

Exercice 1

[1] On considère les  $\{x_1, \dots, x_n\}$  triés dans l'ordre croissant. On cherche à appliquer la méthode de la transformée inverse, et pour cela on s'intéresse à l'inverse généralisée de la fonction de répartition de la loi ci-dessous qui est :

$$F_X(x) = \sum_{k=1}^i p_k$$

Son inverse généralisée  $F^{-1}(U) = \inf_{\{x_1, \dots, x_n\}} \{x \mid F(x) \geq U\}$   
se définit par :

$$F^{-1}(U) = x_i \iff U \in \left[ \sum_{k=1}^{i-1} p_k ; \sum_{k=1}^i p_k \right]$$

Si  $U \in \mathcal{U}_{(0,1)}$ ,  $F^{-1}(U)$  suit la loi ci-dessous

en effet :  $P(F^{-1}(U) = x_i) = P(U \in \left[ \sum_{k=1}^{i-1} p_k ; \sum_{k=1}^i p_k \right]) = p_i$

On a alors le pseudo-code suivant :

- Input :  $p = (p_1, \dots, p_n)$ ,  $x = (x_1, \dots, x_n)$   $x_1 < \dots < x_n$
- Compute  $P = (p_1, p_1 + p_2, \dots, \sum_{k=1}^n p_k)$
- Générer  $U \in \mathcal{U}_{(0,1)}$
- Trouver l'unique  $i_0$  /  $P_{i_0-1} < U < P_{i_0}$
- Return  $x_{i_0}$  (qui est  $F^{-1}(U)$ )

## Exercice 2

i) Les paramètres sont ici :

$$\theta = (\alpha_j, \mu_j, \Sigma_j)_{1 \leq j \leq p}$$

$$\text{ou } \left\{ \begin{array}{l} (\alpha_j)_j \in \left\{ \alpha \in \mathbb{R}^p \mid \sum_{j=1}^p \alpha_j = 1, \alpha_j > 0 \text{ } \forall j=1 \dots p \right\} \\ (\mu_j)_j \in \mathbb{R}^d \\ \Sigma_j \in S_n^+(d) \text{ } \underline{\text{symétrique}} \end{array} \right.$$

ii) Pour une observation  $x$  donnée :

$$P_\theta(x) = \int_Z P_\theta(x, z) dz = \int_Z P_\theta(z) P_\theta(x|z) dz$$

$$= \sum_{j=1}^p \underbrace{P_\theta(z_j)}_{\alpha_j} \underbrace{P_\theta(x|z=j)}_{\star}$$

$$\text{où on a : } \star = P_\theta(x|z=j) := \mathcal{N}(x; \mu_j, \Sigma_j) \\ = \frac{1}{(2\pi)^{d/2}} \frac{1}{\sqrt{\det \Sigma_j}} \exp \left\{ -\frac{1}{2} (x - \mu_j)^T \Sigma_j^{-1} (x - \mu_j) \right\}$$

Il vient alors pour la vraisemblance des  $(x_i - \bar{x})$  iid :

$$P_\theta(x_1 - \bar{x}_n) = \prod_{i=1}^n P_\theta(x_i) = \prod_{i=1}^n \sum_{j=1}^p \frac{\alpha_j}{(2\pi)^{d/2}} \frac{1}{\sqrt{|\Sigma_j|}} \exp \left\{ -\frac{1}{2} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \right\}$$

## 2/ NOTEBOOK .

3/ On s'intéresse donc ici à l'algorithme EM pour le GRN.

À l'étape  $(\theta)_t$ , on cherche à résoudre le programme, à  $\theta^{(t)} = (\alpha^{(t)}, \mu^{(t)}, \Sigma^{(t)})$  donné

$$\theta^{t+1} \in \underset{\alpha_j > 0}{\operatorname{argmax}} \mathbb{E}_{z_i \sim p_t(\cdot | (x_i)_{1 \leq i \leq n})} [\log p_\theta((x_i)_i, (z_i)_i)]$$

$$= \underset{\alpha_j > 0}{\operatorname{argmax}} \sum_{i=1}^n \underbrace{\mathbb{E}_{z_i \sim p_t(\cdot | x_i)} [\log(p_\theta(x_i, z_i))]}_{\text{s.c. } \sum \alpha_j = 1} := F_i(\theta)$$

Car  $p_\theta((x_i)_i, (z_i)_i) = \prod_{i=1}^n p_\theta(x_i, z_i)$  par indépendance des  $((x_i, z_i))_{1 \leq i \leq n}$ .

