Maxime CORLAY

TP2 Computational Statistics

Exercice 1 :

1.

- Tirer $U \sim \mathcal{U}([0,1])$
- Soit $J = \min\{j \in [\![1,n]\!] | U \leq \sum_{k=1}^{j} p_k\}$
- Poser $X = x_J$

2. Testons avec $n = 5$, $x_1 = 1$, $x_2 = 2$, $x_3 = 3$, $x_4 = 4$, $x_5 = 5$, $p_1 = 0{,}1$, $p_2 = 0{,}6$, $p_3 = 0{,}2$, $p_4 = 0{,}05$ et $p_5 = 0{,}05$. Voir le notebook correspondant.

3. Sans surprise, les fréquences obtenues sont très proches des probabilités initiales.

Exercice 2 :

1.

- $\theta = \left((\alpha_j)_{j \in [\![1,m]\!]}, (\mu_j)_{j \in [\![1,m]\!]}, (\Sigma_j)_{j \in [\![1,m]\!]}\right)$ ce qui fait un total de $m + dm + \frac{d(d-1)}{2} m = \frac{m}{2}(d^2 + d + 2)$ paramètres, en notant d la dimension de chaque x_i

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

Pour passer au log, il est peu pratique de garder la somme sur j. On peut introduire des variables latentes Z_i (vecteurs aléatoires) vérifiant $\mathbb{P}(\{Z_{ik} = 0 \ pour\ tout\ k \neq j\} \cap \{Z_{ij} = 1\}) = \alpha_j$.

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

3. Voir le reste sur le notebook correspondant.

Exercice 3 :

Voir dans le notebook.