CORRECTION DE LA FEUILLE N° 9

Exercice 1. *Mélange de lois de Poisson*
C'est le corrigé d'un exercice légèrement différent où on considère un mélange de **deux** *lois de Poisson seulement et où on notait les variables latentes u_i au lieu de z_i. Tous les calculs sont pareils dans le cas général d'un mélange à K composants, à part la question 4d) où il faut utiliser un multiplicateur de Lagrange pour la maximisation en $(\pi_1, \dots, \pi_K)$ afin de prendre en compte la contrainte $\sum_k \pi_k = 1$.*

1. Soit U une variable aléatoire discrète à valeurs dans $\{1, \dots, K\}$ avec probabilités $\mathbb{P}_\theta(Z_i = k) = \pi_k$, $k = 1, \dots, K$. Soient $V_k, k = 1, \dots, K$ des variables aléatoires de loi $V_k \sim \mathrm{Poi}(\lambda_k)$. Supposons que les U et $V_k, k = 1, \dots, K$ sont mutuellement indépendants. Afin de construire une v.a. X d'un mélange de lois de Poisson, on pose
$$X = \sum_{j=1}^{K} V_j \mathbb{1}\{U = j\} = V_U \,. \tag{1}$$

2. On a
$$\begin{aligned}
\mathbb{P}_\theta(X = k) &= \mathbb{P}_\theta(X = k|U = 1)\mathbb{P}_\theta(U = 1) + \mathbb{P}_\theta(X = k|U = 2)\mathbb{P}_\theta(U = 2) \qquad \text{(formule des probabilités totales)}\\
&= \pi\mathbb{P}_\theta\left(\sum_{j=1}^{2} V_j\mathbb{1}\{U=j\} = k \middle| U = 1\right) + (1-\pi)\mathbb{P}_\theta\left(\sum_{j=1}^{2} V_j\mathbb{1}\{U=j\} = k \middle| U = 2\right) \qquad \text{(par (1))}\\
&= \pi\mathbb{P}_\theta\left(V_1 = k|U = 1\right) + (1-\pi)\mathbb{P}_\theta\left(V_2 = k|U = 2\right)\\
&= \pi\mathbb{P}_\theta\left(V_1 = k\right) + (1-\pi)\mathbb{P}_\theta\left(V_2 = k\right) \qquad (\text{car } V_j \perp\!\!\!\perp U)\\
&= \frac{1}{k!}\left(\pi\lambda_1^k \mathrm{e}^{-\lambda_1} + (1-\pi)\lambda_2^k\mathrm{e}^{-\lambda_2}\right) \qquad (\text{car } V_j \sim \mathrm{Poi}(\lambda_j))
\end{aligned}$$

3. La fonction de log-vraisemblance des données (incomplètes) $\mathbf{x}$ est donnée par
$$\begin{aligned}
\ell(\theta) &= \sum_{i=1}^{n} \log\left(\mathbb{P}_\theta(X = x_i)\right) \qquad \text{(cas discret i.i.d.)}\\
&= \sum_{i=1}^{n} \log\left(\frac{1}{x_i!}\left(\pi\lambda_1^{x_i}\mathrm{e}^{-\lambda_1} + (1-\pi)\lambda_2^{x_i}\mathrm{e}^{-\lambda_2}\right)\right)\\
&= -\sum_{i=1}^{n}\log\left(x_i!\right) + \sum_{i=1}^{n}\log\left(\pi\lambda_1^{x_i}\mathrm{e}^{-\lambda_1} + (1-\pi)\lambda_2^{x_i}\mathrm{e}^{-\lambda_2}\right)\,.
\end{aligned}$$
Pour les dérivées partielles de $\ell(\theta)$ on obtient
$$\begin{aligned}
\frac{\partial}{\partial\pi}\ell(\theta) &= \frac{\lambda_1^{x_i}\mathrm{e}^{-\lambda_1} - \lambda_2^{x_i}\mathrm{e}^{-\lambda_2}}{\pi\lambda_1^{x_i}\mathrm{e}^{-\lambda_1} + (1-\pi)\lambda_2^{x_i}\mathrm{e}^{-\lambda_2}}\\
\frac{\partial}{\partial\lambda_1}\ell(\theta) &= \pi\lambda_1^{x_i-1}\mathrm{e}^{-\lambda_1}\sum_{i=1}^{n}\frac{x_i - \lambda_1}{\pi\lambda_1^{x_i}\mathrm{e}^{-\lambda_1} + (1-\pi)\lambda_2^{x_i}\mathrm{e}^{-\lambda_2}}\\
\frac{\partial}{\partial\lambda_2}\ell(\theta) &= (1-\pi)\lambda_2^{x_i-1}\mathrm{e}^{-\lambda_2}\sum_{i=1}^{n}\frac{x_i - \lambda_2}{\pi\lambda_1^{x_i}\mathrm{e}^{-\lambda_1} + (1-\pi)\lambda_2^{x_i}\mathrm{e}^{-\lambda_2}}
\end{aligned}$$

Il est clair que $\nabla \ell(\theta) = 0$ n'admet pas de solution explicite. Alors, l'EMV n'est pas explicite non plus.

