Mélange de lois de Poisson

Considérons un mélange de K lois de Poisson. Pour $1 \leq k \leq K$, nous noterons $\lambda_k > 0$ le paramètre de la k-ème composante et $\pi_k \in (0,1)$ son poids. Notons $\theta = (\pi_1, \ldots, \pi_K, \lambda_1, \ldots, \lambda_K)$ le paramètre inconnu et

$$\Theta = \left\{ \theta = (\pi_1, \ldots, \pi_K, \lambda_1, \ldots, \lambda_K) \,;\, \forall k \in \{1, \ldots, K\}\,,\, \pi_k \in (0,1)\,,\, \lambda_k > 0\,,\, \sum_{k=1}^{K} \pi_k = 1 \right\}.$$

1. Soit $\theta \in \Theta$, expliquer comment construire une variable aléatoire X suivant un mélange de lois de Poisson paramétré par θ.
2. Notons $\mathbb{P}_\theta$ la loi de X. Pour tout $j \geq 0$, calculer $\mathbb{P}_\theta(X = j)$.
3. Soit $\theta \in \Theta$ et $(x_1, \ldots, x_n) \in \mathbb{N}^n$. Calculer $\log \mathbb{P}_\theta(X_{1:n} = x_{1:n})$ où les $(X_i)_{1 \leq i \leq n}$ sont i.i.d de même loi que X.
4. Puisque nous ne pouvons pas maximiser la logvraisemblance explicitement, nous allons utiliser l'algorithme Expectation Maximization.
   (a) Pour tout $\theta \in \Theta$ et tout $k \in \{1, \ldots, K\}$, calculer $\mathbb{P}_\theta(Z = k | X = j)$.
   (b) Calculer la logvraisemblance complète des données.
   (c) Calculer la quantité intermédiaire de l'algorithme EM.
   (d) En déduire la mise à jour d'une itération de l'algorithme EM.
   (e) Détailler le fonctionnement complet de l'algorithme EM
   (f) Cet algorithme converge t'il vers le maximum de vraisemblance ?

Mélange de régressions linéaires

Considérons des covariables $(x_i)_{i=1}^n$, supposées déterministes, où $x_i \in \mathbb{R}^d$ pour $1 \leq i \leq n$ et le modèle suivant paramétré par

$$\theta = (\pi_1, \ldots, \pi_K, \beta_1, \ldots, \beta_K, \sigma_1^2, \ldots, \sigma_K^2)\,.$$

- Les variables $(Z_i)_{1 \leq i \leq n}$ sont non observées, indépendantes et telles que $Z_i \in \{1, \ldots, K\}$ et $\mathbb{P}_\theta(Z_i = k) = \pi_k$ pour $1 \leq k \leq K$, $1 \leq i \leq n$.
- Conditionnellement aux variables $(Z_i)_{1 \leq i \leq n}$, les observations réelles $(Y_i)_{1 \leq i \leq n}$ sont indépendantes et pout tout $1 \leq i \leq n$, la loi conditionnelle de Y_i est gaussienne, de moyenne $x_i^\top \beta_{Z_i}$ et de variance $\sigma_{Z_i}^2$.

1. Écrire la logvraisemblance complète de $(Z_{1:n}, Y_{1:n})$.
2. Soit $\theta^{(p)}$ l'estimateur de θ à l'itération $p \geq 0$ de l'algorithme EM. Donner la quantité intermédiaire $\theta \mapsto Q(\theta, \theta^{(p)})$.
3. Donner la mise à jour des poids π_k, $1 \leq k \leq K$.
4. Donner la mise à jour des β_k, $1 \leq k \leq K$.
5. Donner la mise à jour des variances σ_k^2, $1 \leq k \leq K$.