On a :

$$F_i(\theta) = \sum_{j=1}^p \underbrace{\log(p_\theta(x_i, z_i=j))}_{T_{ij}} \underbrace{p_t(z_i=j | x_i)}_{\tau_{ij}}$$

On calcule la loi a posteriori de  $z_i$  :

$$p_t(z_i=j | x_i) = \frac{p_t(x_i | z_i=j) \underbrace{p_t(z_i=j)}_{\alpha_j^{(t)}}}{\sum_{k=1}^p p_t(x_i | z_i=k) p_t(z_i=k)} \quad (\text{Bayes})$$

$$= \frac{\alpha_j^{(t)} \mathcal{N}(x_i; \mu_j^{(t)}, \Sigma_j^{(t)})}{\sum_{k=1}^p \alpha_k^{(t)} \mathcal{N}(x_i; \mu_k^{(t)}, \Sigma_k^{(t)})} \equiv \tau_{ij}$$

où  $\mathcal{N}(x_i; \mu_j, \Sigma_j)$  est la densité  $\mathcal{N}(\mu_j, \Sigma_j)$  en  $x_i$

→ indépendant de  $\theta$  / dépend uniquement de  $\theta^{(t)}$

On a :

$$\log(P_0(x_i, z_i=j)) = \log(P_0(x_i | z_i=j) P_0(z_i=j))$$

$$= \log(\pi(x_i; \mu_j, \Sigma_j)) + \log(\alpha_j)$$

$$= \text{constante} - \frac{1}{2} \log \det \Sigma_j - \frac{1}{2} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) + \log(\alpha_j)$$

Si bien qu'en fine, le programme se réduit à :

$$\theta^{t+1} \in \underset{\substack{\theta \\ \alpha_j > 0}}{\operatorname{argmax}} \sum_{i=1}^N \sum_{j=1}^p \tau_{ij} \left[ \log(\alpha_j) - \frac{1}{2} \log \det \Sigma_j - \frac{1}{2} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \right]$$

$$= \sum_{j=1}^p \left\{ -\frac{1}{2} \log \det \Sigma_j T_j + T_j \log(\alpha_j) - \frac{1}{2} \sum_{i=1}^n \tau_{ij} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \right\}$$

s.c.  $\sum_{j=1}^p \alpha_j = 1$  où  $T_j = \sum_{i=1}^n \tau_{ij}$

La fonction objectif est concave en tant que somme fine des  $(F_i)_i$ , elles-mêmes étant concaves par combinaison convexe ( $\tau_{ij} \geq 0 \forall i, j$ ,  $\forall i : \sum_{j=1}^p \tau_{ij} = 1$ ) des fonctions concaves ( $\theta \mapsto \log(P_0(x_i, z_i=j))$ ) pour  $i \in \{1, n\}$ . La contrainte est linéaire.

Donc tout point critique du Lagrangien est un maximum global ( $\nabla_\theta L = 0$ )

de Lagrange est, pour  $\lambda \in \mathbb{R}$  :

$$\mathcal{L}(x, \theta, \lambda) = \sum_{j=1}^p \left\{ -\frac{1}{2} \log \det \Sigma_j T_j + T_j \log(\alpha_j) - \frac{1}{2} \sum_{i=1}^n \tau_{ij} (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \right\} + \lambda \left\{ 1 - \sum_{j=1}^p \alpha_j \right\}.$$

Determinons les points critiques :

$$\cdot \nabla_{\mu_j} \mathcal{L} = 0 \Leftrightarrow 2 \sum_{i=1}^n \tau_{ij} \Sigma_j^{-1} (x_i - \mu_j) = 0$$