4. a) Pout tout $x \in \mathbb{N}$ et $j \in \{1, 2\}$, on a par le théorème de Bayes

$$
\begin{aligned}
\mathbb{P}_\theta(U = j|X = x) &= \frac{\mathbb{P}_\theta(U = j)\mathbb{P}_\theta(X = x|U = j)}{\mathbb{P}_\theta(X = x)} = \frac{\mathbb{P}_\theta(U = j)\mathbb{P}_\theta(V_j = x|U = j)}{\frac{1}{x!}\left(\pi\lambda_1^x \mathrm{e}^{-\lambda_1} + (1-\pi)\lambda_2^x \mathrm{e}^{-\lambda_2}\right)} \\
&= \frac{\mathbb{P}_\theta(U = j)\mathbb{P}_\theta(V_j = x)}{\frac{1}{x!}\left(\pi\lambda_1^x \mathrm{e}^{-\lambda_1} + (1-\pi)\lambda_2^x \mathrm{e}^{-\lambda_2}\right)} = \frac{\mathbb{P}_\theta(U = j)\frac{\lambda_j^x}{x!}\mathrm{e}^{-\lambda_j}}{\frac{1}{x!}\left(\pi\lambda_1^x \mathrm{e}^{-\lambda_1} + (1-\pi)\lambda_2^x \mathrm{e}^{-\lambda_2}\right)} \\
&= \begin{cases} \frac{\pi\lambda_1^x \mathrm{e}^{-\lambda_1}}{\pi\lambda_1^x \mathrm{e}^{-\lambda_1} + (1-\pi)\lambda_2^x \mathrm{e}^{-\lambda_2}} & \text{si } j = 1 \\ \frac{(1-\pi)\lambda_2^x \mathrm{e}^{-\lambda_2}}{\pi\lambda_1^x \mathrm{e}^{-\lambda_1} + (1-\pi)\lambda_2^x \mathrm{e}^{-\lambda_2}} & \text{si } j = 2 \end{cases}
\end{aligned}
$$

b) On a

$$
\begin{aligned}
\mathcal{L}(\mathbf{x}, \mathbf{u}; \theta) &= \prod_{i=1}^n \mathbb{P}_\theta(X = x_i, U = u_i) = \prod_{i=1}^n \mathbb{P}_\theta(U = u_i)\mathbb{P}_\theta(X = x_i|U = u_i) \\
&= \prod_{i=1}^n \mathbb{P}_\theta(U = u_i)\mathbb{P}_\theta(V_{u_i} = x_i|U = u_i) = \prod_{i=1}^n \mathbb{P}_\theta(U = u_i)\mathbb{P}_\theta(V_{u_i} = x_i) \\
&= \prod_{i=1}^n \mathbb{P}_\theta(U = u_i)\frac{\lambda_{u_i}^{x_i}}{x_i!}\mathrm{e}^{-\lambda_{u_i}} \\
&= \prod_{i=1}^n \left[\pi\frac{\lambda_1^{x_i}}{x_i!}\mathrm{e}^{-\lambda_1}\right]^{\mathbb{1}\{u_i=1\}} \left[(1-\pi)\frac{\lambda_2^{x_i}}{x_i!}\mathrm{e}^{-\lambda_2}\right]^{\mathbb{1}\{u_i=2\}} .
\end{aligned}
$$

