Emilio PICARD

Computational Statistics
TP - 2

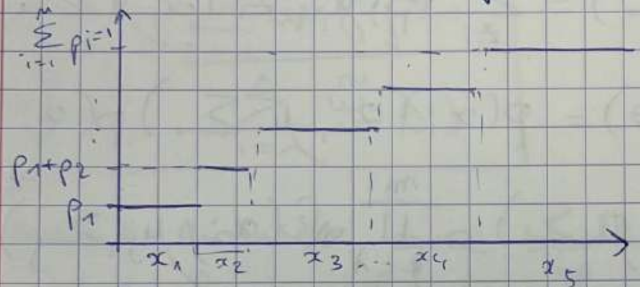
Exercice 1 :

(1) soit $(p_i)_i$ tels que $\sum_{i=1}^{n} p_i = 1$, $\mathbb{P}(X = x_i) = p_i$.

utilisons la fonction de répartition de X.

$F_X(x) = \mathbb{P}(X \leq x) = \sum_{i=1}^{x} p_i$.

On obtient une fonction de la sorte :



pour simuler X selon les $(p_i)_i$ on tire $\mathcal{U}([0,1])$. On note u le tirage.

$\forall i \in \{1, .., n\}$, si $p_{i-1} \leq u \leq p_i$, on simule $X = x_1$.

(2) pseudo-code :

```
simuler U ~ U([0, 1])
cum_prob ← cumsum(p)
for i in range(n)
    si cumprob[i] ≤ u ≤ cumprob[i+1] :
        return x(i)
```

Exercice 2 : (Notations : $X \sim \mathcal{N}(\mu, \Sigma)$, $f_X(x) = \mathcal{N}(x ; \mu, \Sigma)$)

(1) En utilisant les notations de l'énoncé, $\forall j \in \{1, \dots, m\}$
$\theta_j = (\alpha_j, \mu_j, \Sigma_j)$. (paramètres du modèle)

Calculons la vraisemblance de $(\theta, x_1, \dots, x_n)$:

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

$\forall i \in \{1, \dots, n\}$,
$$p(x_i | \alpha, \mu, \Sigma) = \sum_{z_i} p(x_i, z_i | \alpha, \mu, \Sigma)$$

or, $p(x_i, z_i | \theta) = p(x_i | z_i, \mu, \Sigma) \, p(z_i | \alpha)$

et $p(x_i | z_i, \mu, \Sigma) = \prod_{j=1}^{m} \mathcal{N}(x_i ; \mu_j, \Sigma_j)^{z_{ij}}$

donc $p(x_i, z_i | \alpha, \mu, \Sigma) = \prod_{j=1}^{m} \alpha_j \mathcal{N}(x_i ; \mu_j, \Sigma_j)^{z_{ij}}$

d'où $p(x_i | \alpha, \mu, \Sigma) = \sum_{z_i} \left[\prod_{j=1}^{m} \alpha_j \mathcal{N}(x_i ; \mu_j, \Sigma_j)^{z_{ij}} \right]$

$= \alpha_1 \mathcal{N}(x_i ; \mu_1, \Sigma_1) + \alpha_2 \mathcal{N}(x_i ; \mu_2, \Sigma_2) + \dots + \alpha_m \mathcal{N}(x_i ; \mu_m, \Sigma_m)$

$= \sum_{j=1}^{m} \alpha_j \mathcal{N}(x_i ; \mu_j, \Sigma_j)$

Finalement, on obtient :

$$\boxed{\begin{aligned} \mathcal{L}(x_1, \dots, x_n, \theta) &= \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_j \mathcal{N}(x_i ; \mu_j, \Sigma_j) \\ &= \prod_{i=1}^{n} \sum_{j=1}^{m} \frac{\alpha_j}{\sqrt{2\pi |\Sigma_j|}} e^{-\frac{1}{2}(x_i - \mu_j) \Sigma_j^{-1} (x_i - \mu_j)} \end{aligned}}$$

$\square$

(3) $\forall (i,j) \in \{1, \dots, n\} \times \{1, \dots, m\}$,

$$\gamma_{ij} = \frac{\alpha_j \mathcal{N}(x_i ; \mu_j, \Sigma_j)}{\sum_{l=1}^{m} \alpha_l \mathcal{N}(x_i ; \mu_l, \Sigma_l)}$$

calculs de $\hat{\alpha}, \hat{\mu}, \hat{\Sigma}$ (par rapport aux γ_i)

$\forall j \in \{1, \dots, m\}$

$$\begin{cases} \hat{\alpha}_j = \frac{1}{n} \sum_{i=1}^{n} \gamma_{ij} \\ \hat{\mu}_j = \frac{1}{n_j} \sum_{i=1}^{n} \gamma_{ij} x_i \\ \hat{\Sigma}_j = \frac{1}{n_j} \sum_{i=1}^{n} \gamma_{ij} (x_i - \hat{\mu}_j)(x_i - \hat{\mu}_j)^T \end{cases}$$

(n_j : number of observa° in cluster j)
$n_j = \sum_{i=1}^{n} \gamma_{ij}$

Une fois EM exécuté,

(4) la prédiction des paramètres α, μ et Σ sont assez proches des paramètres réels. Ces résultats sont illustrés dans le notebook.