$\Leftrightarrow$

$$\mu_j^{(t+1)} = \frac{\sum_{i=1}^n \tau_{ij} x_i}{\tau_j} \quad \forall j$$

$$\cdot \text{On utilise que } \nabla_{\Sigma_j} \log \det \Sigma_j = -\nabla_{\Sigma_j} \log \det \Sigma_j^{-1} = -\Sigma_j$$

$$\text{et que } (x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j)$$

$$= T_2 \left[ \Sigma_j^{-1} (x_i - \mu_j) (x_i - \mu_j)^T \right]$$

$$\text{et que } \nabla_{\Sigma_j} T_2 (\Sigma_j^{-1} (x_i - \mu_j) (x_i - \mu_j)^T)$$

$$= (x_i - \mu_j) (x_i - \mu_j)^T$$

ou  $x_i - \mu_j$   
est une  
colonne

Pour obtenir

$$\nabla_{\Sigma_j} \mathcal{L} = 0 \Leftrightarrow \frac{T_j}{2} \Sigma_j - \frac{1}{2} \sum_i \tau_{ij} (x_i - \mu_j) (x_i - \mu_j)^T = 0$$

$$\Leftrightarrow \Sigma_j^{(t+1)} = \frac{1}{T_j} \sum_i \tau_{ij} (x_i - \mu_j) (x_i - \mu_j)^T$$

$$\bullet \quad \nabla_{\alpha_j} = 0 \quad (\Rightarrow) \quad \frac{\tau_j}{\alpha_j} = \lambda$$

$$\Leftrightarrow \sum_{j=1}^p \tau_j = \lambda \sum_{j=1}^p \alpha_j = \lambda$$

$$\text{or} \quad \sum_{j=1}^p \tau_j = \sum_{j=1}^p \sum_{i=1}^N \tau_{ij} = \sum_{i=1}^N \sum_{j=1}^p \tau_{ij} = \sum_{i=1}^N \underbrace{\sum_{j=1}^p P(z_j | x_i)}_{=1} = N$$

$$\text{D'où} \quad \frac{\tau_j}{\alpha_j} = N \quad \Leftrightarrow \boxed{\alpha_j^{(t+1)} = \frac{\tau_j}{N} \quad \forall j}$$

• À l'étape  $(E)_{t+1}$ , on cherche à évaluer la loi a posteriori  $P_{\theta^{(t+1)}}(z_i | x_i) \quad \forall i$ , autrement dit on "met à jour" les  $\tau_{ij}$  ("les responsabilités") avec le nouveau  $\theta^{(t+1)}$ . Si bien que le pseudo-code de l'EN devient dans le cas GNN :

$$\underline{\text{Init}} \quad \theta = (\alpha, \Sigma, \mu)$$

$$\underline{\text{Repeat}} \quad (E): \quad \tau_{ij} \leftarrow \frac{\alpha_j \mathcal{N}(x_i; \mu_j, \Sigma_j)}{\sum_{k=1}^p \alpha_k \mathcal{N}(x_i; \mu_k, \Sigma_k)} \quad \forall i, \forall j$$

Until  
Convergence

$$(P): \quad \mu_j \leftarrow \frac{1}{\tau_j} \sum_{i=1}^N \tau_{ij} x_i$$

$$\Sigma_j \leftarrow \frac{1}{\tau_j} \sum_{i=1}^N \tau_{ij} (x_i - \mu_j)(x_i - \mu_j)^T \quad \forall j$$

$$\alpha_j \leftarrow \tau_j / N \quad \forall j$$

### Exercice 3

4/ Si  $q_{\theta}(x_i^{(0)}) = p_{\theta}(x_i)$  en reprenant les notations du bnn (exercice 2) ( $x_i \sim w(\mu_j, \Sigma_j)$ ,  $x_i = z_i = j$ ,  $P(z_i = j) = \alpha_j$ )

On garde, comme dans le cadre du bnn, la forme inférieure sur la log-vraisemblance  $\mathcal{L}(\theta)$ ,

$$\forall i, w_i^{(0)} \log \left( q_{\theta}(x_i^{(0)}) \right) \geq w_i^{(0)} Q_i(\theta, \theta^{(t)})$$

$$\text{car } w_i^{(0)} \leq \mathcal{J}(x_i^{(0)}) / q_{\theta^{(0)}}(x_i^{(0)}) \geq 0$$

$$\text{avec } Q_i(\theta, \theta^{(t)}) = \mathbb{E}_{z_i \sim p(\cdot | x_i, \theta^{(t)})} [\log p_{\theta}(x_i, z_i)]$$

En reprenant le fil de l'exercice 2 q. 3), on peut donc appliquer un algorithme EN [complet] pour répondre à la question. On initialise l'EN à  $\theta^{(0)}$  (celui de Population n.c), et au sein de l'EN, à  $\theta^{(t)}$  (celui de l'EN cette fois) donne :

