theta)] = \\ &\qquad q(x_i, \theta^{(m)}) \cdot \log(\pi \cdot \phi(x_i, \mu, \sigma)) + (1 - q(x_i, \theta^{(m)})) \cdot \log(c(1-\pi)\mathbb{1}_{[-a,a]}(x_i)) \end{aligned}$$

Il reste à sommer pour tout i,

$$\begin{aligned} &\mathbb{E}_{y|x,\theta^{(m)}}[\log p(x, y \mid \theta)] = \\ &\qquad \sum_{i=1}^{n} q(x_i, \theta^{(m)}) \cdot \log(\pi \cdot \phi(x_i, \mu, \sigma)) + (1 - q(x_i, \theta^{(m)})) \cdot \log(c(1-\pi)\mathbb{1}_{[-a,a]}(x_i)) \end{aligned}$$

M–step On cherche à présent à trouver les paramètres $\pi^{(m+1)}$, $\mu^{(m+1)}$ et $\sigma^{(m+1)}$ qui maximise l'espérance calculée juste au dessus. On annule donc les dérivées partielles.

$$\frac{\partial \mathbb{E}_{y|x,\theta^{(m)}}[\log p(x, y \mid \theta)]}{\partial \pi} = \sum_{i=1}^{n} \frac{q(x_i, \theta^{(m)})}{\pi} - \frac{1 - q(x_i, \theta^{(m)})}{1-\pi} = 0$$

$$\frac{\partial \mathbb{E}_{y|x,\theta^{(m)}}[\log p(x, y \mid \theta)]}{\partial \mu} = \sum_{i=1}^{n} q(x_i, \theta^{(m)}) \frac{x_i - \mu}{\sigma} = 0$$

$$\frac{\partial \mathbb{E}_{y|x,\theta^{(m)}}[\log p(x, y \mid \theta)]}{\partial \sigma} = \sum_{i=1}^{n} q(x_i, \theta^{(m)}) \frac{(x_i - \mu)^2}{\sigma^3} - \frac{q(x_i, \theta^{(m)})}{\sigma} = 0$$

Et on trouve,

$$\pi^{(m+1)} = \frac{\displaystyle\sum_{i=1}^{n} q(x_i, \theta^{(m)})}{n}$$

$$\mu^{(m+1)} = \frac{\sum_{i=1}^{n} q(x_i, \theta^{(m)}) x_i}{\sum_{i=1}^{n} q(x_i, \theta^{(m)})}$$

$$\sigma^{(m+1)} = \sqrt{\frac{\sum_{i=1}^{n} q(x_i, \theta^{(m)}) (x_i - \mu^{(m+1)})^2}{\sum_{i=1}^{n} q(x_i, \theta^{(m)})}}$$