# Computationnal Statistics - TP2 - BLANCHARD Théau

## Exercice 1:

1) Soit $X$ une variable aléatoire discrète à valeurs dans $(x_i)_{i\in[\![1,n]\!]}$ telle que $\forall i \in [\![1,n]\!]$, $\mathbb{P}(X=x_i)=p_i$

Sa fonction de répartition est donc: $F_X(t)=\sum_{i=1}^{n} \mathbb{1}_{\{x_i \leq t\}} p_i$.

Son pseudo-inverse est alors:

$$F_X^{-1}(u)=\mathbb{1}_{[0,p_1]}(u)\,x_1+\sum_{i=1}^{n-1}\mathbb{1}_{]\sum_{k=1}^{i} p_k,\sum_{k=1}^{i+1} p_k]}(u)\,x_{i+1}$$

Par le théorème d'inversion, si $U\sim\mathcal{U}([0,1])$, alors $F_X^{-1}(U)$ suit la même loi que $X$.

## Exercice 2:

1) Supposons que $x_i\in\mathbb{R}^d$. Alors chaque gaussienne est décrite par $\mu_j\in\mathbb{R}^d$, $\Sigma_j\in\mathbb{R}^{d\times d}$ et $\alpha_j\in\mathbb{R}$. Soit $d^2+d+1$ paramètres.

Au total on a donc $m(d^2+d+1)$ paramètres et $\theta=(\mu_j,\Sigma_j,\alpha_j)_{j\in[\![1,m]\!]}$.

Soit $i\in[\![1,n]\!]$, on a:

$$f_\theta(x_i)=\sum_{j=1}^{m} f_\theta(z_i=j)\,f_\theta(x_i|z_i=j)=\sum_{j=1}^{m}\alpha_j\,f_\theta(x_i|z_i=j).$$

Alors:

$$\mathcal{L}(x_1,\dots,x_n|\theta)=\prod_{i=1}^{n} f_\theta(x_i)$$

$$=\prod_{i=1}^{n}\sum_{j=1}^{m}\alpha_j\,f_\theta(x_i|z_i=j)$$

Ou encore:

$$\mathcal{L}(x_1,\dots,x_n|\theta)=\prod_{i=1}^{n}\sum_{j=1}^{m}\alpha_j\frac{1}{\sqrt{(2\pi)^d\det(\Sigma_j)}}\exp\left(-\frac{1}{2}(x_i-\mu_j)\,\Sigma_j^{-1}(x_i-\mu_j)\right)$$

3°) On a $f_\theta(X_i, Z_i) = \sum_{j=1}^{m} \mathbb{1}_{\{Z_i=j\}} f_\theta(Z_i=j) \cdot f_\theta(x_i|Z_i)$

$$= \sum_{j=1}^{m} \mathbb{1}_{\{Z_i=j\}} \alpha_j \cdot f_\theta(x_i|Z_i).$$

D'où la log vraisemblance des données complètes :

$$\mathcal{L}((x_i, Z_i)|\theta) = \sum_{i=1}^{m} \log\left(\sum_{j=1}^{m} \mathbb{1}_{\{Z_i=j\}} \alpha_j \cdot f_\theta(x_i|Z_i)\right)$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{m} \mathbb{1}_{\{Z_i=j\}} \left(\log(\alpha_j) + \log(f_\theta(x_i|Z_i))\right)$$

Étape E :

On pose $Q(\theta|\theta^{(t)}) = \mathbb{E}_{Z|X,\theta^t}\left[\log(\mathcal{L}((X_i, Z_i)|\theta))\right]$. D'où

$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbb{E}_{Z_i|X_i,\theta^{(t)}}\left[\mathbb{1}_{\{Z_i=j\}}\right] \left(\log(\alpha_j) + \log f_\theta(x_i|Z_i)\right)$$

Donc si on pose $z_{ij}^{(t)} = \mathbb{E}_{Z|X_i,\theta^{(t)}}\left[\mathbb{1}_{\{Z_i=j\}}\right] = \mathbb{P}(Z_i = j | X_i, \theta^t)$ on a par

la formule de Bayes : $\boxed{z_{ij}^{(t)} = \frac{\alpha_j^{(t)} f_{\theta^{(t)}}(x_i|Z_i=j)}{\sum_{k=1}^{m} \alpha_k^{(t)} f_{\theta^{(t)}}(x_i|Z_i=k)}}$

Étape M :

Alors on peut écrire :

$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n} \sum_{j=1}^{m} z_{ij}^{(t)}\left(\log \alpha_j - \frac{1}{2}(x_i - \mu_j)^T \Sigma_j^{-1}(x_i - \mu_j) - \frac{1}{2}\log\det \Sigma_j - \frac{d}{2}\log(2\pi)\right)$$

On cherche alors $\theta^{(t+1)} \in \underset{\theta}{\operatorname{argmax}}\, Q(\theta|\theta^{(t)})$ tel que $\sum \alpha_j^{(t+1)} = 1$.