Notons $r = \sum_{i=1}^n \mathbb{1}\{u_i = 1\} = n - \sum_{i=1}^n \mathbb{1}\{u_i = 2\}$ le nombre de u_i égal à 1. Alors, la fonction de log-vraisemblance s'écrit

$$
\begin{aligned}
\log\left(\mathcal{L}(\mathbf{x}, \mathbf{u}; \theta)\right) &= \sum_{i:u_i=1} [\log \pi + x_i \log \lambda_1 - \log(x_i!) - \lambda_1] \\
&\quad + \sum_{i:u_i=2} [\log(1-\pi) + x_i \log \lambda_2 - \log(x_i!) - \lambda_2] \\
&= r \log \pi + \log \lambda_1 \sum_{i:u_i=1} x_i - \sum_{i=1}^n \log(x_i!) - r\lambda_1 \\
&\quad + (n-r)\log(1-\pi) + \log \lambda_2 \sum_{i:u_i=2} x_i - (n-r)\lambda_2 .
\end{aligned}
$$

c) Notons $R = \sum_{i=1}^n \mathbb{1}\{U_i = 1\}$. On a

$$
\begin{aligned}
Q(\theta|\theta') &= \mathbb{E}_{\theta'}[\log(\mathcal{L}(\mathbf{x}, \mathbf{U}; \theta))|\mathbf{X} = \mathbf{x}] \\
&= \mathbb{E}_{\theta'}[R|\mathbf{X} = \mathbf{x}] \log \pi + \log \lambda_1 \sum_{i=1}^n x_i \mathbb{E}_{\theta'}[\mathbb{1}\{U_i = 1\}|\mathbf{X} = \mathbf{x}] - \sum_{i=1}^n \log(x_i!) - \mathbb{E}_{\theta'}[R|\mathbf{X} = \mathbf{x}]\lambda_1 \\
&\quad + (n - \mathbb{E}_{\theta'}[R|\mathbf{X} = \mathbf{x}])\log(1-\pi) + \log \lambda_2 \sum_{i=1}^n x_i \mathbb{E}_{\theta'}[\mathbb{1}\{U_i = 2\}|\mathbf{X} = \mathbf{x}] - (n - \mathbb{E}_{\theta'}[R|\mathbf{X} = \mathbf{x}])\lambda_2
\end{aligned}
$$

Or, $\mathbb{E}_{\theta'}[\mathbb{1}\{U_i = 1\}|\mathbf{X} = \mathbf{x}] = \mathbb{P}_{\theta'}(U_i = 1|\mathbf{X} = \mathbf{x})$ et $\mathbb{E}_{\theta'}[R|\mathbf{X} = \mathbf{x}] = \sum_{i=1}^n \mathbb{P}_{\theta'}(U_i = 1|\mathbf{X} = \mathbf{x})$. Notons alors

$$
\begin{aligned}
p_{i,\theta'}(1) &= \mathbb{P}_{\theta'}(U_i = 1|X_i = x_i) = \frac{\pi'(\lambda_1')^{x_i}\mathrm{e}^{-\lambda_1'}}{\pi'(\lambda_l')^{x_i}\mathrm{e}^{-\lambda_l'} + (1-\pi')(\lambda_2')^{x_i}\mathrm{e}^{-\lambda_2'}} \qquad (2) \\
p_{i,\theta'}(2) &= 1 - p_{i,\theta'}(1)
\end{aligned}
$$

qui sont les probabilités calculées à la question a). Alors,

$$Q(\theta|\theta') = \log \pi \sum_{i=1}^n p_{i,\theta'}(1) + \log \lambda_1 \sum_{i=1}^n x_i p_{i,\theta'}(1) - \sum_{i=1}^n \log(x_i!) - \lambda_1 p_{i,\theta'}(1)$$
$$+ \log(1-\pi)\left(n - \sum_{i=1}^n p_{i,\theta'}(1)\right) + \log \lambda_2 \sum_{i=1}^n x_i p_{i,\theta'}(2) - \lambda_2 \left(n - \sum_{i=1}^n x_i p_{i,\theta'}(1)\right).$$

d) Remarquons que les paramètres $\pi, \lambda_1, \lambda_2$ sont bien séparés dans la fonction $Q(\theta|\theta')$, ce qui permet de décomposer le problème de maximiser $\theta \mapsto Q(\theta|\theta')$ en plusieurs problèmes indépendants. Considérons d'abord la maximisation en λ_1. Pour les dérivées on trouve

$$\frac{\partial}{\partial \lambda_1} Q(\theta|\theta') = \frac{1}{\lambda_1} \sum_{i=1}^n x_i p_{i,\theta'}(1) - \sum_{i=1}^n p_{i,\theta'}(1)$$
$$\frac{\partial^2}{\partial^2 \lambda_1} Q(\theta|\theta') = -\frac{1}{\lambda_1^2} \sum_{i=1}^n x_i p_{i,\theta'}(1) < 0.$$