(5) Application :

Après observation des données en 2D, la distribution ne ressemble pas à un ~~mélo~~ modèle de mélange gaussien (pas d'observation de clusters ni de ballon de rugby (forme d'une gaussienne en 2 dimensions)

Exercice 3 :

③ Pour $\mu = 0{,}8$, l'estimation $\mathbb{E}[f(X)]$ est bonne, alors que pour $\mu = 6$, beaucoup de poids sont des 0, et la variance est plus élevée. (voir notebook)

④. dans EM pour un modèle de mélange gaussien, la vraisemblance à maximiser (avant l'ajout des variables latentes z_i) est :

$$\mathcal{L}(x_1 \dots x_n, \theta) = \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_j \mathcal{N}(x_i; \mu_j, \Sigma_j)$$

en prenant la log-vraisemblance :

$$l(x_1, \dots x_n, \theta) = \sum_{i=1}^{n} \log \sum_{j=1}^{m} \alpha_j \mathcal{N}(x_i; \mu_j, \Sigma_j)$$

Dans le Population Monte Carlo Algorithm, nous devons maximiser :

$$\sum_{i=1}^{n} \tilde{\omega}_i \log \sum_{j=1}^{m} \alpha_j \mathcal{N}(x_i; \mu_j, \Sigma_j) \quad (1)$$

pour EM, maximiser $l(x_1, \dots, x_n, \theta)$ revient à maximiser

$$\mathbb{E}_{Z}\left[l(x_1, \dots, x_n, z_1, \dots, z_n, \theta) \right] = \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} \log(\alpha_j \mathcal{N}(x_i; \mu_j, \Sigma_j))$$

et EM assure que ceci converge.

De la même manière, on obtient pour (1) :

$$\sum_{i=1}^{n} \sum_{j=1}^{m} \tilde{\omega}_i \gamma_{ij} \log(\alpha_j \mathcal{N}(x_i; \mu_j, \Sigma_j)) \quad (2)$$

donc par propriété, EM permet d'assurer que ceci converge.

Nous pouvons donc utiliser EM pour maximiser (1).

Dans ce cas, $\forall i \in \{1, \dots, n\}$, $\forall j \in \{1, \dots, m\}$, γ_{ij} devient $w_i \gamma_{ij}$. (2)

On peut retrouver grâce à ça les nouvelles valeurs des paramètres α, μ et Σ.

$$\frac{\partial (2)}{\partial \mu_j} = \frac{\partial}{\partial \mu_j}\left(\sum_{i=1}^{m} \tilde{\omega}_i \sum_{j=1}^{m} \gamma_{ij}\left(\log \alpha_j + \log\left(-\frac{1}{2}(x_i - \mu_j)^T \Sigma_j^{-1}(x_i - \mu_j)\right)\right)\right)$$

$$= \sum_{i=1}^{m} \tilde{\omega}_i \gamma_{ij} \frac{x_i - \mu_j}{\Sigma_j} = 0$$

$$\Leftrightarrow \sum_{i=1}^{m} \tilde{\omega}_i \gamma_{ij} x_i = \sum_{i=1}^{m} \tilde{\omega}_i \gamma_{ij} \mu_j$$

$$\Leftrightarrow \boxed{\hat{\mu}_j = \frac{\sum_{i=1}^{m} \tilde{\omega}_i \gamma_{ij} x_i}{\sum_{i=1}^{m} \tilde{\omega}_i \gamma_{ij}}}$$

De la même manière, avec les mêmes types de calculs que EM pour les modèles de mélanges gaussiens, on obtient :

$$\boxed{\hat{\Sigma}_j = \frac{\sum_{i=1}^{m} \tilde{\omega}_i \gamma_{ij} (x_i - \mu_j)^2}{\sum_{i=1}^{m} \tilde{\omega}_i \gamma_{ij}}} \quad \text{et} \quad \boxed{\hat{\alpha}_j = \frac{1}{m} \sum_{i=1}^{m} w_i \gamma_{ij}}$$

$\square$