D'où le lagrangien $L(\theta, \lambda) = Q(\theta|\theta^{(t)}) + \lambda\left(1 - \sum_{j=1}^{m} \alpha_j\right)$

Alors pour $k \in [\![1, m]\!]$.

- $\frac{\partial L}{\partial \alpha_k} = 0 \iff \frac{1}{\alpha_k} \cdot \sum_{i=1}^{m} z_{ik} - \lambda = 0 \iff \alpha_k = \frac{1}{\lambda} \cdot \sum_{i=1}^{n} z_{ik}$

Or $\sum_{j=1}^{n} \alpha_j = 1$ d'où $1 = \frac{1}{\lambda} \sum_{i=1}^{m} \sum_{k=1}^{n} \tau_{ik}$

D'où $\lambda = \sum_{i=1}^{n} \sum_{k=1}^{m} \tau_{ik} = \sum_{i=1}^{n} \sum_{k=1}^{m} \mathbb{P}(Z_i = k \mid x_i, \theta^t) = m$

Finalement on a donc : $\boxed{\alpha_k^{(t+1)} = \frac{1}{m} \sum_{i=1}^{n} \tau_{ik}^{(t)}}$

- $\frac{\partial L}{\partial \mu_k} = 0 \iff 0 = \frac{1}{2} \sum_{i=1}^{m} \tau_{ik} \Sigma_k^{-1} . (x_i - \mu_k)$

$\iff \left(\sum_{i=1}^{m} \tau_{ik}\right) . \Sigma_k^{-1} \mu_k = \Sigma_k^{-1} \cdot \left(\sum_{i=1}^{m} \tau_{ik} x_i\right)$

Finalement on a donc : $\boxed{\mu_k^{(t+1)} = \frac{\sum_{i=1}^{m} \tau_{ik}^{(t)} x_i}{\sum_{i=1}^{m} \tau_{ik}^{(t)}}}$

- Remarquons d'abord que $(x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) = \mathrm{Tr}\left((x_i - \mu_j)(x_i - \mu_j)^T \Sigma_j^{-1}\right)$

Or $\frac{\partial \mathrm{Tr}(AB)}{\partial B} = A$

De plus si $B \in GL_d(\mathbb{R})$ alors $B . B^{-1} = I_d$ donc $0 = \frac{\partial I_d}{\partial B} = \frac{\partial B}{\partial B} . B^{-1} + B \frac{\partial B^{-1}}{\partial B}$

ce qui donne $\frac{\partial B^{-1}}{\partial B} = - B^{-2}$

D'où $\frac{\partial \mathrm{Tr}(AB^{-1})}{\partial B} = \frac{\partial \mathrm{Tr}(AB^{-1})}{\partial B^{-1}} . \frac{\partial B^{-1}}{\partial B} = - AB^{-2}$.

De plus $\frac{\partial \log \det B}{\partial B} = B^{-T}$.

Donc $\frac{\partial L}{\partial \Sigma_k} = 0 \iff 0 = \sum_{i=1}^{n} \frac{1}{2} \tau_{ik} \left( (x_i - \mu_k)(x_i - \mu_k)^T \Sigma_k^{-2} - \Sigma_k^{-1} \right) = 0$

$\iff \left(\sum_{i=1}^{n} \tau_{ik}\right) \Sigma_k^{-1} = \left(\sum_{i=1}^{n} \tau_{ik} (x_i - \mu_k)(x_i - \mu_k)^T\right) . \Sigma_k^{-2}$

Finalement on a donc $\boxed{\Sigma_k^{(t+1)} = \frac{\sum_{i=1}^{m} \tau_{ik} (x_i - \mu_k)(x_i - \mu_k)^T}{\sum_{i=1}^{m} \tau_{ik}}}$

## Exercice 3 :

4) On constate que maximiser la fonction de (i;i) revient à modifier l'étape E de l'algorithme EM en écrivant

$$Q(\theta|\theta^{(t)}) = \sum_{i=1}^{n}\sum_{j=1}^{m} \tilde{\omega}_i^{(t)} \cdot \tau_{ij}^{(t)} \left( \log(\alpha_j) + \log(\varphi(x_i;\mu_j,\Sigma_j)) \right)$$

Il suffit alors de remplacer dans l'algorithme les $(\tau_{ij})_{i,j}$ par $(\tilde{\omega}_i \tau_{ij})_{i,j}$ de sorte que :

$$\boxed{\tilde{\tau}_{ij}^{(t)} = \frac{\tilde{\omega}_i^{(t)} \alpha_j^{(t)} \varphi(x_i;\mu_j^{(t)},\Sigma_j^{(t)})}{\sum_{k=1}^{m} \alpha_k^{(t)} \varphi(x_i;\mu_k^{(t)},\Sigma_k^{(t)})}}$$

Attention, on n'a plus que $\sum_{j=1}^{m} \tilde{\tau}_{ij} = 1$, il faut donc renormaliser les $(\alpha_j)_j$ au moment de la mise à jour.