

FEUILLE D'EXERCICE N° 9

ALGORITHME EM

Exercice 1. Mélange de lois de Poisson

Considérons un mélange de K lois de Poisson. On notera $\lambda_k > 0$ le paramètre de la loi de Poisson du k -ème composant du mélange, et π_k son poids. Notons $\theta = (\pi_1, \dots, \pi_K, \lambda_1, \dots, \lambda_K)$ le vecteur de paramètres inconnus du modèle qui vit dans

$$\Theta = \left\{ \theta = (\pi_1, \dots, \pi_K, \lambda_1, \dots, \lambda_K) : \pi_k \in]0, 1[, \sum_{k=1}^K \pi_k = 1, \lambda_k > 0, k = 1, \dots, K \right\}.$$

1. Soit $\theta = (\pi_1, \dots, \pi_K, \lambda_1, \dots, \lambda_K) \in \Theta$ fixé. Construire une variable aléatoire X d'un mélange de lois de Poisson de paramètre θ à partir de variables aléatoires V_k de loi de Poisson et une variable aléatoire Z qui modélise l'appartenance de groupe.
2. Calculer $\mathbb{P}_\theta(X = k)$ pour tout $k = 0, 1, \dots$.
3. Soit $\mathbf{x} = (x_1, \dots, x_n)$ une réalisation du vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_n)$ avec des variables aléatoires X_i i.i.d. du mélange de lois de Poisson de paramètre θ . Montrer que l'estimateur de maximum de vraisemblance de θ n'est pas explicite.
4. Pour approcher l'estimateur de maximum de vraisemblance de θ nous allons donc utiliser l'algorithme EM. Notons $\mathbf{z} = (z_1, \dots, z_n)$ les réalisations de la variable cachée Z associées aux observations $\mathbf{x} = (x_1, \dots, x_n)$.
 - a) Calculer les probabilités conditionnelles $\mathbb{P}_\theta(U = k | \mathbf{X} = \mathbf{x})$ pour $k = 1, \dots, K$.
 - b) Donner la fonction de vraisemblance $\mathcal{L}(\mathbf{x}, \mathbf{z}; \theta)$ et la fonction de log-vraisemblance des données complètes (\mathbf{x}, \mathbf{z}) .
 - c) Calculer l'espérance conditionnelle $Q(\theta | \theta') = \mathbb{E}_{\theta'}[\log(\mathcal{L}(\mathbf{x}, \mathbf{Z}; \theta)) | \mathbf{X} = \mathbf{x}]$.
 - d) Trouver le point maximal de $\theta \mapsto Q(\theta | \theta')$ sur Θ .
 - e) Détailler la mise en œuvre de l'algorithme EM.
 - f) Cet algorithme, converge-t-il toujours vers l'estimateur de maximum de vraisemblance de θ ?