Etape M<sub>t</sub> Le programme devient :

$$\theta^{(t+1)} \in \operatorname{argmax}_{\theta} \sum_{i=1}^N w_i^{(0)} \underbrace{\mathbb{E}_{z_i \sim p_{\theta}(\cdot | x_i)} [\log(p_{\theta}(x_i, z_i))]}_{\text{s.c. } \sum \alpha_j = 1} := F_i(\theta) (= Q_i(\theta, \theta^{(0)}))$$

Le calcul des  $F_i$  ne change pas. On a par suite :

$$\theta^{(t+1)} \in \underset{\substack{\theta \\ \alpha_j > 0}}{\operatorname{argmax}} \sum_{i=1}^N w_i^{(t)} \sum_{j=1}^p \tau_{ij} \left[ \log(\alpha_j) - \frac{1}{2} \log \det \Sigma_j - \frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_j) \right]$$

$$= \sum_{j=1}^p \left\{ \log(\alpha_j) (\omega\tau)_j - \frac{1}{2} (\omega\tau)_j \log \det \Sigma_j - \frac{1}{2} \sum_{i=1}^N w_i^{(t)} \tau_{ij} (\mathbf{x}_i - \boldsymbol{\mu}_j)^T \Sigma_j^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_j) \right\}$$

s.c.  $\sum_{j=1}^p \alpha_j = 1$

or  $(\omega\tau)_j = \sum_{i=1}^N w_i^{(t)} \tau_{ij}$

Le problème est toujours convexe ( $w_i^{(t)} \geq 0$ ) :  
on calcule toujours un point critique du Lagrangien :

- $\nabla_{\boldsymbol{\mu}_j} \mathcal{L} = 0 \Leftrightarrow 2 \sum_{i=1}^N \tau_{ij} w_i^{(t)} \Sigma_j^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_j) = 0$

$$\Leftrightarrow \boldsymbol{\mu}_j^{(t+1)} = \frac{\sum_{i=1}^N w_i^{(t)} \tau_{ij} \mathbf{x}_i}{(\omega\tau)_j} \quad \forall j$$

- $\nabla_{\Sigma_j} \mathcal{L} = 0 \Leftrightarrow \frac{(\omega\tau)_j}{2} \Sigma_j - \frac{1}{2} \sum_i w_i^{(t)} \tau_{ij} (\mathbf{x}_i - \boldsymbol{\mu}_j) (\mathbf{x}_i - \boldsymbol{\mu}_j)^T = 0$

$$\Leftrightarrow \Sigma_j^{(t+1)} = \frac{1}{(\omega\tau)_j} \sum_i w_i^{(t)} \tau_{ij} (\mathbf{x}_i - \boldsymbol{\mu}_j) (\mathbf{x}_i - \boldsymbol{\mu}_j)^T \quad \forall j$$

$$\bullet \quad \nabla_{\alpha_j} = 0 \quad (\Rightarrow) \quad \frac{(\omega\tau)_j}{\alpha_j} = \gamma$$

$$\Rightarrow \sum_{j=1}^p (\omega\tau)_j = \gamma \sum \alpha_j = \gamma$$

$$\text{or} \quad \sum_{j=1}^p (\omega\tau)_j = \sum_{j=1}^p \sum_{i=1}^N w_i^{(t)} \tau_{ij} = \sum_{i=1}^N w_i^{(t)} \underbrace{\left( \sum_{j=1}^p \tau_{ij} \right)}_{= \sum_{j=1}^p p(z_i=j | x_i) = n} \\ \Rightarrow \gamma = w^{(t)} := \sum_{i=1}^N w_i^{(t)}$$

D'où

$$\alpha_j^{(t+1)} = \frac{(\omega\tau)_j}{w^{(t)}}$$

en notant que si  
les poids sont normalisés,  
 $w^{(t)} = n$

- à l'étape  $(E)_{t+1}$  reste la même, étant donné  $\theta^{(t)}$ , on met à jour la postérior de  $z$  ie :

$$\tau_{ij} \leftarrow \frac{\alpha_j^{(t)} N(x_i; \mu_j^{(t)}, \Sigma_j^{(t)})}{\sum_{k=1}^p \alpha_k^{(t)} N(x_i; \mu_k^{(t)}, \Sigma_k^{(t)})}$$

On répète  $(E)$  et  $(\eta)$  jusqu'à convergance.

$\theta^{(t)}$  est mis à jour par la valeur finale des paramètres retournée par l'EN.