La fonction $\lambda_1 \mapsto Q(\theta|\theta')$ est alors strictement concave. Par conséquent, son maximum global est atteint au point $\bar\lambda_1$où la dérivée s'annule. Or,

$$\left.\frac{\partial}{\partial \lambda_1} Q(\theta|\theta')\right|_{\lambda=\bar\lambda} = 0 \iff \bar\lambda_1 = \frac{\sum_{i=1}^n x_i p_{i,\theta'}(1)}{\sum_{i=1}^n p_{i,\theta'}(1)} .$$

Pour λ_2 on obtient de la même façon que le maximum global est atteint au point $\bar\lambda_2$ donné par

$$\bar\lambda_2 = \frac{\sum_{i=1}^n x_i p_{i,\theta'}(2)}{\sum_{i=1}^n p_{i,\theta'}(2)} .$$

Quant à la maximisation en π on obtient pour les dérivées :

$$\frac{\partial}{\partial \pi} Q(\theta|\theta') = \frac{1}{\pi} \sum_{i=1}^n p_{i,\theta'}(1) - \frac{1}{1-\pi}\left(n - \sum_{i=1}^n p_{i,\theta'}(1)\right)$$
$$\frac{\partial^2}{\partial^2 \pi} Q(\theta|\theta') = -\frac{1}{\pi^2} \sum_{i=1}^n p_{i,\theta'}(1) - \frac{1}{(1-\pi)^2}\left(n - \sum_{i=1}^n p_{i,\theta'}(1)\right) < 0 .$$

Donc, $\pi \mapsto Q(\theta|\theta')$ est strictement concave avec maximum global en $\bar\pi$ donné par

$$\left.\frac{\partial}{\partial \pi} Q(\theta|\theta')\right|_{\pi=\bar\pi} = 0 \iff \bar\pi = \frac{1}{n} \sum_{i=1}^n p_{i,\theta'}(1) .$$

Il est clair que $\bar\pi$ appartient bien à $[0, 1]$.
Pour résumer, la fonction $\theta \mapsto Q(\theta|\theta')$ est maximale en $\bar\theta = (\bar\pi, \bar\lambda_1, \bar\lambda_2)$ donné par

$$\bar\pi = \frac{1}{n}\sum_{i=1}^n p_{i,\theta'}(1) , \quad \bar\lambda_j = \frac{\sum_{i=1}^n x_i p_{i,\theta'}(j)}{\sum_{i=1}^n p_{i,\theta'}(j)} , \quad j = 1, 2 ,$$

où les $p_{i,\theta'}(j)$ sont données par (2).

e) L'algorithme EM consiste à calculer itérativement $\theta^{(t+1)} = \arg\max_{\theta\in\Theta} Q(\theta|\theta^{(t)})$ pour $t \geq 0$ en commençant par un point initial $\theta^{(0)}$ choisi par l'utilisateur, jusqu'à convergence. Pour savoir si l'algorithme a convergé, on utilise un critère d'arrêt. Par exemple, soit $\varepsilon > 0$ un seuil fixé par utilisateur. On considère que l'algorithme a convergé dès que $\|\theta^{(t+1)} - \theta^{(t)}\| < \varepsilon$, où $\|\cdot\|$ désigne par exemple la norme euclidienne ou la norme sup.

Voici un schéma d'un programme qui met en œuvre l'algorithme EM.

```
# initialisation
choisir un point initial θ^(0)
fixer un seuil ε
poser crit.arret = FALSE
poser t = 0

# algorithme EM
while (crit.arret == FALSE)
      calculer p_{i,θ^(t)}(j) pour j = 1,2 par Equation (2).
      calculer θ^(t+1) = (π^(t+1), λ_1^(t+1), λ_2^(t+1)) par
```

$$\pi^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} p_{i,\theta^{(t)}}(1)\ , \quad \lambda_j^{(t+1)} = \frac{\sum_{i=1}^n x_i p_{i,\theta^{(t)}}(j)}{\sum_{i=1}^n p_{i,\theta^{(t)}}(j)}\ , \quad j = 1, \dots m.$$

```
      if ||θ^(t+1) − θ^(t)|| < ε
           poser crit.arret = TRUE
      poser t = t + 1
```

f) En général, il n'y a pas de garantie que l'algorithme EM converge vers l'EMV. Néanmoins, on sait qu'à chaque itération il augmente la log-vraisemblance, et, sous des conditions assez générales, il converge vers un point critique de la fonction de log-vraisemblance. En revanche, si ce point critique est le maximum global de la log-vraisemblance dépend essentiellement du point initial. En pratique, on lance l'algorithme EM plusieurs fois avec différents points initiaux et on retient la solution avec la plus grande valeur de la log-vraisemblance, en espérant qu'il s'agit bien le maximum global.